Introduction to Matrix Computations

Exercises and Even numbered Solutions

Margot Gerritsen
margot.gerritsen@stanford.edu

September 12, 2014

Contents

Preface

This supplement to the Introduction to Matrix Computations notes provides answers to around half of the exercises, mostly the even numbered ones. The exercises themselves are all included for completeness and the layout is spacious to give some room for note taking.

From time to time, we may refer to one of the reference books, listed below, if a particular answer is provided there in detail. All these books can be found in the Stanford library.

- Jim Demmel, Applied Numerical Linear Algebra
- Tim Davis, Direct Methods for Sparse Linear Systems
- Trefethen and Bau, Numerical Linear Algebra
- Golub and van Loan, Matrix Computations

Chapter 1

Exercises & Solutions

Exercise 1.1

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) If $A^2 + A = I$ then $A^{-1} = I + A$

(b) If all diagonal entries of A are zero, then A is singular (not invertible).

(c) $\|A\|_F^2 = \text{tr}\left(A^T A\right)$. Here $\text{tr}(A^T A)$ is the *trace* of $A^T A$. The trace of a matrix is the sum of its diagonal elements.

Answer for Exercise 1.1

Exercise 1.2

The product of two n by n lower triangular matrices is again lower triangular (all its entries above the main diagonal are zero). Prove it in general and confirm this with a 3 by 3 example.

Answer for Exercise 1.2

Suppose $AB = C$ and A and B are lower-triangular. Recall that, by property of matrix multiplication

$$c_{ij} = \vec{r_i}^T \vec{b_j}$$

Consider the i, j element of C. We only need to show that this element is necessarily 0 if $i < j$, i.e. above the diagonal, as this is the definition of lower-triangular matrix. Note that for a lower-triangular matrix all elements in i-th row after the i-th element are zero, and all elements in the j-th column up to the j-th element are as well. Schematically, we have

$$c_{ij} = \overbrace{\begin{bmatrix} a_1 & a_2 \dots & a_i \end{bmatrix}}^{i \text{ non-zeros}} \overbrace{\begin{bmatrix} 0 & 0 \dots 0 \end{bmatrix}}^{n-i \text{ zeros}}{}^T \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ b_{j+1} \\ b_{j+2} \\ \vdots \\ b_n \end{bmatrix} \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \\ \cdots \end{matrix} \begin{matrix} \left.\vphantom{\begin{matrix} a \\ a \\ a \\ a \end{matrix}}\right\} j-1 \text{ zeros} \\ \\ \\ \\ \left.\vphantom{\begin{matrix} a \\ a \\ a \\ a \end{matrix}}\right\} n-j+1 \text{ nonzeros} \\ \\ \\ \end{matrix} = 0$$

because $j - 1 \geq i$ and $n - j + 1 \leq n - i$.

The requested verification can be done very simply with MATLAB.

Exercise 1.3

If $A = A^T$ and $B = B^T$, which of the following matrices are certainly symmetric?

(a) $A^2 - B^2$

(b) $(A + B)(A - B)$

(c) ABA

(d) $ABAB$

Answer for Exercise 1.3

Exercise 1.4

A *skew-symmetric* matrix is a matrix that satisfies $A^T = -A$. Prove that if A is $n \times n$ and skew-symmetric, than for any $\vec{x}$, we must have that $\vec{x}^T A\vec{x} = 0$.

Answer for Exercise 1.4

Since A is $n \times n$, $\vec{x}$ must be $n \times 1$, which is consistent with our view of all vectors as *column* vectors. Then the product $A\vec{x}$ must also be a column vector of the same size as $\vec{x}$. Call this column vector $\vec{y}$. Then

$$\vec{x}^T A\vec{x} = \vec{x}^T(A\vec{x}) = \vec{x}^T\vec{y}$$

But we recognize $\vec{x}^T\vec{y}$ as an inner product, and so $\vec{x}^T A\vec{x}$ is a scalar quantity. Now, a nice property of scalars is that they equal their own transpose, so for any $\vec{x}$ we have that

$$\vec{x}^T A\vec{x} = (\vec{x}^T A\vec{x})^T$$

By skew symmetry of A, we then have that

$$\vec{x}^T A\vec{x} = (\vec{x}^T A\vec{x})^T = \vec{x}^T A^T\vec{x} = \vec{x}^T(-A)\vec{x} = -(\vec{x}^T A\vec{x})$$

The only scalar that satisfies this equation is 0.

Exercise 1.5

Suppose A is an $n \times n$ invertible matrix, and you exchange its first two rows to create a new matrix B. Is the new matrix B necessarily invertible? If so, how could you find B^{-1} from A^{-1}? If not, why not?

Answer for Exercise 1.5

Exercise 1.6

Let A be an invertible n-by-n matrix. Prove that A^m is also invertible and that

$$(A^m)^{-1} = \left(A^{-1}\right)^m$$

for $m = 1, 2, 3, \ldots$

Answer for Exercise 1.6

We want to show that for any positive integer m, the inverse of A^m is the inverse of A raised to the m-th power. In other words, we want to show that

$$A^m(A^{-1})^m = (A^{-1})^m A^m = I$$

We prove this by induction on m.

(i) First, we look at the base case, i.e., the case $m = 1$:

$$(A^1)^{-1} = A^{-1} = (A^{-1})^1.$$

(ii) Next, we *assume* that the result holds for some arbitrary value $m = k$:

$$\text{Assume } (A^k)^{-1} = (A^{-1})^k$$

(iii) Now, using steps i) and ii), we *show* the result holds for $m = k + 1$:

$$(A^{-1})^{k+1}A^{k+1} = A^{-1}(A^{-1})^k A^k A = A^{-1}(A^k)^{-1}A^k A = A^{-1}IA = I$$

and

$$A^{k+1}(A^{-1})^{k+1} = AA^k(A^{-1})^k A^{-1} = AA^k(A^k)^{-1}A^{-1} = AIA^{-1} = I$$

where we have used the result in part (ii). Thus, the general result follows by the induction principle.

Exercise 1.7

Let A be a 2×2 matrix $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ with $a_{11} \neq 0$ and let $\alpha = a_{21}/a_{11}$. Show that A can be factored into a product of the form

$$\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} \\ 0 & b \end{pmatrix}$$

What is the value of b?

Answer for Exercise 1.7

---

* Exercise 1.8

(a) Show that if the $n \times n$ matrices A, B and $A + B$ are invertible, then the matrix $B^{-1} + A^{-1}$ is also invertible.

(b) Assume that C is a skew-symmetric matrix and that D is a matrix defined as

$$D = (I + C)(I - C)^{-1}$$

Prove that $D^T D = DD^T = I$.

Answer for Exercise 1.8

(a) We must find a matrix, call it M, such that

$$(A^{-1} + B^{-1})M = I$$

Using the fact that A, B and $A+B$ are invertible, we will do some algebraic manipulations to simplify to something we are given. We write

$$\begin{aligned} I &= (A^{-1}+B^{-1})M \\ &= (A^{-1}+IB^{-1})M \\ &= (A^{-1}+A^{-1}AB^{-1})M \\ &= A^{-1}(I+AB^{-1})M \\ &= A^{-1}(BB^{-1}+AB^{-1})M \\ &= A^{-1}(B+A)B^{-1}M \end{aligned}$$

From the last equation, we see that

$$\left(A^{-1}(B+A)B^{-1}\right)M = I$$

Therefore, $M = \left(A^{-1}(A+B)B^{-1}\right)^{-1} = B(A+B)^{-1}A$

Now we verify that multiplying M on the right by the original matrix also gives the identity:

$$\begin{aligned} M(A^{-1}+B^{-1}) &= B(A+B)^{-1}A(A^{-1}+B^{-1}) \\ &= B(A+B)^{-1}(AA^{-1}+AB^{-1}) \\ &= B(A+B)^{-1}(I+AB^{-1}) \\ &= B(A+B)^{-1}(BB^{-1}+AB^{-1}) \\ &= B(A+B)^{-1}(B+A)B^{-1} \\ &= BB^{-1} \\ &= I \end{aligned}$$

Therefore, we have that, indeed $(B^{-1}+A^{-1})^{-1} = B(A+B)^{-1}A$

(b) The trick here is to observe that $(I+C)$ and $(I-C)$ commute, that is,

$$(I+C)(I-C) = (I-C)(I+C)$$

. Check this. Remember though that commuting is generally not guaranteed (so if you believe that matrices commute, you should always show that it is indeed true). Then

$$\begin{aligned} D^TD &= [(I+C)(I-C)^{-1}]^T(I+C)(I-C)^{-1} \\ &= [(I-C)^T]^{-1}(I+C)^T(I+C)(I-C)^{-1} \\ &= (I-C^T)^{-1}(I+C^T)(I+C)(I-C)^{-1} \\ &= (I+C)^{-1}(I-C)(I+C)(I-C)^{-1} \\ &= (I+C)^{-1}(I+C)(I-C)(I-C)^{-1} \\ &= I \end{aligned}$$

Going from the first line to the second line, we have used properties of the transpose as well as the fact that we can interchange transposition and inversion. Going from the second line to the third line, we have used the property for the transpose of a sum of matrices. Going from the third line to the fourth line, we've used the fact that C is skew-symmetric. Going from the fourth line to the fifth line, we're using the fact that $(I + C)$ and $(I - C)$ commute.

To show $DD^T = I$ we can use similar manipulations:

$$\begin{aligned} DD^T &= (I + C)(I - C)^{-1}[(I + C)(I - C)^{-1}]^T \\ &= (I + C)(I - C)^{-1}[(I - C)^T)^{-1}(I + C)^T] \\ &= (I + C)(I - C)^{-1}(I - C^T)^{-1}(I + C^T) \\ &= (I + C)(I - C)^{-1}(I + C)^{-1}(I - C) \\ &= (I + C)[(I + C)(I - C)]^{-1}(I - C) \\ &= (I + C)[(I - C)(I + C)]^{-1}(I - C) \\ &= (I + C)(I + C)^{-1}(I - C)^{-1}(I - C) \\ &= I \end{aligned}$$

where again we have used the fact that $(I - C)$ and $(I + C)$ commute.

*** Exercise 1.9**

Given an $n \times n$ matrix A with column vectors $\vec{a_1}, \vec{a_2}, \ldots, \vec{a_n}$, construct a matrix B such that the matrix AB has the columns $\vec{u_1}, \vec{u_2}, \ldots, \vec{u_n}$ with the following properties:

(a) $\vec{u_i} = \vec{a_i}, \quad i \neq j$

(b) $\vec{u_j} = \sum_{k=1}^{j} \alpha_k \vec{a_k}$,

where j is a *fixed* integer such that $1 \leq j \leq n$.

Answer for Exercise 1.9

Exercise 1.10

The well-known Fibonacci number $0, 1, 1, 2, 3, 4, \ldots$ can be computed using the formula:

$$F_{n+2} = F_{n+1} + F_n, \quad F_0 = 0, \quad F_1 = 1.$$

Show that the following holds for $n > 1$:

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{n-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} F_{n-1} \\ F_n \end{bmatrix}.$$

Answer for Exercise 1.10

The recursive formula for Fibonacci numbers is given by:

$$F_{n+2} = F_{n+1} + F_n, \quad F_0 = 0, F_1 = 1.$$

We can use inductive logic. Start from matrix below for n

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{n-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} F_{n-1} \\ F_n \end{bmatrix},$$

and prove that it is correct for $n + 1$

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

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{n-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} F_{n-1} \\ F_n \end{bmatrix} = \begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix}$$

$$\implies \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^{n} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix}.$$

Now we should check this system for $n = 1$:

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \implies F_1 = 1, F_2 = 1.$$

Exercise 1.11

If $A = LU$, where L is lower triangular with ones on the diagonal and U is upper triangular, show that L and U are unique.

Answer for Exercise 1.11

** Exercise 1.12

For any $n \times n$ matrix A. Show that the 2-norm and the Frobenius norm of the matrix satisfy

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\|A\|_2$$

Answer for Exercise 1.12

The proof is available in the introductory chapters (see section about norms) in Golub & van Loan.

Exercise 1.13

The matrix A is given by the factorization

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}$$

Show (**without** multiplying these two factors) that A is invertible, symmetric, tridiagonal, and positive definite.

Answer for Exercise 1.13

Just a quick note: in this question you are asked about positive definiteness. This is discussed in later chapters in the book. You may want to return to this subquestion once you've completed the chapter on eigenvalues and eigenvectors.

---

* Exercise 1.14

Suppose that the $n \times n$ matrix A has the property that the elements in each of its rows sum to 1. An example of such matrix is $\begin{bmatrix} 1/3 & 0 & 2/3 \\ 0 & 1 & 0 \\ 1/2 & 1/2 & 0 \end{bmatrix}$. Show that for any such $n \times n$ matrix A, the matrix $A - I$ (with I the $n \times n$ identity matrix) is singular.

Answer for Exercise 1.14

There are several ways to approach this problem. Here, we use the fact that if the nullspace dimension is larger than 0 (that is, if the nullspace contains more than just the zero vector), the matrix must be singular. So, we look for a vector $\vec{x}$ such that $(A - I)\vec{x} = \vec{0}$. $\vec{x} = \vec{1}$ (all entries equal to 1) is such a vector. Check this. Since $\mathcal{N}(A - I)$ contains a nonzero vector, $A - I$ must be singular.

---

* Exercise 1.15

(a) If $n \times n$ real matrix A is such that $A^2 = -I$, show that n must be even.

(b) If the $n \times n$ real matrix A is such that $A^2 = I$, show that A cannot be skew-symmetric.

Answer for Exercise 1.15

Exercise 1.16

We define the 1-norm of a vector $\vec{x}$ as $\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i|$, i.e. the sum of the magnitudes of the entries. The 1-norm of an $m \times n$ matrix A is defined as $\|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|$, i.e. the largest column sum. It can be shown that for these norms, $\|A\vec{x}\|_1 \le \|A\|_1 \|\vec{x}\|_1$ (we will not prove this here, but you can use it as given, whenever needed). Compute the 1-norm of the following matrices and vectors

(a) $\vec{x} = (1, 2, 3, \ldots, n)^T$.

(b) $\vec{x} = \left(\frac{1}{n}, \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right)^T$.

(c) αI where I is the $n \times n$ identity matrix.

(d) $J - I$ where J is the $n \times n$ matrix filled with 1s and I is the $n \times n$ identity.

(e) $A = \begin{bmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \\ & & & -1 & 2 \end{bmatrix}$, where A is an $n \times n$ matrix.

(f) $A = \begin{bmatrix} 1/n & \cdots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \cdots & 1/n \end{bmatrix}$, where A is an $n \times n$ matrix.

Answer for Exercise 1.16

(a) $\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i| = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$.

(b) $\|\vec{x}\|_1 = \sum_{i=1}^{n} |x_i| = \sum_{i=1}^{n} \frac{1}{n} = n\left(\frac{1}{n}\right) = 1$.

(c) Every column has sum exactly equal to $|\alpha|$, thus the max column sum is $|\alpha|$.

(d) Every column sums to $n - 1$, so the max column sum is $n - 1$.

(e) We have

$$\sum_{i=1}^{n}|a_{ij}| = 4, \quad 2 \leq j \leq n-1$$
$$\sum_{i=1}^{n}|a_{ij}| = 3, \quad j = 1, n$$

Thus,

$$\|A\|_1 = \max_{1\leq j\leq n}\sum_{i=1}^{n}|a_{ij}| = \max(3,4) = 4.$$

(f)

$$\|A\|_1 = \max_{1\leq j\leq n}\sum_{i=1}^{n}|a_ij| = \max_{1\leq j\leq n}\sum_{i=1}^{n}\frac{1}{n}\max_{1\leq j\leq n}1 = 1.$$

Chapter 2

Exercises & Solutions

Exercise 2.1

Compute the LU decomposition without pivoting (that is, without row exchanges) of the 4×4 matrix

$$A = \begin{bmatrix} 4 & 2 & -1 & 4 \\ 3 & 2 & 1 & 1 \\ -1 & 2 & 4 & 3 \\ 1 & 3 & -2 & 4 \end{bmatrix}.$$

Clearly show the intermediate matrices A', A'', A''', and C_1, C_2, C_3.

Answer for Exercise 2.1

Exercise 2.2

Solve $A\vec{x} = \vec{b}$ for the matrix from exercise 2.1 and $\vec{b} = \begin{bmatrix} 1 \\ 1 \\ 6 \\ 13 \end{bmatrix}$.

Answer for Exercise 2.2

We first compute the LU decomposition of the matrix. This is asked for in exercise 2.1. We assume now that we have this decomposition. Note that the matrices are not super pretty (but that's real life - I hardly ever encounter matrices that only have integers in them!).

$$A\vec{x} = \vec{b} \Rightarrow (LU)\vec{x} = \vec{b} \Rightarrow L(U\vec{x}) = \vec{b} \Rightarrow L\vec{y} = \vec{b},$$

where we define $U\vec{x} = \vec{y}$, in which $\vec{y}$ is a new (unknown) vector. In order to find $\vec{x}$ from the equation $U\vec{x} = \vec{y}$, we need $\vec{y}$ first, which can be computed from $L\vec{y} = \vec{b}$ as follows:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{3}{4} & 1 & 0 & 0 \\ \frac{-1}{4} & 5 & 1 & 0 \\ \frac{1}{4} & 5 & \frac{21}{10} & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 6 \\ 13 \end{bmatrix}$$

$$\begin{aligned} \Rightarrow y_1 &= 1, \\ \frac{3}{4}y_1 + y_2 &= 1, \\ \frac{-1}{4}y_1 + 5y_2 + y_3 &= 6, \\ \frac{1}{4}y_1 + 5y_2 + \frac{21}{10}y_3 + y_4 &= 13. \end{aligned}$$

We can solve for y_1, y_2, y_3, y_4 using forward substitution to get $y_1 = 1, y_2 = 1/4, y_3 = 5, y_4 = 1$. We can now recover $\vec{x}$ by solving $U\vec{x} = \vec{y}$ as follows:

$$\begin{bmatrix} 4 & 2 & -1 & 4 \\ 0 & \frac{1}{2} & \frac{7}{4} & -2 \\ 0 & 0 & -5 & 14 \\ 0 & 0 & 0 & -\frac{82}{5} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{4} \\ 5 \\ 1 \end{bmatrix}$$

$$\begin{aligned} \Rightarrow 4x_1 + 2x_2 - x_3 + 4x_4 &= 1, \\ \frac{1}{2}x_2 + \frac{7}{4}x_3 - 2x_4 &= \frac{1}{4}, \\ -5x_3 + 14x_4 &= 5, \\ -\frac{82}{5}x_4 &= 1. \end{aligned}$$

We can solve for x_4, x_3, x_2, x_1 using forward substitution to get $x_4 = -5/82, x_3 = -48/41, x_2 = 357/82, x_1 = -177/82$.

Exercise 2.3

Compute the LU decomposition of the matrix from exercise 2.1 using partial pivoting. Partial pivoting is a strategy in which rows are swapped to create the largest possible pivot at each stage of Gaussian Elimination.

Answer for Exercise 2.3

---

* Exercise 2.4

The minor diagonal of a matrix is the diagonal from the lower left corner of the matrix to the upper right corner. If a matrix has only zeros above this diagonal (so in the left corner of the matrix), the matrix is called a *reverse lower triangular* matrix. A 4×4 example is the matrix

$$R = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 3 \\ 0 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 \end{bmatrix}.$$

Let T be an arbitrary $n \times n$ reverse lower triangular matrix with non-zero elements on its minor diagonal ($n > 1$).

(a) Prove that T of this form is non-singular (prove this for any n).

(b) Explain why the LU factorization of T (for any n) can **not** be found if row exchanges are not permitted. This shows that even though a matrix is non-singular, a LU decomposition without row re-ordering may not exist.

(c) Prove, for any n, that if row exchanges are permitted, the LU factorization of PT can be computed, where P is the permutation matrix that performs the appropriate row exchanges. This shows that, with pivoting, it is always possible to find LU for a re-ordered form of the matrix. Note that perhaps the simplest way of proving this is to actually find the LU factorization.

Answer for Exercise 2.4

(a) Consider T is non-singular if

$$T\vec{x} = \vec{0} \Longleftrightarrow \vec{x} = \vec{0}.$$

First we will prove that $T\vec{x} = \vec{0} \implies \vec{x} = \vec{0}$. Assume $T\vec{x} = \vec{0}$. Let $t_{i,j}$ be the element (i, j) of matrix T. Since $t_{i,j} = 0$ for all $i < j$. We have

$$T\vec{x} = \begin{bmatrix} t_{1,n}x_n \\ t_{2,n-1}x_{n-1} + t_{2,n}x_n \\ \vdots \\ \sum_{k=1}^{n} t_{n,k}x_k \end{bmatrix}.$$

$$T\vec{x} = \vec{0} \implies t_{1,n}x_n = 0 \implies x_n = 0 \text{ (since } t_{1,n} \neq 0)$$

$t_{2,n-1}x_{n-1}+t_{2,n}x_n = 0$ and $x_n = 0 \implies t_{2,n-1}x_{n-1} = 0 \implies x_{n-1} = 0$ (since $t_{2,n-1} \neq 0$).

Repeating this process, at each step we obtain an equation of the form $t_{i,n-i+1}x_i = 0$, which implies that $x_i = 0$ since $t_{i,n-i+1}$ lies on the minor diagonal and is given to be non-zero. So $T\vec{x} = \vec{0} \implies \vec{x} = \vec{0}$.
Clearly $\vec{x} = 0 \implies T\vec{x} = \vec{0}$. So, $T\vec{x} = \vec{0} \Longleftrightarrow \vec{x} = \vec{0}$. Thus, T is non-singular.

(b) We use a proof by contradiction. Assume the LU factorization of T can be found even if row exchanges are not permitted. Then

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & t_{1,n} \\ & & \cdots & t_{2,n-1} & t_{2,n} \\ & & \vdots & & \\ 0 & t_{n-1,2} & \cdots & t_{n-1,n-1} & t_{n-1,n} \\ t_{n,1} & t_{n,2} & \cdots & t_{n,n-1} & t_{n,n} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ l_{2,1} & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & & \\ l_{n-1,1} & & \cdots & 1 & \\ l_{n,1} & l_{n,2} & \cdots & l_{n,n-1} & 1 \end{bmatrix} \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n-1} & u \\ 0 & u_{2,1} & & & \\ & & \ddots & & \\ & & & u_{n-1,n-1} & u_n \\ 0 & & & 0 & u \end{bmatrix}$$

By performing the matrix multiplication on the right hand side, and computing the elements, we get

$$u_{1,1} = 0 \text{ and } t_{1,n} = l_{n,1}u_{1,1},$$

which implies $t_{1,n} = 0$. But $t_{1,n}$ lies on the minor diagonal and is given to be non-zero. Hence, We derived a contradiction to our assumption.

(c) Let P be the matrix that has ones on its minor diagonal and zeros everywhere else. Premultiplying P to T replaces row i with row $n - i + 1$ and hence produces an upper triangular matrix $PT = U$, which is the LU decomposition of T with its rows swapped and $L = I_{n\times n}$ is the identity matrix.

Exercise 2.5

MATLAB Exercise. We will solve 1D heat equation, discussed in section **??**, numerically for different boundary conditions. Set the source term $f(x) = 0$ with boundary conditions $T(0) = 0$ and $T(1) = 2$.

(a) Give the exact solution to this differential equation.

(b) Discretize the 1D equation using the second order discretization scheme with $h = 1/N$. Write an expression for each equation in the system.

(c) Formulate the resulting matrix-vector equation $A\vec{T} = \vec{c}$ for $N = 5$, and use Gaussian Elimination (by hand) to check that the matrix A is nonsingular.

(d) Now set the source $f(x) = -10\sin(3\pi x/2)$. Keep the boundary conditions as before.

  (i) Verify that $T(x) = (2 + 40/(9\pi^2))x + 40/(9\pi^2)\sin(3\pi x/2)$ is the exact solution that solves the differential equation with this source and boundary conditions.

  (ii) For this source function, solve the system $A\vec{T} = \vec{c}$ using MATLAB for $N = 5$, $N = 10$, $N = 25$ and $N = 200$. Plot all computed solutions together in one figure with the exact solution. Clearly label your graph. Note that the curves may overlap. Use a legend. *Note: We want you to arrive at a discrete solution for N=200 points here not because we need it for accuracy, but because it will make it hard to construct the matrix in MATLAB element-by-element, which we do not want you to do. Use some of MATLAB's built in functions for creating these matrices. Use the diag command to construct tridiagonal matrices (consult help diag for usage). If you are ambitious, construct A as a sparse matrix using sp-diags, since the sparse representation is more efficient for matrices with few nonzero elements.*

Answer for Exercise 2.5

Exercise 2.6

In the first chapter, we discretized the 1-dimensional heat equation with Dirichlet boundary conditions:

$$\frac{d^2T}{dx^2} = 0, 0 \leq x \leq 1$$
$$T(0) = 0, T(1) = 2$$

The discretization leads to the matrix-vector equation $A\vec{T} = \vec{b}$, with

$$A = \begin{pmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & -2 & 1 \\ 0 & \dots & 0 & 1 & -2 \end{pmatrix}, b = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ -2 \end{pmatrix} \tag{2.1}$$

Here A is an $(N-1) \times (N-1)$ matrix.

(a) Find the LU factorization of A for $N = 10$ using `Matlab`. Is `Matlab` using any pivoting to find the LU decomposition? Find the inverse of A also. As you can see the inverse of A is a dense matrix.
Note: The attractive sparsity of A has been lost when computing its inverse, but L and U are sparse. Generally speaking, banded matrices have L and U with similar band structure. Naturally we then prefer to use the L and U matrices to compute the solution, and not the inverse. Finding L and U matrices with least "fill-in" ("fill-in" refers to nonzeros appearing at locations in the matrix where A has a zero element) is an active research area, and generally involves sophisticated matrix re-ordering algorithms.

(b) Compute the determinants of L and U for $N = 1000$ using `Matlab`'s determinant command. Why does the fact that they are both nonzero imply that A is non-singular? How could you have computed these determinants really quickly yourself without `Matlab`'s determinant command?

Answer for Exercise 2.6

(a) The solution can be found using Matlab command `lu`. To check whether Matlab is using any pivoting we can see what permutation matrix P is returned by the `lu` command. We can see it if we use `[L, U, P] = lu(A)` as in the code given below.

For the above matrix, with $N = 10$ we get L, U, P as follows:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1/2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2/3 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -3/4 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -4/5 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -5/6 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -6/7 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -7/8 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -8/9 & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -3/2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -4/3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5/4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -6/5 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -7/6 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -8/7 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -9/8 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -10/9 \end{pmatrix}$$

and P is the identity matrix. Hence we know that `Matlab` is not using any pivoting in this case. While L and U maintain the sparsity of A, the A^{-1} does not.

$$A^{-1} = \begin{pmatrix} 0.9 & 0.8 & 0.7 & 0.6 & 0.5 & 0.4 & 0.3 & 0.2 & 0.1 \\ 0.8 & 1.6 & 1.4 & 1.2 & 1.0 & 0.8 & 0.6 & 0.4 & 0.2 \\ 0.7 & 1.4 & 2.1 & 1.8 & 1.5 & 1.2 & 0.9 & 0.6 & 0.3 \\ 0.6 & 1.2 & 1.8 & 2.4 & 2.0 & 1.6 & 1.2 & 0.8 & 0.4 \\ 0.5 & 1.0 & 1.5 & 2.0 & 2.5 & 2.0 & 1.5 & 1.0 & 0.5 \\ 0.4 & 0.8 & 1.2 & 1.6 & 2.0 & 2.4 & 1.8 & 1.2 & 0.6 \\ 0.3 & 0.6 & 0.9 & 1.2 & 1.5 & 1.8 & 2.1 & 1.4 & 0.7 \\ 0.2 & 0.4 & 0.6 & 0.8 & 1.0 & 1.2 & 1.4 & 1.6 & 0.8 \\ 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 & 0.9 \end{pmatrix}$$

Matlab code for solving part a)

```
% Assignment 4 - problem a , finding LU and inverse of A of a
% finite difference matrix to solve T_xx = f(x), 0<=x<=1
% with T( x=0) = 0 , T( x=1) = 2
```

```
clear all
N = 10;

% Find the points of discretization
h = 1/N;
% Interval size of discretization
x = 0:h:1; % x = [x_0, x_1, ..., x_N]
x = x(2:end-1); % x = [x_1, x_2, ..., x_{N-1}]
% Construct the tri-diagonal matrix A
A = diag(ones(N-2,1),1) - 2*eye(N-1) + diag(ones(N-2,1),-1);

% Instead of the line above, use the line below to take advantage of the sparsity of
% We can construct A as a sparse matrix directly (this will save us some memory spac
% and speed up computation)
% A = spdiags(ones(N-1,2), [-1;1], N-1, N-1) - 2*speye(N-1);
% Note: to answer this question, we don't need to setup the RHS, b

[L, U, P ] = lu(A);
sym(L)
sym(U)
A_inv = inv(A)
```

(b) Determinants can be computed using Matlab command `det`. The determinant of L is 1 and that of U is -1000. The same answer could have been obtained by hand by simply multiplying the diagonal elements of the triangular matrices.

The fact that the determinants of L and U are both nonzero implies that A is nonsingular since,

$|A| = |L||U| = 0$

Exercise 2.7

In this chapter we introduced the LU decomposition of A, where L is unit-lower triangular, in that it has ones along the diagonal, and U is upper triangular. However, in the case

of symmetric matrices, such as the discretization matrix, it is possible to decompose A as LDL^T, where L is still unit-lower triangular and D is diagonal. This decomposition clearly shows off the symmetric nature of A.

(a) Find the LDL^T decomposition for the matrix given in Exercise 2.6. Show that L is bidiagonal. How do D and L relate to the matrix U in the LU decomposition of A?
Hint: Think about how D and L^T relate to U.
Note: Computing LDL^T this way does not work out for any symmetric matrix, it only happens to work for this matrix in particular.

*(b) (i) To solve $A\vec{x} = \vec{b}$ we can exploit this new decomposition. We get $LDL^T\vec{x} = \vec{b}$ which we can now break into three parts: Solve $L\vec{y} = \vec{b}$ using forward substitution, now solve $D\vec{z} = \vec{y}$, and then solve $L^T\vec{x} = \vec{z}$ using back substitution. Write a `Matlab` code that does exactly this for arbitrary N for the A in Exercise 2.6.

(ii) Solve a system of the same form as Exercise 2.6 for A of size 10 and of size 1000 with $\vec{b}$ having all zeros except 2 as the last entry in both cases, and verify the correctness of your solution using `Matlab`'s `A\b` operator and the `norm` command.

Answer for Exercise 2.7

Exercise 2.8

For tridiagonal matrices, such as those given in Exercise 2.6, a fast algorithm was developed called the Thomas algorithm. Go here for a description of this algorithm:

http://en.wikipedia.org/wiki/Tridiagonal matrix algorithm

(a) Implement the algorithm in `Matlab` for the system $A\vec{T} = \vec{b}$ with comments to show your understanding of the algorithm. Use your implementation to compute solutions for the system given in Exercise 2.6 for $N = 400$ and $N = 1000$. Compare the 2-norm of the solutions found using Thomas and backslash (`A\b`) in `Matlab` for both $N = 400$ and $N = 1000$.

*(b) A $n \times n$ matrix A is said to be strictly diagonally dominant if

$$|a_{ii}| > \sum_{j=1}^{i-1} |a_{ij}| + \sum_{j=i+1}^{n} |a_{ij}| \quad \text{for all } i \text{ between 1 and } n$$

where a_{ij} denotes the entry in the i^{th} row and j^{th} column. Show that strictly diagonally dominance guarantees the Thomas algorithm to work.

*(c) The matrix in exercise (2.6) is not strictly diagonally dominant. Nevertheless, the Thomas algorithm works, which we hope you found out in part a of this problem. How can you explain that it does?

Answer for Exercise 2.8

(a) Here is `Matlab` code for solving part a

```
%% Problem 2 Thomas Algorithm
%% Finite difference method to solve T_xx = f(x), 0<=x<=1
%% with T(x=0) = 1, T_x(x=1) = 0
clear all; clc

N = 400;
h = 1/(N);
N1 = N-1;
x_a = (0:h:1); %% Grid points
x_u = x_a(2:N); %% solution unknown at these grid points excluding the point xN
%% Assign b = [f1 f2.... fN-1], where f(x) = sin(1/2pi*x)
d = h^2*sin(0.5*pi*x_u);

%% use the fact TN = 1, in the eqn ( T(N-2) - 2T(N-1) + TN ) = (h^2)f(N-1)
%% The last eqn becomes (T(N-2)-2T(N-1))=(h^2)f(N-1)-TN = (h^2)f(N-1)- 1
d(N1) = d(N1) - 1;

% %% Construction of matrix A
% %% Regular Version
% %% main diagonal [-2 -2 .... -2 -2]
% md = [-2*ones(N1,1)];
% %% upper diagonal [1 1.......1]
% ud = [ones(N1-1,1)];
% %% lower diagonal [1 1.......1]
% ld = ud;
% A = diag(md) + diag(ud,1) + diag(ld,-1);

%% Sparse Matrix Version
MD = sparse(1:N1,1:N1,-2,N1,N1);
UD = sparse(1:N1-1,2:N1,1,N1,N1);
LD = sparse(2:N1,1:N1-1,1,N1,N1);
```

```
A = MD + UD + LD;

%% Computation time can be measured using tic-toc, cputime or etime
% t = clock; T_lu = A\d; etime(clock,t)
% t= cputime; T_lu = A\d; cputime-t
Time for \
tic; T_lu = A\d; t= toc

%% Thomas
Time for Thomas
Tt = 0*T_lu;
b = diag((MD));
tic;
for i = 2:N1
m = 1/b(i-1);
b(i) = b(i)-m;
d(i) = d(i) - m*d(i-1);
end
Tt(N1) = d(N1)/b(N1);
for k = N1-1:-1:1
Tt(k) = (d(k) - Tt(k+1))/b(k);
end
t = toc

%% compare the solutions
hold on;
h1 = plot(x_u, T_lu,go);
h2 = plot(x_u, Tt,r.);
title(Comparision of solutions by \\ and Thomas algorithms)
legend([h1 h2] , \\, Thomas)
xlabel(x)
ylabel(T)
```

For $N = 400$, the 2-norm difference is 1.6232×10^{-14}; for $N = 1000$, the 2-norm difference is 2.2452×10^{-14}.
Notice that the results of Thomas algorithm can be improved if the codes are modified to have fewer numerical computations. If the second approach provided in the solution is used, I get the 2-norm differences are 0 for both $N = 400$ and $N = 1000$.

*(b) This is an interesting question that requires a careful proof. What we really need to show is that if the matrix diagonally dominant, the Thomas algorithm can never encounter a zero pivot, so it does not have to check for zero pivots. Using the notations

in the link, we want to show

$$b_i - c'_{i-1}a_i \neq 0$$

Let $\beta_i = b_i - c'_{i-1}a_i$, then from $c'_i = \frac{c_i}{\beta_i}$, we have

$$\beta_i = b_i - c'_{i-1}a_i = b_i - a_i\frac{c_{i-1}}{\beta_{i-1}}$$

Taking absolute values, we have

$$\left|a_i\frac{c_{i-1}}{\beta_{i-1}}\right| = |a_i|\left|\frac{c_{i-1}}{\beta_{i-1}}\right| = |b_i| > |a_i| + |c_i|$$

So if we can show $\left|\frac{c_{i-1}}{\beta_{i-1}}\right| < 1$, then we would have,

$$|a_i| > |a_i|\left|\frac{c_{i-1}}{\beta_{i-1}}\right| = |b_i| > |a_i| + |c_i|$$

which is a contradiction.
So in order to prove that $\beta_i \neq 0$ for all i, we need to show that $\left|\frac{c_i}{\beta_i}\right| < 1$ for all i. We do this by induction. For $i = 1$, we have

$$\left|\frac{c_1}{\beta_1}\right| = \left|\frac{c_1}{b_1}\right| < 1$$

by strict diagonal dominance. Now suppose $\left|\frac{c_i}{\beta_i}\right| < 1$, then

$$|\beta_i| = \left|b_i - a_i\frac{c_{i-1}}{\beta_{i-1}}\right| \geq |b_i| - \left|a_i\frac{c_{i-1}}{\beta_{i-1}}\right| > |b_i| - |a_i| > |c_i|$$

or $\left|\frac{c_i}{\beta_i}\right| < 1$. Note that the last two inequalities come from the induction hypothesis and diagonal dominance, respectively.

*(c) Our matrix is not diagonally dominant, but it is almost so! Lets see if this so-called "weak diagonal dominance" is enough. For our example, we have $|b_i| \geq |a_i| + |c_i|$, $a_i = c_i$ for $i = 2, \ldots, N-1$, and $|b_1| > |c_1|$ and $|b_n| > |a_n|$.
Once again, we'll do a proof by contradiction. Suppose $\beta_i = 0$. Then

$$0 = \beta_i = b_i - a_i\frac{c_{i-1}}{\beta_{i-1}} \implies b_i = a_i\frac{c_{i-1}}{\beta_{i-1}}$$

Taking absolute values, we have

$$\left|a_i\frac{c_{i-1}}{\beta_{i-1}}\right| = |a_i|\left|\frac{c_{i-1}}{\beta_{i-1}}\right| = |b_i| \geq |a_i| + |c_i|$$

So if we can show once again that $\left|\frac{c_{i-1}}{\beta_{i-1}}\right| < 1$, then we will have the same contradiction as before. For $i = 1$, we have

$$\left|\frac{c_1}{\beta_1}\right| = \left|\frac{c_1}{b_1}\right| < 1$$

Now suppose $\left|\frac{c_i}{\beta_i}\right| < 1$, then

$$|\beta_i| = \left|b_i - a_i\frac{c_{i-1}}{\beta_{i-1}}\right| \geq |b_i| - \left|a_i\frac{c_{i-1}}{\beta_{i-1}}\right| > |b_i| - |a_i| > |c_i|$$

or $\left|\frac{c_i}{\beta_i}\right| < 1$. Note that only the very last inequality changes due to only weak diagonal dominance but it doesn't affect the strict inequality from the induction hypothesis. The initial struct dominance in the first row gives us the needed condition the rest of the way.

*** Exercise 2.9**

The column vectors $\vec{u}$ and $\vec{v}$ each have n elements. We form the $n \times n$ matrix A as $A = \vec{u}\vec{v}^T$. We now perform Gaussian elimination on A to find the matrix U. Show that U has either $n - 1$ or n zero rows.

Answer for Exercise 2.9

Exercise 2.10

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement false, it is sufficient to give on counter example. To prove a statement true, provide a general proof.

If the matrix A is symmetric and tridiagonal, pivoting is never required in Gaussian elimination.

Answer for Exercise 2.10

False. A counterexample is

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

Pivoting is required in the second step.

Chapter 3

Exercises & Solutions

Exercise 3.1

Assume that the vector $\vec{b}$ in the system $A\vec{x} = \vec{b}$ is perturbed by $\overrightarrow{\delta b}$. The solution will now change to $\vec{x} + \overrightarrow{\delta x}$. In the note, we derived an upper bound for the relative error of $\vec{x}$ in terms of the condition number of A and the relative error in $\vec{b}$:

$$\frac{\|\overrightarrow{\delta x}\|_2}{\vec{x}} \leq \|A\|_F \|A^{-1}\|_F \frac{\|\overrightarrow{\delta b}\|_2}{\vec{b}}.$$

Derive a **lower** bound for the relative error of $\vec{x}$.

Answer for Exercise 3.1

Exercise 3.2

In the notes, we discussed how a perturbation in the matrix A can lead to perturbations in the solution $\vec{x}$, and separately, how a perturbation in the vector $\vec{b}$ can perturb $\vec{x}$. How might the solution $\vec{x}$ to the linear system $A\vec{x} = \vec{b}$ change when we perturb both A and $\vec{b}$?

Answer for Exercise 3.2

Consider the perturbed linear system

$$(A + \partial A)\vec{y} = \vec{b} + \partial\vec{b}.$$

The perturbed solution $\vec{y}$ can be expressed as $\vec{x} + \partial\vec{x}$ for some $\partial\vec{x}$, so we have

$$(A + \partial A)(\vec{x} + \partial\vec{x}) = \vec{b} + \partial\vec{b}.$$

Assuming $A + \partial A$ is invertible, we solve for $\partial\vec{x}$ to obtain

$$\partial\vec{x} = (A + \partial A)^{-1}(\vec{b} + \partial\vec{b} - (A + \partial A)\vec{x}) \tag{3.1}$$

$$= (A + \partial A)^{-1}(\vec{b} + \partial\vec{b}) - \vec{x}. \tag{3.2}$$

If ∂A is small, then $(A + \partial A)^{-1} \approx A^{-1} - A^{-1}\partial A A^{-1}$. We substitute this expression into (3.2)

$$\begin{aligned}\partial\vec{x} &\approx (A^{-1} - A^{-1}\partial A A^{-1})(\vec{b} + \partial\vec{b}) - \vec{x} \\ &= A^{-1}\partial\vec{b} - A^{-1}\partial A A^{-1}(\vec{b} + \partial\vec{b})\end{aligned}$$

and drop the second-order term $A^{-1}\partial A A^{-1}\partial\vec{b}$, because for small perturbations this will be much smaller than the first-order terms. We get

$$\begin{aligned}\partial\vec{x} &\approx A^{-1}\partial\vec{b} - A^{-1}\partial A A^{-1}\vec{b} \\ &= A^{-1}(\partial\vec{b} - \partial A\vec{x}).\end{aligned}$$

We take norms to obtain

$$\|\partial\vec{x}\| \leq \|A^{-1}\|(\|\partial\vec{b}\| + \|\partial A\|\|\vec{x}\|).$$

We divide both sides by $\|\vec{x}\|$

$$\frac{\|\partial\vec{x}\|}{\|vecx\|} \leq \|A^{-1}\|\left(\frac{\|\partial\vec{b}\|}{\|\vec{x}\|} + \|partialA\|\right)$$

and use the fact that $\frac{1}{\|\vec{x}\|} \leq \frac{\|A\|}{\|\vec{b}\|}$

$$\begin{aligned}\frac{\|\partial\vec{x}\|}{\|\vec{x}\|} &\leq \|A^{-1}\|\left(\|A\|\frac{\|\partial\vec{b}\|}{\|\vec{b}\|} + \|\partial A\|\right) \\ &= \kappa(A)\left(\frac{\|\partial\vec{b}\|}{\|\vec{b}\|} + \frac{\|\partial A\|}{\|A\|}\right).\end{aligned}$$

Exercise 3.3

The system $A\vec{x} = \vec{b}$, with $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \vec{b} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}$ has the solution $\vec{x} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. The vector $\vec{b}$ is now perturbed. The new vector is equal to $\vec{b} + \partial\vec{b}$ with $\partial\vec{b} = \begin{bmatrix} 0.001 \\ 0 \end{bmatrix}$. Estimate the maximum perturbation we can expect in the solution $\vec{x}$, measured in the vector 2-norm. For matrix-norms, you may use the Frobenius norm.

Answer for Exercise 3.3

Exercise 3.4

In the previous chapter, we solved a heat equation using numerical discretization for the nonuniform source term given by $f(x) = -10\sin(3\pi x/2)$.

(a) Find the condition number of the matrix A using MATLAB for $N = 5$, $N = 10$, $N = 25$ and $N = 200$. Use the Frobenius norm to compute the condition number, that is, set $\kappa(A) = \|A\|_F \|A^{-1}\|_F$ (use the MATLAB function cond(A,'fro')).

(b) Plot the condition numbers and comment on any pattern that you might find.

(c) Perturb the matrix A by ∂A for $\partial A = 0.1A$. Solve the perturbed system of equations $(A + \partial A)\vec{T} = \vec{c}$ and plot the solution. Comment on what you see. Can you relate it to the condition number of A?

Answer for Exercise 3.4

We use the following MATLAB code to compute the condition number of the matrix A for $N = 5,\ N = 10,\ N = 25,$ and $N = 200$.

```
N = [5; 10; 25; 200]
cond_numbers = zeros(4, 1);
```

```
% alpha is a perturbation parameter
alpha = 0.1;
f = @(x) -10*sin(3*pi*x/2);

% Set colors for plotting solutions
c = ['r', 'b', 'k', 'g'];
for k=1:length(N)
  h = 1/(N(k));
  % Construct sparse matrix A
  A = spdiags(ones(N(k)-1,2), [-1;1], N(k)-1, N(k)-1) - 2*speye(N(k)-1);

  % Perturbing the matrix by fraction alpha
  A = A + alpha*A;
  % Find the discretization points
  x = linspace(0, 1, N(k)+1);
  xind = x(2:end-1);
  % Form the right hand side c
  c  = (h^2)*(f(xind));
  c(N(k)-1) = c(N(k)-1) - 2;

  % Solve the matrix-vector system, and plot the solution
  T = A\(c');
  Tsol = [0; T; 2];
  plot(x', Tsol, c(k));
  hold on

  % Compute the Frobenius norm condition number
  cond_numbers(k) = cond(A,'fro');
end

% Display condition numbers
cond_numbers
%{
   1.2931e+01
   7.6943e+01
   7.8663e+02
   1.4558e+05
%}

legend({'N=5', 'N=10', 'N=25', 'N=200'}, 'Location', 'SouthEast')
hold off
```

```
figure(3);
plot(cond_numbers);
```

In Figure 3.1, we see that the condition number increases with N. The rate of increase is greater than linear. This rate typically depends on the norm used to define the condition number. That the condition number increases with N seems intuitive: the larger the system, the more manipulations take place in solving a system and the easier it is for perturbations to accumulate. In other words, the system should become more sensitive to perturbations.

In figure 3.2 we show the solution to the perturbed system for different values of N. The solution does not change all that much here, even though the condition number seems large. A few things to keep in mind. First of all, the condition number defines an upper limit to possible perturbations: it represents a worst case scenario. The other thing to keep in mind is that what is a large condition number for a matrix of a given sense is hard to define. Here, even though the condition number for $N = 200$ seems pretty decent, it's actually not so bad for system of that size. If a condition number for a matrix of that size is 10^{20} though, we would certainly not like it! It is all relative, and we should always do a perturbation study (or a couple of them) to assess potential problems.

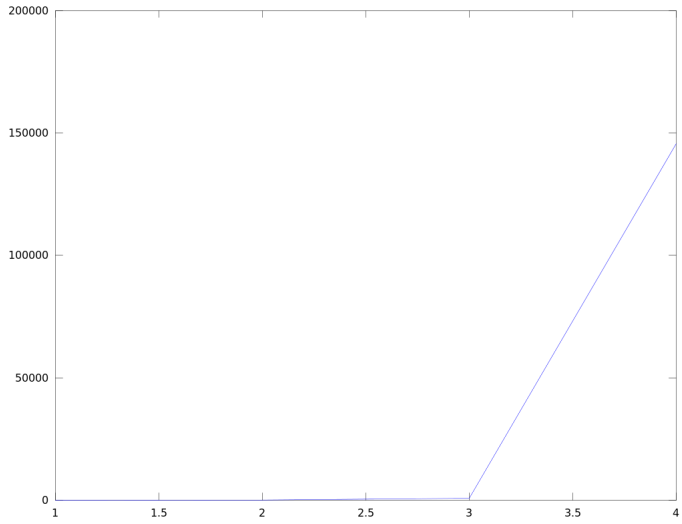

Figure 3.1: Variation of condition number with N.

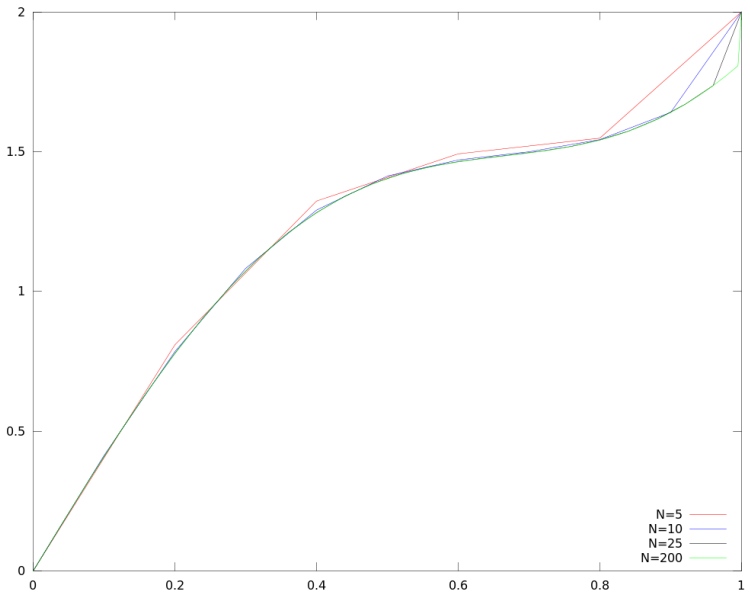

Figure 3.2: Solution to the **uniformly** perturbed system $(A+\alpha A)T = b$ for different values of N.

Exercise 3.5

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement is false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) $\kappa(\alpha A) = \alpha \cdot \kappa(A)$, where $\alpha > 0$

(b) Catastrophic cancellation can only occur during Gaussian Elimination of a matrix if that matrix is ill-conditioned.

Answer for Exercise 3.5

(a) False. Consider

$$(\alpha A)^{-1} = \frac{1}{\alpha} A^{-1}.$$

Taking $\alpha > 1$, we show that the statement above does not hold in general,

$$\kappa(\alpha A) = \|\alpha A\| \|(\alpha A)^{-1}\| = \alpha \|A\| \frac{1}{\alpha} \|A^{-1}\| = \|A\| \|A^{-1}\| = \kappa(A) \neq \alpha \ \kappa(A).$$

This is an important result: it means that the condition number of a matrix *does not change* when the matrix is scaled.

(b) False. This is discussed in the notes. Catastrophic cancellation can occur for well-conditioned matrices also and therefore we often use partial pivoting to avoid the use of tiny pivots.

Exercise 3.6

If a matrix is close to being singular, its condition number is very high. Give an example of a 4×4 matrix for which this is the case.

Answer for Exercise 3.6

A matrix is close to singular, so it is badly scaled or the rows/columns are close to being linearly dependent. Consider

$$A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1.001 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}.$$

$\kappa(A) = \|A^{-1}\|_F \|A\|_F = 4901.0$ which is quite high for such a small matrix.

Exercise 3.7

Show that any $n \times n$ permutation matrix P has a condition number (computed with the Frobenius norm) that is equal to n.

Answer for Exercise 3.7

Chapter 4

Exercises & Solutions

Exercise 4.1

Let L be the set of vectors $\vec{x}$ in $\mathbb{R}^4$ for which

$$x_1 + x_2 + x_3 = 0. \tag{4.1}$$

Find a basis for L. What is the dimension of L?

Answer for Exercise 4.1

Exercise 4.2

Find a basis for the subspace of $\mathbb{R}^4$ spanned by the vectors

$$\left\{ \begin{bmatrix} 0 \\ 2 \\ 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

Answer for Exercise 4.2

We create a matrix with rows equal to the given vectors, and get

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

This matrix is almost upper triangular (except the last row), and the last row is clearly a linear combination of the second and third rows. This means that the matrix has a row space with dimension 3 and the first three rows form a basis for this subspace of $\mathbb{R}^4$. If the matrix had not been already almost upper triangular, we would have performed Gaussian Elimination to find any dependencies between rows.

Exercise 4.3

(a) Find a basis for the column space and row space of matrix A given by

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

(b) Construct a matrix B such that the null space of B is identical to the row space of A.

Answer for Exercise 4.3

Exercise 4.4

We are given the LU decomposition of a matrix A as

$$A = LU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 \\ 3 & 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

(a) What is the rank $r(A)$ of matrix A?

(b) Find a basis for the null space of A.

(c) Find a basis for the column space of A.

(d) For $\vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ do solution(s) to $A\vec{x} = \vec{b}$ exist? Clearly motivate your answer. Note that it is not necessary to compute the solution(s).

Answer for Exercise 4.4

(a) First, recall that $r(A) = \dim(\mathcal{R}(\mathcal{A})) = \dim(\clubsuit(\mathcal{A}))$. Now, since we are already given the LU decomposition of A, we can tell the rank of matrix A by counting the number of non-zero rows in U, which corresponds to the number of linearly independent rows in A, i.e. $\dim(\clubsuit(A))$. There are 3 non-zero rows in U, hence $r(A) = 3$.

(b) There are a couple of ways to get the basis for the null space of A. (The method presented here was not the only way to get full credit.)
We recall that for $A \in \mathbb{R}^{m \times n}$, the null space $\mathcal{N}(A)$ is defined as,

$$\mathcal{N}(A) = \{\vec{x} \in \mathbb{R}^n | A\vec{x} = \vec{0}\}.$$

We know from the lectures that $\mathcal{N}(U) = \mathcal{N}(A)$. For completeness, this can be proven as follows (but not necessary to do this on the midterm solutions): First, we know that $\mathcal{N}(U) \subseteq \mathcal{N}(A)$ since if $\vec{x} \in \mathcal{N}(U)$, then $U\vec{x} = \vec{0}$ and $A\vec{x} = LU\vec{x} = L\vec{0} = \vec{0}$, so $\vec{x} \in \mathcal{N}(A)$. Then, we know that L is nonsingular as it is a triangular matrix with non-zero entries on the diagonal. In other words, L has full rank. Because $A = LU$, we then know that $\mathcal{N}(A) \subseteq \mathcal{N}(U)$. We can also convince ourselves of the latter in the following way: suppose that we have a $\vec{x} \in \mathcal{N}(A)$ and $\vec{x} \notin \mathcal{N}(U)$, then $U\vec{x} = \vec{y} \neq \vec{0}$ and we get a contradiction $A\vec{x} = LU\vec{x} = L\vec{y} \neq \vec{0}$, since L has full column rank.

Using that $\mathcal{N}(A) = \mathcal{N}(U)$, we can find a basis for the null space of A by working with

U directly. Suppose, $\vec{x} \in \mathcal{N}(U)$, $\vec{x} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \end{bmatrix}^T$. Then,

$$U\vec{x} = \begin{bmatrix} 1 & 2 & 0 & 1 & 2 & 1 \\ 0 & 0 & 2 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

We get the following set of equations

$$\begin{aligned} x_1 + 2x_2 + x_4 + 2x_5 + x_6 = 0 &\implies x_1 = -2x_2 - x_4 - 2x_5 - x_6, \\ 2x_3 + 2x_4 = 0 &\implies x_3 = -x_4, \\ x_6 = 0. & \end{aligned}$$

Choosing the free parameters, $x_2 = r,\ x_4 = s,\ x_5 = t$, we plug in to the equations above to get

$$\begin{aligned} x_1 &= -2r - s - 2t \\ x_2 &= r \\ x_3 &= -s \\ x_4 &= s \\ x_5 &= t \\ x_6 &= 0. \end{aligned}$$

Hence, every vector in the null space of A can written as,

$$\vec{x} = \begin{bmatrix} -2r - s - 2t \\ r \\ -s \\ s \\ t \\ 0 \end{bmatrix} = r \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + s \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = r\vec{a} + s\vec{b} + t\vec{c}.$$

Since, $r,\ s,$ and t are free parameters, the entire null space of A is comprised of linear combinations of the three vectors $\vec{a},\ \vec{b},\ \vec{c}$, i.e.

$$\mathcal{N}(A) = \left\{ \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \right\}.$$

[Note that depending on the choice of the free parameters, the resultant basis vectors could differ. However, we only need to make sure that the vectors are linearly independent and they indeed span the null space.]

You can easily, verify that the vectors $\vec{a}$, $\vec{b}$, $\vec{c}$ are in the null space of U by multiplying each one of them by U. Also, by observation it is easy to see that these three vectors are linearly independent. Finally, the dimension of the null space is $\dim(\mathcal{N}(A)) = n - r(A) = 6 - 3 = 3$ by the subtle theorem. Hence, we indeed found a basis for the null space of A.

(c) The idea here is to multiply L and U to get A, and then the columns of A corresponding to the pivot positions of U will form the column space of A. Therefore

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

The column space can therefore be taken as columns $\{1, 3, 6\}$ or $\{1, 4, 6\}$ of A, and column 3 may be scaled by $\frac{1}{2}$. Your answer here affects the next part so this was taken into account. (Note - many students took the columns from U and not A and this is incorrect.)

(d) There are a few ways to do this. The easiest way is if you notice that

$$\vec{b} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 0 \\ 2 \\ 2 \\ 4 \end{bmatrix}$$

which means that it is indeed spanned by the columns of A. A more systematic way to arrive at this is by actually solving for the coefficients of the linear combinations of the columns in the column space of A which will give $\vec{b}$

$$\alpha_1 \begin{bmatrix} 1 \\ 2 \\ 2 \\ 3 \end{bmatrix} + \alpha_2 \begin{bmatrix} 0 \\ 2 \\ 2 \\ 4 \end{bmatrix} + \alpha_3 \begin{bmatrix} 1 \\ 2 \\ 3 \\ 7 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

There is no need to fully solve this system, only need to row reduce the augmented

matrix and show that there is no inconsistent row.

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 2 & 2 & 2 & 1 \\ 2 & 2 & 3 & 1 \\ 3 & 4 & 7 & 1 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & -1 \\ 0 & 2 & 1 & -1 \\ 0 & 4 & 4 & -2 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 4 & 0 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 2 & 0 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Yet another way to do this is by looking at

$$\begin{aligned} Ax &= b \\ LUx &= b \\ Ly &= b \\ Ux &= y \end{aligned}$$

You can solve for y by performing forward substitution or computing the inverse of L and multiplying this to b. You should get

$$y = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

When we've given this assignment to students, several made mistakes in computing the inverse of L. The most common mistake was to take L and simply multiply all off-diagonal elements by -1. Careful though! This could be done for the elementary building blocks of L (the C_i's introduced in the notes) but it can *not* by done on L itself. You can easily convince yourself by multiplying the matrix you would get in this case with L. You will not get the identity.

At this point you can explicitly solve $Ux = y$ to get the solution or simply observe that wherever U has all zero rows (in this case only the last row) y also has zero component (the last two components). Therefore the system has at least one solution, in fact infinitely many solutions.

Exercise 4.5

Consider the matrix

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

(a) Find a basis for the null space of A. What is the rank of A?

(b) Find the solution(s), if any can be found, to the equation $A\vec{x} = \vec{b}$ with $\vec{b} = \begin{bmatrix} 0 \\ -1/2 \\ -1 \end{bmatrix}$.

(c) Prove that A^k is singular for all integer $k > 0$.

Answer for Exercise 4.5

Exercise 4.6

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement is false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) If the vectors $\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m$ in $\mathbb{R}^n$ span a subspace S in $\mathbb{R}^n$, the dimension of S is m.

(b) If matrices A and B have the same row space, the same column space and the same nullspace, then A must be equal to B.

*(c) If for the $m \times n$ matrix A, the equation $A\vec{x} = \vec{b}$ always has at least one solution for every choice of $\vec{b}$, then the only solution to $A^T\vec{y} = \vec{0}$ is $\vec{y} = \vec{0}$.

*(d) If the m vectors $\vec{x}_i, i = 1, \ldots, m$ are orthogonal, they are also independent. Note here that $\vec{x}_i \in \mathbb{R}^n$ and $n > m$.

Answer for Exercise 4.6

(a) False.

Take for example $\vec{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\vec{x}_2 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$. Then $m = 2$ but the span of $\{\vec{x}_1, \vec{x}_2\}$ only has dimension 1.

The crux here is to recognize that is just says the vectors *span* then subspace. It does not say the vectors form a basis. The vectors form a spanning set and this spanning set can be redundant.

(b) False.

A can be a multiple of B. Remember that the row and column space are spanned by the rows and columns, respectively. If the row spaces are the same, for example, it does not mean the rows are identical, it just means that both sets of rows span the same space.

*(c) True.

Since $A\vec{x} = \vec{b}$ has at least one solution for any $\vec{b}$, the columnspace must be all of $\mathbb{R}^m$. This implies that $m \leq n$, and that $r(A) = m$. Now, for an $m \times n$ matrix A, the dimension of the nullspace is $n - r$, where r is the rank of A. Remember that the rank of A is the same as the rank of A^T. For the $n \times m$ matrix A^T, the dimension of the nullspace is therefore $m - r = m - m = 0$. This implies the only solution of $A^T\vec{y} = \vec{0}$ is $\vec{y} = \vec{0}$ as the nullspace is just the zero vector itself (the dimension of the nullspace is 0).

*(d) True.

If $c_1, \ldots, c_m$ are scalars satisfying $c_1\vec{x}_1 + \ldots + c_m\vec{x}_m = \vec{0}$, then we may write

$$(c_1\vec{x}_1 + \ldots + c_m\vec{x}_m)^T\vec{x}_i = \vec{0}^T\vec{x}_i = \vec{0},$$

or, equivalently,

$$c_1(\vec{x}_1^T\vec{x}_i) + \ldots + c_i(\vec{x}_i^T\vec{x}_i) + \ldots + c_m(\vec{x}_m^T\vec{x}_i) = \vec{0}.$$

Since the vectors $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ are orthogonal, all the inner products in the previous line vanish with the exception of $c_i(\vec{x}_i^T\vec{x}_i)$. Thus, we have

$$c_i(\vec{x}_i^T\vec{x}_i) = \vec{0}$$

Now, since $\vec{x}_i^T\vec{x}_i \neq 0$ (we implicitly know that the $\vec{x}_i$ are nonzero for the conclusion to hold), we must have $c_i = 0$. Since this argument holds for all $i = 1, \ldots, m$, we deduce that the set of vectors $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_m\}$ is linearly independent.

Note that we here implicitly assumed that none of the vectors are $\vec{0}$.

Exercise 4.7

A projector P is an $n \times n$ matrix that satisfies $P = P^2$. Show that $I - P$, with I the identity matrix, is also a projector, and that the range of $I - P$ exactly equals the null space of P.

Answer for Exercise 4.7

Exercise 4.8

An $n \times n$ matrix A has a property that the elements in each of its rows sum to 1. Let P be any $n \times n$ permutation matrix. Prove that $(P - A)$ is singular.

Answer for Exercise 4.8

$(P - A)$ is a matrix such that the elements in each of its rows sum to 0. This is because the permutation matrix P has exactly one entry 1 in each row and 0's elsewhere.

Consider that $(P-A)$ is singular if any only if $\mathcal{N}(P-A)$ contains a nonzero vector. Recall that $\mathcal{N}(P - A) = \left\{\vec{x} \in \mathbb{R}^n | (P - A)\vec{x} = \vec{0}\right\}$. Therefore, in order to show that $\mathcal{N}(P - A)$ contains at least one nonzero vector, we have to find some vector $\vec{v} \neq \vec{0}$ such that $(P-A)\vec{v} = \vec{0}$.

We explore the fact that the elements in each of the rows of $(P - A)$ sum to 0. Let c_{ij} be the (i, j)-th entry of matrix $(P - A)$. Then we have

$$\vec{0} = \begin{bmatrix} \sum_{j=1}^{n} c_{1j} \\ \sum_{j=1}^{n} c_{2j} \\ \vdots \\ \sum_{j=1}^{n} c_{nj} \end{bmatrix} = (P - A) \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

Now let $\vec{v} = [1, 1, \ldots, 1]^T$ be a vector with 1s in all its entries. Then $(P - A)\vec{v} = \vec{0}$, which means that $\vec{v}$ is in the null space of $(P - A)$. Since $\vec{v} \neq \vec{0}$, we found a nonzero vector in $\mathcal{N}(P - A)$, that is, $\dim \mathcal{N}(P - A) > 0$. Thus, $(P - A)$ is singular.

Exercise 4.9

(a) The nonzero column vectors $\vec{u}$ and $\vec{v}$ have n elements. An $n \times n$ matrix A is given by $A = \vec{u}\vec{v}^T$ (**Note:** this is different from the innerproduct (also sometimes known as the dot product), which we would write as $\vec{v}^T\vec{u}$). Show that the rank of A is 1.

(b) Show that the converse is true also. That is, if the rank of a matrix A is 1, then we can find two vectors $\vec{u}$ and $\vec{v}$, such that $A = \vec{u}\vec{v}^T$.

Answer for Exercise 4.9

Exercise 4.10

Prove that all bases of a given vector subspace must have the same number of vectors.

Answer for Exercise 4.10

This proof can be found in the book by Strang.

Exercise 4.11

Let U and V be subspaces of $\mathbb{R}^n$.

(a) The *intersection* of U and V is the set

$$U \cap V := \{x \in \mathbb{R}^n \mid x \in U \text{ and } x \in V\}.$$

Is $U \cap V$ a subspace for any U and V?

(b) The *union* of U and V is the set

$$U \cup V := \{x \in \mathbb{R}^n \mid x \in U \text{ or } x \in V\}.$$

Is $U \cup V$ a subspace for any U and V?

Answer for Exercise 4.11

Exercise 4.12

Let A be an $m \times n$ matrix with rank $r \leq \min\{m, n\}$. Depending on $m,\ n$ and r, a system $A\vec{x} = \vec{b}$ can have none, one, or infinitely many solutions.
For what choices of $m,\ n$ and r do each of the following cases hold? If no such $m,\ n$ and r can be found explain why not.

(a) $A\vec{x} = \vec{b}$ has no solutions, regardless of $\vec{b}$

(b) $A\vec{x} = \vec{b}$ has exactly 1 solution for any $\vec{b}$

(c) $A\vec{x} = \vec{b}$ has infinitely many solutions for any $\vec{b}$

(**Hint**: think of what conditions the column vectors and column space of A should satisfy.)

Answer for Exercise 4.12

(a) $A\vec{x} = \vec{b}$ has no solutions, regardless of $\vec{b}$
Answer: Impossible. Since for $\vec{b} = 0$ we always have a solution $\vec{x} = 0$.

(b) $A\vec{x} = \vec{b}$ has exactly 1 solution for any $\vec{b}$
Answer: For any $\vec{b} \in \mathbb{R}^m$ there is exactly one solution to $A\vec{x} = \vec{b}$, this happens if and only if the columns of A form a basis of space $\mathbb{R}^m$. That is, if and only if the column space of A is the full space $\mathbb{R}^m$ and the columns are independent. That is, if and only if $m = n = r$. (Hence if and only if A is square and nonsingular.)

(c) $A\vec{x} = \vec{b}$ has infinitely many solutions for any $\vec{b}$
Answer: For any $\vec{b} \in \mathbb{R}^m$ there are infinite solutions to $A\vec{x} = \vec{b}$. That is, for any $\vec{b} \in \mathbb{R}^m$ we can find a solution to $A\vec{x} = \vec{b}$ and add to this solution any element of the nullspace of A. Therefore we need the nullspace of A to have an infinite number of vectors in it. In other words, the nullspace dimension should be at least 1, i.e. $n - r \geq 1$. Since there is at least one solution for any $\vec{b}$, the columns of A must span the full space $\mathbb{R}^m$. Note that the columns of A live in $\mathbb{R}^m$. So we need $r = m$.
On the other hand, if the column space of A is the full space ($r = m$) and the columns are dependent ($r < n$), then for any $\vec{b} \in \mathbb{R}^m$ there are infinite solutions to $A\vec{x} = \vec{b}$. So the converse is also true.
Therefore a necessary and sufficient condition is that the column space of A is the full space ($m = r$) and the columns are dependent ($r < n$). That is, the statement is true if and only if $m = r < n$.

Remark
Note that the equation $A\vec{x} = \vec{b}$ could have

- exactly 1 solution for any $\vec{b}$ (part (b))
- infinity many solutions for any $\vec{b}$ (part (c))
- no solution for some $\vec{b}$ and has 1 solution for some $\vec{b'}$
- no solution for some $\vec{b}$ and has infinitely many solutions for some $\vec{b'}$

Exercise: Think about the conditions that make the 3^{rd} and 4^{th} cases true.

Exercise 4.13

Consider a matrix product AB, where A is $m \times n$ and B is $n \times p$. Show that the column space of AB is contained in the column space of A. Give an example of matrices A, B such that those two spaces are not identical.
Definition. A vector space U is *contained* in another vector space V (denoted as $U \subseteq V$) if every vector $\vec{u} \in U$ ($\vec{u}$ in vector space U) is also in V.
Definition. We say that two vector spaces are *identical* (equal) if $U \subseteq V$ **and** $V \subseteq U$.
(e.g. V is identical to itself since $V \subseteq V$ and $V \subseteq V$.)

Answer for Exercise 4.13

Exercise 4.14

Let V and W be 3 dimensional subspaces of $\mathbb{R}^5$. Show that V and W must have at least one nonzero vector in common.

Answer for Exercise 4.14

Solution I

Let bases of V and W be $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$, respectively. Then $\text{span}\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is a subspace of $\mathbb{R}^5$. Since the dimension of $\mathbb{R}^5$ is 5, the six vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1, \vec{w}_2, \vec{w}_3$ cannot be independent. Thus, there exist constants $\alpha_i, \beta_i,\ i = 1, 2, 3$ (these constants cannot all be zeros) such that

$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \alpha_3\vec{v}_3 + \beta_1\vec{w}_1 + \beta_2\vec{w}_2 + \beta_3\vec{w}_3 = \vec{0} \in \mathbb{R}^5.$$

Rearranging the terms, we can write

$$\vec{u} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \alpha_3\vec{v}_3 = -(\beta_1\vec{w}_1 + \beta_2\vec{w}_2 + \beta_3\vec{w}_3) \in \mathbb{R}^5.$$

If $\vec{u}$ is zero, we must have $\alpha_i = 0,\ i = 1, 2, 3$, since $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is independent, and $\beta_i = 0,\ i = 1, 2, 3$, since $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is independent. However, by the argument above, we cannot have all the constants to be zeros, thus $\vec{u} \neq 0$.

We have $\vec{u} \in V$, since $\vec{u} = \alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \alpha_3\vec{v}_3$. Similarly $\vec{u} \in W$, since $\vec{u} = -(\beta_1\vec{w}_1 + \beta_2\vec{w}_2 + \beta_3\vec{w}_3)$. Thus, we have a nonzero vector $\vec{u} \in \mathbb{R}^5$, which is both in V and in W.

Solution II

We use proof by contradiction. Let bases of V and W be $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$, respectively. We assume that V and W are disjoint. Then any linear combination of $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ cannot be expressed as a linear combination of $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$. Equivalently, the equation

$$c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + c_4\vec{w}_1 + c_5\vec{w}_2 + c_6\vec{w}_3 = 0$$

can only have a trivial solution $\vec{c} = [c_1, c_2, c_3, c_4, c_5, c_6]^T = \vec{0}$.

Let $A = [\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1, \vec{w}_2, \vec{w}_3]$. So A is a 5×6 matrix. The linear equation $A\vec{x} = \vec{0}$ is a homogeneous underdetermined system, so it must have a non-trivial solution. This contradiction proves that V and W cannot be disjoint.

Exercise 4.15

Show that the two sets of vectors $S_1 = \left\{ \begin{bmatrix} 3 \\ 1 \\ -2 \end{bmatrix}, \begin{bmatrix} -1 \\ 3 \\ -4 \end{bmatrix} \right\}$ and $S_2 = \left\{ \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} \right\}$ span the same subspace.

Answer for Exercise 4.15

Exercise 4.16

The 4×3 matrix A is given by $A = \begin{bmatrix} 1 & 2 & -3 \\ -1 & 1 & 1 \\ -1 & 0 & 2 \\ 3 & 8 & -10 \end{bmatrix}$, and the vector $\vec{d}$ by $\vec{d} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$.

(a) Find the bases for the row space, null space, and column space of A.

(b) Is it possible to find a vector $\vec{b}$ for which $A\vec{x} = \vec{b}$ has an infinite number of solutions? Explain. If it is possible, find such a $\vec{b}$.

(c) Show that $A\vec{x} = \vec{d}$, with $\vec{d}$ as given above, cannot be solved exactly. Then explain clearly how a solution may be found that minimizes $\|\vec{d} - A\vec{x}\|_2$, and give the general expression of this solution in terms of A and $\vec{d}$. **Note:** you need not calculate this solution exactly.

Answer for Exercise 4.16

(a) We apply Gaussian Elimination. So, there are 3 independent rows (and so 3 independent columns). The first three rows of A form a basis for the row space, all columns of A form a basis for the column space. Since the rank of A is 3, the dimension of the

nullspace is zero and so the nullspace only contains the zero vector. Thus, the bases of the row space, column space, and nullspace of A are

$$\left\{ \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \right\}, \left\{ \begin{bmatrix} 1 \\ -1 \\ -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 0 \\ 8 \end{bmatrix}, \begin{bmatrix} -3 \\ 1 \\ 2 \\ -10 \end{bmatrix} \right\}, \left\{ \vec{0} \right\},$$

respectively.

(b) It is not possible to find such a vector. If $\vec{b}$ is in the column space, then $A\vec{x} = \vec{b}$ will have a unique solution, which follows immediately from the zero-dimensional nullspace.

(c) First, show that the vector $\vec{d}$ is not in the column space. The most straight forward way to do this is to do a Gaussian elimination on the augmented matrix $\left[A|\vec{d}\right]$. For the second part, see Chapter **??** on the derivation of the normal equations.

---

* Exercise 4.17

Indicate whether the following if-statement is TRUE or FALSE and motivate your answers clearly. To show a statement is false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

Let $\vec{x}$ and $\vec{y}$ be vectors in $\mathbb{R}^n$, and P be the projection matrix that projects a vector onto some subspace of $\mathbb{R}^n$. If $\vec{x}$ and $\vec{y}$ are orthogonal, then $P\vec{x}$ and $P\vec{y}$ are also orthogonal.

Answer for Exercise 4.17

Exercise 4.18

We are given the following matrix A:

$$A = \begin{bmatrix} c & 4 & 0 \\ 4 & c & 3 \\ 0 & 3 & c \end{bmatrix}$$

(a) For $c = 0$, find a basis for the row space of A.

(b) Find all vectors $\vec{b}$ for which $A\vec{x} = \vec{b}$ does NOT have a solution. These vectors will be dependent on c.

Answer for Exercise 4.18

(a)

$$A = \begin{bmatrix} 0 & 4 & 0 \\ 4 & 0 & 3 \\ 0 & 3 & 0 \end{bmatrix}$$

Since the first and third rows are dependent and independent of the second row, the basis is given by

$$\vec{b}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{b}_2 = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix}$$

(b) Since A is square, to find vectors $\vec{b}$ that do not have a solution, we first find values c that make the matrix singular. To do this we first perform Gaussian elimination:

$$U = \begin{bmatrix} c & 4 & 0 \\ 0 & c - \frac{16}{c} & 3 \\ 0 & 0 & c - \frac{9}{c - \frac{16}{c}} \end{bmatrix}$$

We already know $c = 0$ produces a singular matrix so let's look at the second pivot. For $c = 4$, the second pivot is 0 but we can perform a row swap and achieve a full rank U. Finally, we go to the last pivot:

$$0 = c - \frac{9}{c - \frac{16}{c}} \rightarrow 0 = c^3 - 16c - 9c = c(c^2 - 25) = c(c + 5)(-5).$$

So we conclude that $c = 0, 5, -5$ are the values that make A singular. Now we plug in to find the vectors $\vec{b}$. For $c = 0$, the vectors that span the column space are

$$\vec{b}_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad \vec{b}_2 = \begin{bmatrix} 4 \\ 0 \\ 3 \end{bmatrix}$$

For $c = 5$, the columns that span the column space are (using the first two columns of U),

$$\vec{b}_1 = \begin{bmatrix} 5 \\ 4 \\ 0 \end{bmatrix}, \quad \vec{b}_2 = \begin{bmatrix} 4 \\ 5 \\ 3 \end{bmatrix}$$

And for $c = -5$, the columns that span the column space are

$$\vec{b}_1 = \begin{bmatrix} -5 \\ 4 \\ 0 \end{bmatrix}, \quad \vec{b}_2 = \begin{bmatrix} 4 \\ -5 \\ 3 \end{bmatrix}$$

The vectors that don't have solutions are the ones that are not in the span of those vectors depending on the value c.

Exercise 4.19

We are given the following matrix A:

$$A = \begin{bmatrix} 1 & 1 & 1 & 2 \\ 1 & 2 & -1 & 1 \\ -1 & -4 & 5 & 1 \end{bmatrix}$$

(a) Find the LU decomposition of A. Give both L and U.

(b) Find all $\vec{x}$ for which $A\vec{x} = \vec{0}$.

(c) Does $A\vec{x} = \vec{b}$ have a solution for $\vec{b} = \begin{bmatrix} 2 \\ 3 \\ -5 \end{bmatrix}$?

Answer for Exercise 4.19

Chapter 5

Exercises & Solutions

Exercise 5.1

Indicate whether the following statement is TRUE or FALSE and motivate your answer clearly. To show a statement false, it is sufficient to give one counter example. To show a statement true, provide a general proof.

if Q is an orthogonal matrix, then $\|Q\vec{x}\|_2 = \|\vec{x}\|_2$ for any $\vec{x}$ (you may assume that Q and $\vec{x}$ are real).

Answer for Exercise 5.1

Exercise 5.2

(a) Using Gram-Schmidt orthogonalization, create an orthonormal basis for $\mathbb{R}^2$ from the vectors

$$\vec{a}_1 = \begin{bmatrix} 3 \\ -4 \end{bmatrix}, \quad \text{and } \vec{a}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

(b) Find the QR decomposition of the matrix

$$A = \begin{bmatrix} 3 & 1 & 1 \\ -4 & 1 & -1 \end{bmatrix},$$

where Q is a 2×2 matrix and R is 2×3.

Answer for Exercise 5.2

(a) $\vec{q}_1 = \frac{\vec{a}_1}{\|\vec{a}_1\|_2} = \frac{1}{\sqrt{3^2+(-4)^2}} \begin{bmatrix} 3 \\ -4 \end{bmatrix} = \begin{bmatrix} 3/5 \\ -4/5 \end{bmatrix}$.

$\vec{w}_2 = a_2 - (\vec{q}_1^T \vec{a}_2)\vec{q}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \left(\begin{bmatrix} 3/5 & -4/5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right) \begin{bmatrix} 3/5 \\ -4/5 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} -3/25 \\ 4/25 \end{bmatrix} = \begin{bmatrix} 28/25 \\ 21/25 \end{bmatrix}$

Consider $\|\vec{w}_2\|_2 = \sqrt{\left(\frac{28}{25}\right)^2 + \left(\frac{21}{25}\right)^2} = \frac{7}{5}$. So $\vec{q}_2 = \frac{\vec{w}_2}{\|\vec{w}_2\|} = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$.

Therefore, the orthogonal basis is $\left\{ \begin{bmatrix} 3/5 \\ -4/5 \end{bmatrix}, \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} \right\}$.

(b) The first two columns of matrix A are vectors $\vec{a}_1$ and $\vec{a}_2$, see above. We know that $\vec{a}_1, \vec{a}_2$ form a basis of $\mathbb{R}^2$. So $Q = \begin{bmatrix} \vec{q}_1 & \vec{q}_2 \end{bmatrix} = \begin{bmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{bmatrix}$

We now want to find R such that $A = QR$. $A = QR$ gives $R = Q^T A = \begin{bmatrix} 5 & -1/5 & 7/5 \\ 0 & 7/5 & 1/5 \end{bmatrix}$.

Exercise 5.3

(a) (i) Use `Matlab` command `A=rand(4)` to generate a random 4-by-4 matrix and then use the function `qr` to find an orthogonal matrix Q and an upper triangular matrix R such that $A = QR$. Compute the determinants of A, Q and R.

(ii) Set `A=rand(n)` for at least 5 different n's in `Matlab` for computing the determinant of Q where Q is the orthogonal matrix generated by `qr(A)`. What do you observe about the determinants of the matrices Q? Show, with a mathematical proof, that the determinant of any orthogonal matrix is either 1 or -1.

(iii) For a square $n \times n$ matrix matrix B, suppose there is an orthogonal matrix Q and an upper-triangular matrix R such that $B = QR$. Show that if a vector x is a linear combination of the first k column vectors of B with $k \leq n$, then it can also be expressed as a linear combination of the first k columns of Q.

*(b) (i) Assume $\{\vec{v}_1, \vec{v}_2 \cdots \vec{v}_n\}$ is an orthonormal basis of $\mathbb{R}^n$. Suppose there exists a unit vector $\vec{u}$ such that $\vec{u}^T \vec{v}_k = 0$ for all $k = 2, 3 \cdots n$, show that $\vec{u} = \vec{v}_1$ or $\vec{u} = -\vec{v}_1$.

(ii) Prove that if C is non-singular and $C = QR$, where Q is an orthogonal matrix and R is an upper-triangular matrix with diagonal elements all positive, then the Q and R are unique.
Hint: Use proof by contradiction.

Answer for Exercise 5.3

---

* Exercise 5.4

Suppose that $A = QR$ is the QR-factorization of the $m \times n$ matrix A, with $m > n$, and $r(A) = n$. Show that the projection $\vec{x}$ of a vector $\vec{b}$ onto the column space of A is given by $R\vec{x} = Q^T\vec{b}$.

Answer for Exercise 5.4

The required projection can be found by solving the normal equations. So, $A^TA\vec{x} = A^T\vec{b}$. Substitute $A = QR$ and use the fact that Q is orthogonal (so $QQ^T = I$) we have

$$A^TA\vec{x} = A^T\vec{b} \rightarrow R^TQ^TQR\vec{x} = R^TQ^T\vec{b} \rightarrow R\vec{x} = Q^T\vec{b}.$$

Note that since $r(A) = n$, the $n \times n$ matrix R is invertible.

Exercise 5.5

The matrix A is given by $A = \begin{bmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \end{bmatrix}$. Find a 2×2 orthogonal matrix Q and a 2×3 upper triangular matrix R such that $A = QR$.

Answer for Exercise 5.5

Exercise 5.6

Find an orthonormal basis for the column space of the following matrix:

$$A = \begin{bmatrix} 1 & -6 \\ 3 & 6 \\ 4 & 8 \\ 5 & 0 \\ 7 & 8 \end{bmatrix}$$

Answer for Exercise 5.6

Take

$$\vec{q}_1 = \frac{\vec{a}_1}{\|\vec{a}_1\|} = \begin{bmatrix} 1/10 \\ 3/10 \\ 4/10 \\ 5/10 \\ 7/10 \end{bmatrix}$$

Take

$$\begin{aligned} \vec{q}_2 &= \vec{a}_2 - \vec{a}_1^T \vec{q}_1 \cdot \vec{q}_1 \\ \vec{a}_1^T \vec{q}_1 &= (-6)\frac{1}{10} + 6\frac{3}{10} + 8\frac{4}{10} + 0\frac{5}{10} + 8\frac{7}{10} = 10 \\ \implies \vec{w}_2 &= \begin{bmatrix} 7/10 \\ -3/10 \\ -4/10 \\ 5/10 \\ -1/10 \end{bmatrix}. \end{aligned}$$

Since $\vec{w}_2$ has length 1, we find $\vec{q}_2 = \vec{w}_2$ immediately.

Chapter 6

Exercises & Solutions

Exercise 6.1

Let P be a $n \times n$ projection matrix. Prove that $|P| = \pm 1$.

Answer for Exercise 6.1

Exercise 6.2

A is an $n \times n$ matrix. Prove that $|A| = |A^T|$.

Answer for Exercise 6.2

We can use several approaches to prove this. Our first approach is to use the general LU-decomposition: we know that for any matrix A, we have $PA = LU$, where P is a permutation matrix, and L and U are lower- and upper-triangular matrices, respectively. Note that we cannot just set $A = LU$ as it is not guaranteed we can find an LU-decomposition of a matrix without pivoting (we may encounter a zero pivot). A permutation matrix is orthogonal (it is itself just a permutation of the identity matrix), so $P^{-1} = P^T$, and $|P| = \pm 1$. So, we can write $A = P^T LU$, which gives $|A| = |P^T LU| = |P^T||L||U|$. We know

that $|L| = 1$, and $|U| = |U^T|$ (do you see why?), and also that $|P| = |P^T|$. This latter statement is easy to understand: P is a permutation of the identity matrix. Suppose that it required p row/column swaps to create P from I. If p is even, $|P| = 1$. If p is odd, $|P| = -1$. It will take just as many swaps to create P^T from I and so $|P| = |P^T|$. This gives $|A^T| = |U^T|\,|L||P| = |U|\,|P|$, and $|A| = |P^T|\,|L|\,|U| = |P||U|$, and clearly $|A| = |A^T|$.

Exercise 6.3

Prove that the determinant of an orthogonal matrix Q is given by $|Q| = \pm 1$.

Answer for Exercise 6.3

Exercise 6.4

In this exercise, we look at a geometric interpretation of the determinant in two dimensions.

Let $\vec{u} \in \mathbb{R}^2$ and $\vec{v} \in \mathbb{R}^2$ be the columns of a 2×2 matrix A, that is,

$$A = \begin{bmatrix} \vec{u} & \vec{v} \end{bmatrix} = \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}.$$

Prove that $|\det(A)|$ is the area of the parallelogram whose vertices are the origin, $\vec{u}$, $\vec{v}$, and $\vec{u} + \vec{v}$ (see Figure 6.1).

Answer for Exercise 6.4

If we take the line segment between the origin and $\vec{u}$ to be the base of the parallelogram, then the area of the parallelogram is

$$\text{area} = (\text{base})(\text{height}) = (\|\vec{u}\|)(\|\vec{v}\|\sin(\theta)),$$

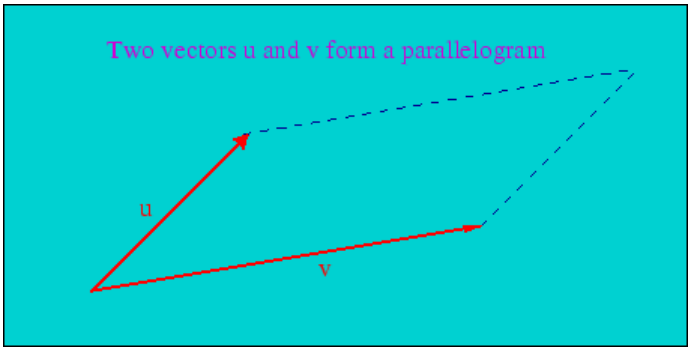

Figure 6.1: The parallelogram formed by two vectors $\vec{u}$ and $\vec{v}$.

where θ is the angle between $\vec{u}$ and $\vec{v}$. We square both sides to obtain

$$\begin{aligned}
\text{area}^2 &= \|\vec{u}\|^2\|\vec{v}\|^2\sin^2(\theta) \\
&= \|\vec{u}\|^2\|\vec{v}\|^2(1-\cos^2(\theta)) \\
&= \|\vec{u}\|^2\|\vec{v}\|^2\left(1-\left(\frac{\vec{u}^T\vec{v}}{\|\vec{u}\|\|\vec{v}\|}\right)^2\right) \\
&= \|\vec{u}\|^2\|\vec{v}\|^2-(\vec{u}^T\vec{v})^2.
\end{aligned}$$

We expand this expression in terms of the components of $\vec{u}$ and $\vec{v}$ to obtain

$$\begin{aligned}
\text{area}^2 &= (u_1^2+u_2^2)(v_1^2+v_2^2)-(u_1v_1+u_2v_2)^2 \\
&= u_1^2v_2^2+u_2^2v_1^2-2u_1v_1u_2v_2 \\
&= (u_1v_1-u_2v_2)^2 \\
&= \det(A)^2.
\end{aligned}$$

We take the square root of both sides to get the required result area $= |\det(A)|$.

Exercise 6.5

The $m \times m$ tridiagonal matrix A_m is given by

$$\begin{bmatrix} b & c & & & \\ a & b & c & & \\ & \ddots & \ddots & \ddots & \\ & & a & b & c \\ & & & a & b \end{bmatrix}$$

We write the determinant of the matrix as $D_m = \det(A_m)$.

(a) Show that $D_m = bD_{m-1} - acD_{m-2}$.

(b) Write the difference equation in 1. as a first-order system of two equations of the form $x^{(m)} = Ax^{(m-1)}$, where $x^{(m)} = \begin{bmatrix} D_m \\ D_{m-1} \end{bmatrix}$.

Answer for Exercise 6.5

---

* Exercise 6.6

(a) Using the system found in part 2 of Exercise 6.5, show that

$$D_m = \frac{r^m}{\sin q} \sin\left((m+1)q\right),$$

where $r = \sqrt{ac}$ and $2r\cos q = b$, given that $ac > 0$ and $\left|\frac{b}{2r}\right| \leq 1$.

(b) If $a_j = \frac{j\pi}{m+1}$ for $j = 1, \ldots, m$, show that

$$I_j = b + 2\sqrt{ac}\cos a_j$$

are the eigenvalues of A_m. Note that you do not need the solution to part 2 of Exercise 6.5 to answer this question.

Answer for Exercise 6.6

(a) The determinant of A_m is unique, so if the given sequence D_m satisfies the above system then it must be the determinant of A_m. That D_m solves the system can be verified directly. The product with the second row is trivial, so consider the first row:

$$\begin{aligned} bD_{m-1} - acD_{m-2} &= b\frac{r^{m-1}}{\sin q}\sin mq - ac\frac{r^{m-2}}{\sin q}\sin(mq-q) \\ &= 2r\cos q\frac{r^{m-1}}{\sin q}\sin mq - r^2\frac{r^{m-2}}{\sin q}\sin(mq-q) \\ &= 2r\cos q\frac{r^{m-1}}{\sin q}\sin mq - r\cos q\frac{r^{m-1}}{\sin q}\sin mq + \frac{r^m}{\sin q}\cos mq\sin q \\ &= r\cos q\frac{r^{m-1}}{\sin q}\sin mq + \frac{r^m}{\sin q}\cos mq\sin q \\ &= \frac{r^m}{\sin q}\sin(mq+q) \end{aligned}$$

The second to last and last equalities come from the sine difference and sum identities applied to $mq-q$ and $mq+q$. Therefore D_m satisfies the system from part 2 of Exercise 6.5, and so $D_m = \det A_m$ as desired.

(b) Notice that any eigenvalue λ and eigenvector $\vec{v}$ must satisfy

$$v_{k+2} + \left(\frac{b-\lambda}{c}\right)v_{k+1} + \left(\frac{a}{c}\right)v_k = 0$$

for $k = 0, \ldots, m-1$ with $v_0 = v_{m+1} = 0$. This is a system of second order difference equations which has a solution of the form $v_k = \xi r^k$ for some constants ξ, r. The characteristic equation of the system is

$$r^2 + \left(\frac{b-\lambda}{c}\right)r + \left(\frac{a}{c}\right) = 0,$$

which leads to solutions of the form $v_k = \alpha r_1^k + \beta r_2^k$, where r_1, r_2 are the roots of the characteristic equation. Applying the boundary conditions, we find

$$\begin{aligned} \alpha + \beta &= 0 \\ \alpha r_1^{n+1} + \beta r_2^{n+1} &= 0 \end{aligned}$$

So

$$\left(\frac{r_1}{r_2}\right)^{n+1} = -\frac{\beta}{\alpha} = 1$$

which is satisfied by

$$r_1 = r_2 e^{i2\pi j/(m+1)}$$

for $1 \leq j \leq m$. Substituting back into the characteristic equation yields

$$r_1 = \sqrt{\frac{a}{c}} e^{i\pi j/(m+1)},\ r_2 = \sqrt{\frac{a}{c}} e^{-i2\pi j/(m+1)} \implies \lambda = b + 2\sqrt{ac}\cos a_j$$

for $1 \leq j \leq m$. So the eigenvalues of A_m are

$$I_j = b + 2\sqrt{ac}\cos a_j$$

as desired.

Exercise 6.7

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement is false it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) If the determinant of an $n \times n$ matrix A is nearly 0, then the matrix is ill-conditioned

(b) For any $n \times n$ unitary matrix U, $\det(U) = \det(U^*)$.

(c) Every $n \times n$ permutation matrix P has a determinant that is equal to either -1 or $+1$.

Answer for Exercise 6.7

(a) False.

We know that the identity matrix is very well-conditioned (very far from being singular!). Now, take $A = \alpha I$, where $\alpha = 10^{-10}$, for example. The condition number of A is still small (in an exercise in the chapter on conditioning we found that condition numbers are not sensitive to scaling factors), but the determinant is equal to 10^{-10n}, which is extremely small for large n.

(b) False.

For unitary matrices, this is not necessarily true. A counterexample would be the matrix

$$A = \frac{1}{2}\begin{bmatrix} 1+i & 1-i \\ 1-i & 1+i \end{bmatrix},$$

which is unitary.

(c) True.

Permutation matrices are permutations of the identity matrix. If the permutation requires p row or column swaps, and p is even, then the determinant of the permutation matrix is $+1$. If p is odd, the determinant is -1. This follows directly from one of the properties of the determinant discussed in the chapter.

Chapter 7

Exercises & Solutions

Exercise 7.1

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement is false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) For a non-singular $n \times n$ matrix A, the *Jacobi* method for solving $A\vec{x} = \vec{b}$ will always converge to the correct solution, if it converges.

(b) The *Gauss-Seidel* iteration used for solving $A\vec{x} = \vec{b}$ will always converge for any $n \times n$ invertible matrix A.

Answer for Exercise 7.1

Exercise 7.2

(a) Find an invertible 2×2 matrix for which the *Jacobi* method does not converge.

(b) Find an invertible 10×10 non-diagonal matrix for which the *Jacobi* method converges very quickly.

Answer for Exercise 7.2

(a) Take a diagonally inferior matrix (as opposed to diagonally dominant). For example, we would take:

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

To show it does not converge, we can look at the iteration matrix $M^{-1}N$. For *Jacobi* we have

$$M^{-1}N = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & -2 \\ -2 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$$

It is easy to see that the Frobenius norm of this amplification matrix is larger than 1, and we expect trouble. We can set up an iteration to show that this iteration does not converge, as long as we don not take the initial guess to be equal to the exact solution to $A\vec{x} = \vec{b}$. For example, we can try to solve for $\vec{b} = \begin{bmatrix} 3 \\ 0 \end{bmatrix}$, which has an exact solution $\vec{x} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. We get as iterates (starting with $\vec{x}^{(0)}$ equal to $\vec{b}$).

$$\vec{x}^{(1)} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} + \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix} = \begin{bmatrix} -3 \\ 6 \end{bmatrix}$$
$$\vec{x}^{(2)} = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -3 \\ 6 \end{bmatrix} + \begin{bmatrix} -3 \\ 0 \end{bmatrix} = \begin{bmatrix} 9 \\ -6 \end{bmatrix}, \quad \text{etc.}$$

which is clearly diverging.

(b) We expect that for a matrix that is strongly diagonally dominant, the *Jacobi* method would converge very quickly. For example, we take

$$A = \begin{bmatrix} 100 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 100 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 100 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 100 & 1 \\ 0 & 0 & 0 & \cdots & 1 & 100 \end{bmatrix}_{10\times 10}$$

For this matrix, the amplification matrix is

$$M^{-1}N = \begin{bmatrix} 0.01 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0.01 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0.01 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0.01 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0.01 \end{bmatrix}_{10\times 10} \begin{bmatrix} 0 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 0 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -1 \\ 0 & 0 & 0 & \cdots & -1 & 0 \end{bmatrix}_{10\times 10}$$

$$= \begin{bmatrix} 0 & -0.01 & 0 & \cdots & 0 & 0 \\ -0.01 & 0 & -0.01 & \cdots & 0 & 0 \\ 0 & -0.01 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -0.01 \\ 0 & 0 & 0 & \cdots & -0.01 & 0 \end{bmatrix}_{10\times 10}$$

which has a Frobenius norm much smaller than 1 (actually the Frobenius norm is around 0.04). Hence, the convergence is guaranteed and will be quite fast.

Exercise 7.3

The matrix A given by

$$A = \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix}.$$

Show that neither the Jacobi method nor the Gauss-Seidel method converges when used to solve the equation $A\vec{x} = \vec{b}$, with $\vec{b}$ an arbitrary vector in $\mathbb{R}^2$.

Answer for Exercise 7.3

Exercise 7.4

Show that both the Jacobi and Gauss-Seidel iterations converge for the matrix $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$. Which iteration converges faster, and by what factor?

Answer for Exercise 7.4

Consider the Jacobi method,

$$M = D = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}, N = M - A = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \rightarrow G = M^{-1}N = \begin{bmatrix} 0 & -1/2 \\ -1/3 & 0 \end{bmatrix}$$

Since $\|G\|_F = 0.60 < 1$, so Jacobi method converges.

Consider the Gauss-Seidel method,

$$M = D + L = \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix}, N = \begin{bmatrix} 0 & -1 \\ 0 & 0 \end{bmatrix} \rightarrow G = M^{-1}N = \begin{bmatrix} 0 & -1/2 \\ 0 & 1/6 \end{bmatrix}$$

Since $\|G\|_F = 0.53 < 1$, so Gauss-Seidel method converges.

---

* Exercise 7.5

One of your friends invited a new iterative scheme for solving the system of equations $A\vec{x} = \vec{b}$ for real and nonsingular $n \times n$ matrices A. The scheme is given by

$$\vec{x}^{(k+1)} = (I + \beta A)\vec{x}^{(k)} - \beta\vec{b}, \quad \text{with } \beta > 0.$$

(a) Show that **if** this scheme converges, it converges to the desired solution of the system of equations. In other words, your friend seems to be on to something.

(b) Derive an equation for the error $\vec{e}^{(k)} = \vec{x}^{(k)} - \vec{x}^*$, where $\vec{x}^*$ is the exact solution, for each iteration step k.

Answer for Exercise 7.5

Chapter 8

Exercises & Solutions

Exercise 8.1

(a) Find the best straight-line fit to the following measurements, and graph your solution:

$$y_1 = 2 \text{ at } t_1 = -1, \quad y_2 = 0 \text{ at } t_2 = 0,$$
$$y_3 = -3 \text{ at } t_1 = 1, \quad y_4 = -5 \text{ at } t_4 = 2,$$

What is the norm of the residual?

(b) Suppose that instead of a straight line, we fit the data above by the parabolic function:

$$y_i = a_2 x_i^2 + a_1 x_i + a_0$$

Derive the over-determined system $A\vec{x} = \vec{b}$ to which least squres could be applied to find this quadratic fit.

(c) Let's look at the general problem of making n observations $y_i, i = 1, 2, \ldots, n$ at n different times t_i. You can extend what you did in the last two parts to find polynomial fits of degree k ($y_i = a_k t_i^k + a_{k-1} t_i^{k-1} + \cdots + a_1 t_i + a_0$) by using least squares. If $k < n-1$, what would the over-determined system $A\vec{x} = \vec{b}$ look like for this general case?

*(d) Prove that for $k = n - 1$, the system $A\vec{x} = \vec{b}$ will no longer be over-determined and we can find a unique fit by solving $A\vec{x} = \vec{b}$ instead of the normal equations.

(e) Consider the systems you solved for in part 3 and 4. For $0 < k < n$, how does the norm of the residual change as we increase k?

Answer for Exercise 8.1

Exercise 8.2

(a) From the notes, we know that $P = A\left(A^TA\right)^{-1}A^T$ is an orthogonal projection matrix that projects onto the column space of A. Prove that P is symmetric and $P^2 = P^TP = P$.

*(b) In general, a matrix which satisfies $P^2 = P$ is a projector. Show that if a projector matrix is symmetric, then it is an orthogonal projector.

**(c) Show that regardless of rank of A, the equation $A^TA\vec{x} = A^T\vec{b}$ always has at least one solution

Answer for Exercise 8.2

(a)

$$P^T = \left(A\left(A^TA\right)^{-1}A^T\right)^T = \left(A^T\right)^T\left(\left(A^TA\right)^{-1}\right)^TA^T = A\left(\left(A^TA\right)^T\right)^{-1}A^T$$
$$= A\left(A^TA\right)^{-1}A^T = P$$

So P is symmetric. Also $P^2 = A\left(A^TA\right)^{-1}A^TA\left(A^TA\right)^{-1}A^T = AI\left(A^TA\right)^{-1}A^T = P$.

*(b) An orthogonal projection is a matrix such that for any $\vec{x}$, $\vec{x} - P\vec{x}$ is orthogonal to $P\vec{x}$. From part (a), we know that

$$P^2 = P = P^T, \quad \text{and} \quad P^TP = P$$

This now gives

$$\left(\vec{x} - P\vec{x}\right)^TP\vec{x} = \left(\vec{x}^T - \vec{x}^TP^T\right)P\vec{x} = \vec{x}^TP\vec{x} - \vec{x}^TP^TP\vec{x} = \vec{x}^TP\vec{x} - \vec{x}^TP\vec{x} = 0$$

, which shows the projection is orthogonal.

(c) There are a couple of ways to approach this problem. Here is just one. Others can be found in the reference books.

If A is an $m \times n$ matrix, and $\vec{b}$ is a vector in $\mathbb{R}^m$, then since $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are orthogonal complements of the space $\mathbb{R}^m$, we can write any $\vec{b}$ in terms of the m basis vectors of these two spaces. Segregating the parts of $\vec{b}$ in $\mathcal{R}(A)$ and $\mathcal{N}(A)$, we have

$$\vec{b} = \vec{b}_r + \vec{b}_n,$$

where $\vec{b}_r \in \mathcal{R}(A)$ and $\vec{b}_n \in \mathcal{N}(A)$. Note that since $\vec{b} - \vec{b}_r = \vec{b}_n \in \mathcal{N}\left(A^T\right) \perp \mathcal{R}(A), \vec{b}_r$ is the orthogonal projection of $\vec{b}$ onto the column space of A, we can find at least one $\vec{x}'$, such that $A\vec{x}' = \vec{b}_r$, and we can write

$$A^T\left(A\vec{x} - \vec{b}\right) = A^T\left(A\vec{x} - \vec{b}_r - \vec{b}_n\right) = A^T\left(A\vec{x} - \vec{b}_r\right) - A^T\vec{b}_n = A^T\left(A\vec{x} - \vec{b}_r\right) = \vec{0}$$

Ans we know that $A\vec{x}' = \vec{b}_r$, so $\vec{x}'$ is an answer to this equation.

Exercise 8.3

We measured the temperature T below the ground on an unusually cold day a few weeks ago. The temperature was measured as a function of the distance from the ground surface. The outside temperature was a balmy 2 deg. Celsius. The data from the measurements are given in the table below:

Distance (m)	Temperature (C)
0	2.0
5	2.2
10	5.8
15	10.4
20	11.0
25	13.8
30	22.4
35	28.4
40	33.3

(a) Write a matlab function that fits the data to a polynomial of degree n using the method of least squares. Make sure that your function allows you to specify n. (Do not use matlab built-in functions `polyfit` or `polyval` except perhaps to check that your code is correct.) Plot the data. On the same axes, plot the polynomial fit for $n = 1$ and $n = 2$. Be sure to clearly label your fit curves.

(b) Calculate the residual error in your fitted values and the observed data for n ranging from 0 to 8.

Plot the 2-norm of this residual error against n.

Comment on what does this result says about how to choose the order of your polynomial fit.

(c) We improve our drilling and sensing methodology and decide that we can drill to 45m and 50m below ground with minimal effort. We want to estimate the temperature at this new data point.

(i) Provide a table of n versus the predicted temperature at these new data points. It turns out that the temperatures are 48.9 deg. Celsius at 45m and 57.9 deg. Celsius at 50m below ground, respectively.

(ii) Plot the 2-norm of the prediction error only at these two points versus n.

(iii) Comment on what does this result says about how to choose the order of your polynomial fit? Be sure to use what you learned from the previous problem.

Answer for Exercise 8.3

---

* Exercise 8.4

We know that the normal equations for the $m \times n$ matrix A given by $A^T A\vec{x} = A^T\vec{b}$ are solved by the vector $\vec{x}$ for which $\|\vec{b} - A\vec{x}\|_2$ is minimal. Show that if A has full column rank, the normal equations give a unique solution $\vec{x}$.

Answer for Exercise 8.4

If A is full rank, then R is $n \times n$ as well as full rank and invertible in the reduced-form QR factorization. So

$$A^T A\vec{x} = A^T\vec{b} \rightarrow R^T Q^T QR\vec{x} = R^T R\vec{x} = R^T Q^T\vec{b}$$

Since R^T is invertible, we can remove it from the equation and obtain

$$R\vec{x} = Q^T\vec{b},$$

which will only have one solution because R is nonsingular.

Chapter 9

Exercises & Solutions

Exercise 9.1

We are interested in finding the fixed points (the points at which the time derivatives are zero) of the following system of equations:

$$\frac{dx_1}{dt} = x_1(a - bx_2)$$
$$\frac{dx_2}{dt} = -x_2(c - dx_1)$$

for $a = 3,\ b = 1,\ c = 2,\ d = 1$. We can use the Newton-Raphson method to find these fixed points, simply by setting the derivatives zero in the given system of equations.

(a) In the scalar case, Newton-Raphson breaks down at points at which the derivative of the nonlinear function is zero. In general, where can it break down for systems of nonlinear equations? For the system given above, find the troublesome points.

(b) Find the fixed points of the above system analytically.

(c) Find all fixed points using repeated application of the Newton-Raphson method. You will have to judiciously choose your starting points (but of course, you are not allowed to use the known roots as starting points!). You may use MATLAB to program the method if you like.

Answer for Exercise 9.1

Exercise 9.2

This exercise involves higher order Taylor expansions.

(a) Assuming that a single-variable function is twice-differentiable you can approximate it in a point x with its second order Taylor expansion at a near-by point x_0

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{1}{2}(x - x_0)^2 f''(x_0)$$

Based on this, approximate $f'(x)$ with a linear function in the proximity of x_0. Then explain how you could apply Newton-Raphson to find the critical points of a scalar function ($f'(x) = 0$).

(b) Now extend this idea to derive an algorithm for finding extrema of functions $f(\vec{x}) = f(x_1, x_2, \ldots, x_n)$ of more than one variable by exploiting the second order Taylor expansion, which for such multi-variable functions is given by

$$f(\vec{x}) = f(\vec{x}_0) + (\nabla f(\vec{x}_0))^T (\vec{x} - \vec{x}_0) + \frac{1}{2} (\vec{x} - \vec{x}_0)^T H(x_0) (\vec{x} - \vec{x}_0),$$

where $\vec{x} = [x_1, x_2, \ldots, x_n]^T$ and $\nabla f(\vec{x}) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T$.
To find what the Hessian is, see http://en.wikipedia.org/wiki/Hessian˙matrix .

*(c) Newton-Raphson is a method for solving a system of nonlinear equations:

$$f(x_1, x_2, \ldots, x_n) = \begin{bmatrix} f_1(x_1, x_2, \ldots, x_n) \\ f_2(x_1, x_2, \ldots, x_n) \\ \vdots \\ f_n(x_1, x_2, \ldots, x_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

If you recall, we used linear approximations of f_i to derive this method. Now use the quadratic approximation introduced in the last section for $f_i(\vec{x}) = f_i(x_1, x_2, \ldots, x_n)$ to reduce these equations f_i to a quadratic form and then design a general algorithm for solving the nonlinear system of equations based on this approximation. (no need to program this up!)

Answer for Exercise 9.2

(a) Based on the approximation given by the second order Taylor expansion in the vicinity of x_0, we can estimate $f'(x)$ in the following manner

$$f'(x) = f'(x_0) + (x - x_0)f''(x_0)$$

If we introduce $g(x) = f'(x)$, then we see that we have a linear approximation for $g(x)$ and we can apply Newton-Raphson to this function in order to find the zeros of it

$$x^{(k+1)} = x^{(k)} - \frac{g\left(x^{(k)}\right)}{g'\left(x^{(k)}\right)} = x^{(k)} - \frac{f'\left(x^{(k)}\right)}{f''\left(x^{(k)}\right)}$$

(b) The extrema of such functions can be found by equating the gradient to zero. It was proven in workshop that $\nabla_{\vec{x}}\left(\vec{x}^T\vec{c}\right) = \nabla_{\vec{x}}\left(\vec{c}^T\vec{x}\right) = \vec{c}$ and $\nabla_{\vec{x}}\left(\vec{x}^T B\vec{x}\right) = 2B\vec{x}$ and also since $\vec{x}_0$ is a constant, it does not make a difference if we take the gradient with respect to $\vec{x} - \vec{x}_0$ or $\vec{x}$. Thus

$$\nabla f\left(\vec{x}\right) = \nabla f\left(\vec{x}_0\right) + H\left(\vec{x}_0\right)\left(\vec{x} - \vec{x}_0\right)$$

Moreover, from the definition of the Hessian matrix, we can write $J\left(\nabla f(\vec{x})\right) = H(\vec{x})$. Thus, if we take $g(\vec{x}) = \nabla f(\vec{x})$, then

$$g(\vec{x}) = g(\vec{x}_0) + J(\vec{x}_0)(\vec{x} - \vec{x}_0)$$

Note that J is the Jacobian of $g(\vec{x})$. Our goal is to find where $g(\vec{x}) = \nabla f(\vec{x}) = 0$, thus we use Newton-Raphson to find the zeros of this function by solving the following equation in each iteration step

$$J\left(\vec{x}^{(k)}\right)\left(\vec{x}^{(k+1)} - \vec{x}^{(k)}\right) = -g\left(\vec{x}^{(k)}\right)$$

which is the same as solving

$$H\left(\vec{x}^{(k)}\right)\left(\vec{x}^{(k+1)} - \vec{x}^{(k)}\right) = -\nabla f\left(\vec{x}^{(k)}\right)$$

to find $\vec{x}^{(k+1)}$ from $\vec{x}^{(k)}$.

*(c) The quadratic approximation yields

$$f_i\left(\vec{x}\right) = f_i\left(\vec{x}_0\right) + \left(\nabla f_i\left(\vec{x}_0\right)\right)^T\left(\vec{x} - \vec{x}_0\right) + \frac{1}{2}\left(\vec{x} - \vec{x}_0\right)^T H_i\left(x_0\right)\left(\vec{x} - \vec{x}_0\right),$$

for $1 \leq i \leq n$. We now want to find where $f_i(\vec{x}) = 0$. Hence, in the same way we derived Newton-Raphson from a linear approximation let us set $f_i(\vec{x}) = 0$ and find $\vec{x}$.

In other words, in every iteration step, we want to solve this equation for $1 \leq i \leq n$

$$\begin{aligned} F_i^{(k)}\left(\vec{x}^{(k+1)}\right) &= f_i\left(\vec{x}^{(k)}\right) + \left(\nabla f_i\left(\vec{x}^{(k)}\right)\right)^T \left(\vec{x}_{(k+1)} - \vec{x}^{(k)}\right) \\ &\quad + \frac{1}{2}\left(\vec{x}_{(k+1)} - \vec{x}^{(k)}\right)^T H_i\left(x^{(k)}\right)\left(\vec{x}_{(k+1)} - \vec{x}^{(k)}\right) \\ &= 0 \end{aligned}$$

By having the definition of $f_i(\vec{x})$, and by having the vector $\vec{x}^{(k)}$, this equation is a quadratic equation of $x_1, x_2, \ldots, x_n$. Furthermore, we have n of these equations. So this is a system of nonlinear equations of quadratic form with n equations and n unknowns. Therefore, by naming each new equation in the k^{th} step of iteration $F_i^{(k)}$, we need to solve the following system to find $\vec{x}^{(k+1)}$

$$F^{(k)}(x_1, x_2, \ldots, x_n) = \begin{bmatrix} F_1^{(k)}(x_1, x_2, \ldots, x_n) \\ F_2^{(k)}(x_1, x_2, \ldots, x_n) \\ \vdots \\ F_n^{(k)}(x_1, x_2, \ldots, x_n) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Note that $F^{(k)}$ should be updated in each iteration step and each $F_i^{(k)}$ is a quadratic function of $x_1, x_2, \ldots, x_n$. We can solve this new nonlinear system of equations by a number of methods, one being Newton-Raphson. In this case, the general algorithm would look like this

Algorithm 1 Algorithm for solving nonlinear equations $f(\vec{x}) = \vec{0}$

Choose $\vec{x}^{(1)}$
$k = 0$
while $\|\vec{x}^{(k+1)} - \vec{x}^{(k)}\|_2 >$ tolerance$_2$ **do**
  $k = k + 1$
  for $i = 1, \ldots, n$ **do**
    Find $F_i^{(k)}$
  end for
  $\vec{y}^{(1)} = \vec{x}^{(k)}$
  $j = 0$
  while $\|\vec{y}^{(j+1)} - \vec{y}^{(j)}\|_2 >$ tolerance$_1$ **do**
    $j = j + 1$
    Solve $J\left(F^{(k)}\left(\vec{y}^{(j)}\right)\right)\left(\vec{y}^{(j+1)} - \vec{y}^{(j)}\right) = -F^{(k)}\left(\vec{y}^{(j)}\right)$ to find $\vec{y}^{(j+1)}$
  end while
end while

Note that we have a nested loop in this method which uses the Newton-Raphson method to find the solution to the quadratic system of equations. This is the main drawback of this method; it is computationally expensive because unlike the Newton-Raphson method, this method reduces the system of equations to another nonlinear system which needs to be solved computationally. Maybe, if one is able to utilize analytical measures to solve the quadratic system of equations, then this method would become an efficient method comparing to Newton-Raphson.

Exercise 9.3

The vector function $\vec{f}(\vec{x})$ is given by

$$\vec{f}(\vec{x}) = \begin{bmatrix} x_1^2 - x_1 x_2 \\ x_1 x_2 - x_2^2 \end{bmatrix}.$$

Give the Newton-Rhapson algorithm for solving $\vec{f}(\vec{x}) = \vec{0}$. Perform two steps in this algorithm with the start vector $\vec{x} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. If you were to continue this process, what convergence criterion or criteria would you use?

Answer for Exercise 9.3

Chapter 10

Exercises & Solutions

Exercise 10.1

Consider the symmetric matrix

$$A = \begin{bmatrix} 9 & -7 \\ -7 & 17 \end{bmatrix}$$

(a) Compute the eigenvalues and eigenvectors of A.

(b) Show that the eigenvectors of A are orthogonal.

Answer for Exercise 10.1

Exercise 10.2

We consider the matrix

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

(a) Show that the eigenvalues of A are $\lambda_1 = -3, \lambda_2 = -2, \lambda_3 = 0$ with corresponding normalized eigenvectors $\vec{y}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix}-1\\-1\\-1\end{bmatrix}, \vec{y}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\0\\1\end{bmatrix}, \vec{y}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix}-1\\2\\-1\end{bmatrix}$. Is A diagonalizable? Motivate your answer.

(b) Does the limit $\lim_{k\to\infty} A^k$ exist (in other words, is it finite)? Motivate your answer.

(c) Obtain the solution, as a function of time t, to the differential equation $\frac{d\vec{x}}{dt} = A\vec{x}$ with initial condition $\vec{x}(0) = \begin{bmatrix}1\\1\\1\end{bmatrix}$, and discuss its behavior as $t \to \infty$.

Answer for Exercise 10.2

(a)
$$0 = \det(A - \lambda I) = -\lambda^3 - 5\lambda^2 - 6\lambda = -\lambda(\lambda^2 + 5\lambda + 6) = -\lambda(\lambda + 2)(\lambda + 3)$$
so the eigenvalues are $\lambda = 0, -2, -3$. Solving $(A - \lambda I)\vec{x} = \vec{0}$ for $\lambda = 0, -2, -3$ will give the eigenvectors given.

A is diagonalizable because there are 3 distinct eigenvalues to this 3×3 system which means there are 3 linearly independent eigenvectors.

(b) Since $A = Y\Lambda Y^{-1}$ because A is diagonalizable,
$$A^k = Y\Lambda^k Y^{-1}$$
But
$$\Lambda^k = \begin{bmatrix}(-3)^k & & \\ & (-2)^k & \\ & & 0^k\end{bmatrix}$$
is blowing up so A^k is blowing up.

(c)
$$\vec{x}(t) = \exp(At)\vec{x}(0) = Y\exp(\Lambda t)Y^T\vec{x}(0)$$
since A is diagonalizable and Y is an orthogonal matrix. We know $\vec{x}(0) - \sqrt{3}\vec{y}_1$, so $Y^T x(0) = -\sqrt{3}e_1$, where $e_1 = \begin{bmatrix}1\\0\\0\end{bmatrix}$. Therefore,
$$\vec{x}(t) = -\sqrt{3}e^{\lambda_1 t}\vec{y}_1 = \begin{bmatrix}e^{-3t}\\e^{-3t}\\e^{-3t}\end{bmatrix}$$

Clearly this solution vanishes as $t \to \infty$.

Exercise 10.3

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) If a symmetric matrix has repeated eigenvalues, it must be singular.

*(b) Any $n \times n$ real skew-symmetric matrix A, with n is odd, is singular. (A *skew-symmetric* matrix A satisfies $A^T = -A$.)

Answer for Exercise 10.3

Exercise 10.4

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) If the $n \times n$ matrix B is formed from the $n \times n$ matrix A by swapping two rows of A, then B and A have the same eigenvalues.

(b) Any invertible $n \times n$ matrix A can be diagonalized.

(c) A singular matrix must have repeated eigenvalues.

(d) If the $n \times n$ matrices A and B are diagonalizable, so is the matrix AB.

(e) Let A be an $n \times n$ matrix.

  (i) The eigenvalues of A and A^T are the same.

(ii) The eigenvalues and eigenvectors of A^TA and AA^T are the same.

Answer for Exercise 10.4

(a) False.

Consider

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 0 & 2 \\ 1 & 0 \end{bmatrix}$$

The eigenvalues of A are 1 and 2, but the eigenvalues of B are $\sqrt{2}$ and $-\sqrt{2}$.

(b) False.

Consider

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

$\det(A) = 1$, so A is invertible. The eigenvalues of A are $\lambda_1 = \lambda_2 = 1$; the corresponding eigenvectors $\vec{v}$ must satisfy $(A - 1I)\vec{v} = \vec{0}$, or

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

Since the matrix is upper triangular, we see by inspection that its column space has exactly one basis vector, as does its null space. Since the null space of $A - 1I$ has only one basis vector, A has only one eigenvector, and therefore is not diagonalizable.

(c) False.

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

The eigenvector matrix can be any invertible 2×2 matrix and A will be singular, but not have repeated eigenvalues.

(d) False.

Consider the two diagonalizable matrices (they have distinct eigenvalues)

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

Multiplication gives

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

which is not diagonalizable as shown in part (b)

(e) (i) True.

Suppose λ is an eigenvalue for A, then we have

$$0 = \det(A - \lambda I) = \det((A - \lambda I)^T) = \det(A^T - \lambda I),$$

and thus λ is an eigenvalue for A^T. The converse is shown by reversing the steps above.

(ii) False.

If two diagonalizable matrices have the same eigenvectors and eigenvalues, then they are the same matrix. But a matrix A that satisfies $AA^T = A^T A$ is called "normal" and not all matrices are "normal". Here is some matlab code to illustrate this:

```
A = rand(2)
[v1,d1]=eig(A'*A)
[v2,d2]=eig(A*A')

A =
     0.8147     0.1270
     0.9058     0.9134
v1 =
     0.5821    -0.8131
    -0.8131    -0.5821
d1 =
     0.1840          0
          0     2.1506
v2 =
    -0.8648     0.5021
     0.5021     0.8648
d2 =
     0.1840          0
          0     2.1506
```

Exercise 10.5

For this problem we assume that eigenvalues and eigenvectors are all real valued.

(a) Let A be an $n \times n$ symmetric matrix. Let $\vec{q}_i$ and $\vec{q}_j$ be the eigenvectors of A corresponding to the eigenvalues λ_i and λ_j respectively. Show that if $\lambda_i \neq \lambda_j$, then $\vec{q}_i$ and $\vec{q}_j$ are orthogonal.

(b) Let A be an $n \times n$ matrix. We say that A is **positive definite** if for any non-zero vector $\vec{x}$, the following inequality holds

$$\vec{x}^T A \vec{x} > 0.$$

Show that the eigenvalues of a positive definite matrix A are all positive.

**(c) Let A be an $n \times n$ matrix. Show that

$$\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i,$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A (λ_i's do not have to be all different).

[Hint 1: One way to prove this is to use the fact that any square matrix with real eigenvalues can be decomposed in the following way (called Schur decomposition)

$$A = QRQ^T,$$

where R is an upper triangular matrix and Q is an orthogonal matrix.]

[Hint 2: The following property of trace might be useful: given two matrices $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times m}$, the trace of their product, $\mathrm{tr}(AB)$, is *invariant under cyclic permutations*, i.e. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.
Note that this implies $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$ for any matrices A, B, C with appropriately chosen dimension.]

Answer for Exercise 10.5

Exercise 10.6

A rabbit population r and a wolf population w are related according to the following system of differential equations:

$$\begin{aligned} \frac{dr}{dt} &= 5r - 2w \\ \frac{dw}{dt} &= r + 2w \end{aligned}$$

Here t denotes time.

(a) If initially (at $t = 0$) there were 100 rabbits and 50 wolves, find the numbers of rabbits and wolves as a function of time t.

(b) Design a matrix A for the above system such that the populations converge to a finite non-zero limit as t goes to infinity.

Answer for Exercise 10.6

(a) If we write $\vec{x}(t) = \begin{bmatrix} r(t) \\ w(t) \end{bmatrix}$, then the system can be written as

$$\frac{d}{dt}\vec{x}(t) = \begin{bmatrix} 5 & -2 \\ 1 & 2 \end{bmatrix} = A\vec{x}(t).$$

We know that the solution to this system is given by $\vec{x}(t) = e^{At}\vec{x}_0$. By solving the characteristic polynomial $\det(A - \lambda I) = 0$, we find that the eigenvalues of A are $\lambda_1 = 4, \lambda_2 = 3$ with corresponding eigenvectors $\vec{v}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Using the diagonalization $A = Y\Lambda Y^{-1}$, we have

$$\begin{aligned} \vec{x}(t) &= e^{At}\vec{x}_0 \\ &= Ye^{\Lambda t}Y^{-1}\vec{x}_0 \\ &= \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{4t} & 0 \\ 0 & e^{3t} \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 100 \\ 50 \end{bmatrix} \\ &= \begin{bmatrix} 2e^{4t} - e^{3t} & -2e^{4t} + 2e^{3t} \\ e^{4t} - e^{3t} & -e^{4t} + 2e^{3t} \end{bmatrix} \begin{bmatrix} 100 \\ 50 \end{bmatrix} \\ &= \begin{bmatrix} 100e^{4t} \\ 50e^{4t} \end{bmatrix} \end{aligned}$$

(b) There are many choices of coefficients that result in a steady equilibrium population. In general, a coefficient matrix A with one zero and one negative eigenvalue will result

in convergence to a non-zero equilibrium. one such matrix is

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -6 \end{bmatrix},$$

which produce the solution

$$\vec{x}(t) = \begin{bmatrix} 60 + 40e^{-5t} \\ 10 + 40e^{-5t} \end{bmatrix},$$

which approaches $\begin{bmatrix} 60 \\ 10 \end{bmatrix}$ as $t \to \infty$.

Exercise 10.7

(a) If $P^2 = P$, show that

$$e^P \approx I + 1.718P$$

(b) Convert the equation below to a matrix equation and then by using the exponential matrix find the solution in terms of $y(0)$ and $y'(0)$:

$$y'' = 0.$$

(c) Show that $e^{A+B} = e^A e^B$ is not generally true for matrices.

Answer for Exercise 10.7

Exercise 10.8

This exercise is built on the Heat Equation, which was introduced as an example in early chapters of the book. We look at the matrix

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

(a) Find the eigenvalues of A and the corresponding eigenvectors. A is symmetric, so the eigenvectors should be orthogonal. Check that this is the case.

(b) Give the algorithm of the power method that can be used to find the largest eigenvalue (in absolute value) of A.

(c) Execute this algorithm, either by hand or with matlab. As initial vector take one of the unit vectors, produce a figure showing the computed approximations as a function of iteration step. Repeat with an initial vector equal to an eigenvector corresponding to either of the two smallest eigenvalues. Observe that the algorithm still converges. but slower than before (round-off error accumulation helps in this case as we discussed in class).

Note: The matrix is well-conditioned and the round-off error will not accumulate particularly fast. Make sure you force the iteration to run for a while (we have found at least 60 iterations works) because you are only testing the relative error in successive iterates of the eigenvalues.

(d) Instead of looking at this 3×3 matrix, take 10×10 (or larger if you like) matrix with the same tridiagonal structure (-2 on the main diagonal and 1 on the subdiagonals). Find all eigenvalues of this larger matrix using the QR iterations. Check your answers against the eigenvalues computed by matlab using the `eig` command. Again, motivate your convergence criterion.

(e) Since the heat equation discretization was an example of numerically solving a differential equation, the plotting of solutions was rather important. Since we just computed eigenvectors, we might be interested to see what the eigenvectors look like when plotted. Consider 102×102 discretization for the heat equation so that the matrix A is 100×100. Use the `eig` command in matlab to obtain the eigenvectors $\vec{v}^{(1)}, \ldots, \vec{v}^{(100)}$ for $-A$ (this is $-A$ so there are positive 2s on the diagonal). For $j = 1, \ldots, 5$, plot $\sqrt{101}\vec{v}^{(j)}$ as we have done in previous homeworks (remember the boundary values).

Now verify for yourself that the functions $u_j(t) = \sqrt{2}\sin(j\pi t)$ for $j = 1, 2, 3, \ldots$ satisfy the differential equation,

$$-\frac{d^2u_j}{dt^2} = \lambda u_j, \quad \lambda = (j\pi)^2, \quad u_j(0) = u_j(1) = 0, \quad \int_0^1 u_j(t)^2 dt = 1.$$

How do the eigenvectors $\vec{v}^{(j)}$ of $-A$ compare to the "eigenfunctions" $u_j(t)$ of the differential equation? What significance does the factor $\sqrt{101}$ have when plotting $\sqrt{101}\vec{v}^{(j)}$ (think about a Riemann sum approximation of the above integral)?

Answer for Exercise 10.8

(a) Using the following code we find the eigenvalues d eigenvectors v and check that the inner product between the different eigenvectors are 0.

```
n=3;
A = -2*diag(ones(n,1)) + diag(ones(n-1,1),-1) + diag(ones(n-1,1),1);
[v,d] = eig(A)
v*v'


v =

    0.5000   -0.7071   -0.5000
   -0.7071    0.0000   -0.7071
    0.5000    0.7071   -0.5000


d =

   -3.4142         0         0
         0   -2.0000         0
         0         0   -0.5858


ans =

    1.0000    0.0000    0.0000
    0.0000    1.0000    0.0000
    0.0000    0.0000    1.0000
```

(b) (i) k = 0 : Set v_0 so that $\|v_0\|= 1$ and λ_0 to an arbitrary quantity.

(ii) Compute $v_{k+1} = \frac{Av_k}{\|av_k\|_2}$ and $\lambda_{k+1} = v_k^T A v_k$.

(iii) If $\frac{|\lambda_{k+1}-\lambda_k|}{|\lambda_k|} < \epsilon$, stop. Otherwise repeat step 2.

(c) The following code can be used

```
A = -2*diag(ones(3,1)) + diag(ones(2,1),-1) + diag(ones(2,1),1);
```

```
[v,d] = eig(A); d=diag(d);
v1 = v(:,2); v2 = A*v1;
lam1 = 1; lam2 = v1*v2;
v2 = v2/norm(v2);
err = abs(lam2-lam1)/abs(lam1);
i=1;
while abs(lam2-lam1)/abs(lam1) > 10^-6 || i<60
v1 = v2; v2 = A*v1;
lam1 = lam2; lam2 = v1*v2;
v2 = v2/norm(v2);
i=i+1;
err(i) = abs(lam2-lam1)/abs(lam1);
end
```

The algorithm takes 83 iterations to converge to $\lambda_{max} = 3.4142$. With an relative error tolerance of 10^{-6}, the roundoff error takes a minimum of 58 iterations to accumulate so that the algorithm can then run on its own without being forced to continue.

Starting with the third eigenvector only takes 27 iterations to converge.
Forcing the iteration to continue seems rather arbitrary. It is also exploiting knowledge about the problem. A better way to generate convergence is via a perturbation of the eigenvector. Convergence to one of the minor eigenvectors is not stable so if we detect convergence, we should perturb our solution vector and see if we come back to the equilibrium. This would to prevent us from having to wait for roundoff error to work in our favor. This problem was really to just see how roundoff error can work in our favor.

(d) The QR iteration:

```
function [A,k] = qrmeth(A)
      k = 1;  %initialize counter
      tol = 1e-5;
              %set tolerance
      while 1<3
         [Q,R] = qr(A);
         A = R*Q;
              %get the next A
         if(norm(Q*A-A*Q) < tol)
              %convergence condition is when Q*A is approx to A*Q
              %as this would mean A is approximately diagonal
             break
end
```

```
k = k+1; end
end
```

Applying it to our matrix:

```
clear all
  close all
  m = 10; %setting dimension of A
  A = diag(-2*ones(m,1)) + diag(ones(m-1,1),1) + diag(ones(m-1,1),-1);
          %creating A
  [D,k] = qrmeth(A);
          %find e-vals of D and the number of iterations
  V = eig(A)
          %the e-vals of A using eig
  lam = diag(D)
          %extracts diagonal from D
  diff = lam-V
          %difference between these two e-vals
```

The results:

```
V=
     -3.9190
     -3.6825
     -3.3097
     -2.8308
      -2.2846
     -1.7154
     -1.1692
     -0.6903
     -0.3175
     -0.0810


lam =
     -3.9190
     -3.6825
     -3.3097
     -2.8308
     -2.2846
     -1.7154
     -1.1692
     -0.6903
```

```
    -0.3175
    -0.0810
 diff =
    1.0e-07 *
     0.9251
    -0.9250
    -0.0001
     0.0000
    -0.0000
    -0.0000
    -0.0000
    -0.0000
     0.0000
    -0.0000
```

(e) Here are the code and plots for the eigenfunctions:

```
n=100;
 A = -2*eye(n) + diag(ones(n-1,1),-1) + diag(ones(n-1,1),1);
 [v,d] = eig(-A);
 d = diag(d);
 t = linspace(0,1,n+2);

 k=5;
 figure
 subplot(2,1,1)


 plot(t,sqrt(n+1)*[zeros(1,k);v(:,1:k);zeros(1,k)]*diag(sign(v(1,1:k))))
title(Discretized Eigenvectors)
legend(j=1,j=2,j=3,j=4,j=5);
subplot(2,1,2)
plot(t,sqrt(2)*sin(pi*t*(1:k)))
title(ODE Eigenfunctions)
```

The eigenvectors are discretized versions of the eigenfunctions. This makes sense because the Poisson matrix was supposed to represent a discretization of the second derivative. Note also that we have normalized the eigenvectors as $\|v_j\|_2 = 1$. The factor $\sqrt{101}$ serves to put the eigenvector on the same scale as the eigenfunctions:

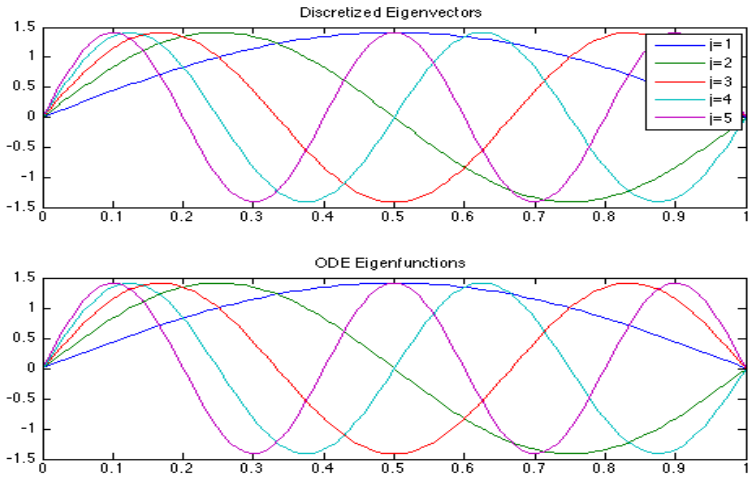

Figure 10.1: Plot of discretized and exact eigenvectors.

$$1 = \int_0^1 u_j(t)^2 dt \approx \frac{1}{n+1}\sum_{i=1}^{n+1} u_j(t_i)^2 \approx \frac{1}{n+1}\sum_{i=1}^{n}(\sqrt{101}v_j(t_i))^2 = \sum_{i=1}^{n+1} v_j(t_i)^2 = 1$$

Exercise 10.9

Consider the matrix

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

(a) For which values(s) of α is A singular, and what is the rank of A

(b) For $\alpha = -1$, prove, without computing the eigenvalues and eigenvectors of A, that A is diagonalizable and that the eigenvectors of A are orthogonal.

(c) For $\alpha = -1$, give the orthogonal matrix Q and diagonal matrix Λ for which $A = Q\Lambda Q^{-1}$.

(d) Prove that for any α, the general solution $\frac{d\vec{u}}{dt} = A\vec{u}$ with $\vec{u}(0) = \vec{u}_0$ is given by $\vec{u} = e^{At}\vec{u}_0$, with the exponential matrix defined as $e^{At} = I + tA + \frac{1}{2}(tA)^2 + \frac{1}{3!}(tA)^3 + \ldots$.

(e) For $\alpha = -1$, what is the behavior of the solution to $\frac{d\vec{u}}{dt} = A\vec{u}$ with $\vec{u}(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as $t \to \infty$?

Answer for Exercise 10.9

Exercise 10.10

In this problem we consider a small mass-spring system, consisting of 2 masses that are connected to each other and to fixed walls by three identical springs as shown in the figure.

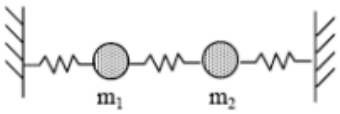

Figure 10.2: Mass-Spring System.

At time $t = 0$, we give the first mass m_1 a small displacement. As a result the mass-spring system will start to move in an oscillatory fashion. If the masses are both equal to 1, the system of equation that describes the displacement of the masses is given by

$$\frac{d^2\vec{u}}{dt^2} = A\vec{u},$$

where

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix},$$

with u_1 and u_2 are the displacements of m_1 and m_2, respectively, and

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

We take the initial conditions (we need two of them for a second order differential equation) to be

$$\vec{u}(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \frac{d\vec{u}}{dt}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

(a) Find the matrices Y and Λ such that $A = Y\Lambda Y^{-1}$, with Y orthogonal, and Λ diagonal.

(b) Using this decomposition of A, show that we can transform the system of differential equations to

$$\frac{d^2\vec{z}}{dt^2} = \Lambda\vec{z}, \quad \vec{z} = Y^T\vec{u}.$$

The system for $\vec{z}$ is now "decoupled" so you can solve each component individually since they do not depend on each other. Solve the system of equations for $\vec{z}$ and from it find the displacements u_1 as a function of time t.

(c) If the masses are not equal, the system of differential equations describing the motion of the mass-spring system is instead given by

$$M\frac{d^2\vec{u}}{dt^2} = A\vec{u}, \quad M = \begin{bmatrix} m_1 & 0 \\ 0 & m_2 \end{bmatrix}. \tag{10.1}$$

We guess solution will be of the form $\vec{u}(t) = e^{i\omega t}\vec{v}$, where ω and $\vec{v}$ are quantities determined by M and A. Plug in $\vec{u}(t) = e^{i\omega t}\vec{v}$ and show that finding ω and $\vec{v}$ corresponds to the so-called "generalized eigenvalue problem"

$$A\vec{v} = \lambda M\vec{v},$$

with $\lambda = (i\omega)^2$.

Now set $m_1 = 1$ and $m_2 = 2$. The characteristic polynomial for the generalized eigenvalue problem is given by $\det(A - \lambda M) = 0$. Solve the characteristic polynomial to find the two generalized eigenvalues λ_1, λ_2 (or if you wish, the two values of ω since $\lambda = (i\omega)^2$) and the two generalized eigenvectors $\vec{v}_1, \vec{v}_2$. Then give the solution to the differential equation in (10.1).

Answer for Exercise 10.10

(a)

$$Y = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}, \qquad \Lambda = \begin{pmatrix} -3 & 0 \\ 0 & 1 \end{pmatrix}$$

(b) Let's make the change of variables $\vec{z} = Y^T\vec{u}$. Then the ODE becomes

$$\frac{d^2\vec{u}}{dt^2} = Y\Lambda\vec{z}.$$

If we pre-multiply by Y^T, we get $\Lambda\vec{z}$ on the RHS so we have to see if in fact

$$Y^T\frac{d^2\vec{u}}{dt^2} = \frac{d^2Y^T\vec{u}}{dt^2} = \frac{d^2\vec{z}}{dt^2}$$

If you know multivariable calculus, this is well known to you. Otherwise, here is how the statement is verified:

$$Y^T\frac{d^2\vec{u}}{dt^2} = Y^T\begin{bmatrix} \frac{d^2\vec{u}_1}{dt^2} \\ \frac{d^2\vec{u}_2}{dt^2} \end{bmatrix} = \begin{bmatrix} y_{11}\frac{d^2\vec{u}_1}{dt^2} & y_{21}\frac{d^2\vec{u}_2}{dt^2} \\ y_{12}\frac{d^2\vec{u}_1}{dt^2} & y_{22}\frac{d^2\vec{u}_2}{dt^2} \end{bmatrix}$$
$$= \begin{bmatrix} y_{11}u_1(t) & y_{21}u_2(t) \\ y_{12}u_1(t) & y_{22}u_2(t) \end{bmatrix} = \frac{d^2Y^T\vec{u}}{dt^2} = \frac{d^2\vec{z}}{dt^2}$$

With the transformed system, we now have two decoupled ODEs:

$$\begin{array}{llll} \frac{d^2z_1}{dt^2} = -3z_1, & z_1(0) = \frac{1}{\sqrt{2}}, & z_1'(0) = 0, \\ \frac{d^2z_2}{dt^2} = -z_1, & z_2(0) = \frac{1}{\sqrt{2}}, & z_2'(0) = 0. \end{array}$$

The solutions to these ODEs give us

$$\vec{z} = \begin{bmatrix} \frac{1}{\sqrt{2}}\cos(\sqrt{3}t) \\ \frac{1}{\sqrt{2}}\cos(t) \end{bmatrix}$$

Converting back to $\vec{u}$ gives

$$\vec{u} = Y\vec{z} = \begin{bmatrix} \frac{1}{2}\cos(t) + \frac{1}{2}\cos(\sqrt{3}t) \\ \frac{1}{2}\cos(t) - \frac{1}{2}\cos(\sqrt{3}t) \end{bmatrix}$$

(c) If $\vec{u} = e^{i\omega t}\vec{v}$, then

$$e^{i\omega t}A\vec{v} = M\frac{d^2\vec{u}}{dt^2} = (i\omega)^2e^{i\omega t}M\vec{v}.$$

With $\lambda = (i\omega)^2$, we divide out $e^{i\omega t}$ to get $A\vec{v} = \lambda M\vec{v}$, as desired.

For $m_1 = 1$ and $m_2 = 2$, the characteristic polynomial is

$$2\lambda^2 + 6\lambda + 3 = 0,$$

with solutions $\lambda = -\frac{3}{2} \pm \frac{\sqrt{3}}{2}$. Solving for the eigenvectors gives

$$\lambda_1 = -\frac{3}{2} - \frac{\sqrt{3}}{2}, \quad \vec{v}_1 = \begin{bmatrix} 1+\sqrt{3} \\ -1 \end{bmatrix}, \lambda_2 = -\frac{3}{2} + \frac{\sqrt{3}}{2}, \quad \vec{v}_2 = \begin{bmatrix} 1-\sqrt{3} \\ -1 \end{bmatrix}$$

We know that $\lambda = (i\omega)^2 = -\omega^2$ which means that $\omega = \pm\sqrt{-\lambda}$ since λ is negative. Our solution thus has the form

$$\vec{u}(t) = \vec{v}_1(c_1 e^{i\sqrt{-\lambda_1}t} + c_2 e^{-i\sqrt{-\lambda_1}t}) + \vec{v}_2(c_3 e^{i\sqrt{-\lambda_2}t} + c_4 e^{-i\sqrt{-\lambda_2}t}).$$

If we use the initial conditions we get the conditions

$$\begin{aligned} \vec{u}(0) &= \vec{v}_1(c_1 + c_2) + \vec{v}_2(c_3 + c_4) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \\ \vec{u}'(0) &= \vec{v}_1(c_1 - c_2) + \vec{v}_2(c_3 - c_4) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \end{aligned}$$

Since the eigenvectors are linearly independent we get from the second equation that $c_1 = c_2$ and $c_3 = c_4$. Plugging this into the first equation gives us the equation system

$$\vec{u}(0) = \begin{bmatrix} 1+\sqrt{3} & 1-\sqrt{3} \\ -1 & -1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}.$$

This has the solutions $c_1 = \frac{\sqrt{3}}{12}$ and $c_3 = -\frac{\sqrt{3}}{12}$. If we now use that $\cos(x) = \frac{e^{ix}+e^{-ix}}{2}$ we get that the solution is

$$\begin{aligned} &\vec{u}(t) = 2\vec{v}_1 c_1 \cos(\sqrt{-\lambda_1}t) + 2\vec{v}_2 c_3 \cos(\sqrt{-\lambda_2}t) \\ &= \frac{\sqrt{3}}{6} \begin{bmatrix} (1+\sqrt{3})\cos\left(\left(\frac{3}{2} + \frac{\sqrt{3}}{2}\right)t\right) - (1-\sqrt{3})\cos\left(\left(\frac{3}{2} - \frac{\sqrt{3}}{2}\right)t\right) \\ \cos\left(\left(\frac{3}{2} - \frac{\sqrt{3}}{2}\right)t\right) - \cos\left(\left(\frac{3}{2} + \frac{\sqrt{3}}{2}\right)t\right) \end{bmatrix} \end{aligned}$$

As a last comment we should note the following:

$$A\vec{v} = \lambda M\vec{v} \quad \Leftrightarrow \quad M^{-1}A\vec{v} = \lambda\vec{v}.$$

That is, those generalized eigenvalues and eigenvectors are of the plain variety for the matrix $M^{-1}A$. Now, we dont typically use $M^{-1}A$ because M might be difficult to invert or not even invertible, hence why mathematicians came up with the generalized eigenvalue problem. This happened to not be the case here, though. However, recognizing this equivalence we could have used our diagonalization technique from part (b) for this problem.

Exercise 10.11

We revisit the heat equation again. Let's consider

$$\frac{d^2T}{dx^2} = 0, \text{ for } 0 \leq x \leq 1, \text{ with } T(0) = 1, T(1) = 2.$$

We discretize the equation in the points $x_i = ih$, with $i = 0, 1, \ldots, N$ and $N = \frac{1}{h}$.

(a) Take N = 4. Give the discrete system of equations for the given boundary conditions and the standard second order discretization in the form $A\vec{T} = \vec{b}$, with $\vec{T} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \end{bmatrix}$.

(b) Show that the discretization used in (a) is indeed second order.

*(c) One of the eigenvalues of the matrix A found in (a) is close to -0.5. Give the algorithm of the shifted inverse power method that can be used to find an approximation to this eigenvalue, with an appropriate convergence criterion. You do not have to execute the algorithm.

*(d) What is the convergence rate of the algorithm given in part (d)? Motivate your answer clearly.

Answer for Exercise 10.11

Exercise 10.12

Consider the matrix

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

(a) For which values(s) of α is A singular, and what is the rank of A

(b) For $\alpha = -1$, prove, without computing the eigenvalues and eigenvectors of A, that A is diagonalizable and that the eigenvectors of A are orthogonal.

(c) For $\alpha = -1$, give the orthogonal matrix Q and diagonal matrix Λ for which $A = Q\Lambda Q^{-1}$.

(d) Prove that for any α, the general solution $\frac{d\vec{u}}{dt} = A\vec{u}$ with $\vec{u}(0) = \vec{u}_0$ is given by $\vec{u} = e^{At}\vec{u}_0$, with the exponential matrix defined as $e^{At} = I + tA + \frac{1}{2}(tA)^2 + \frac{1}{3!}(tA)^3 + \ldots$.

(e) For $\alpha = -1$, what is the behavior of the solution to $\frac{d\vec{u}}{dt} = A\vec{u}$ with $\vec{u}(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as $t \to \infty$?

Answer for Exercise 10.12

(a) We compute $\det(A) = -2(4+\alpha) + -2\alpha = -4\alpha - 8$. Recall that A is singular iff $\det(A) = 0$. By setting $0 = \det(A) = -4\alpha - 8$, we have that A is singular iff $\alpha = -2$.

(b) For $\alpha = -1$, A is symmetric. Using Schur form, we show in Section **??** that A is diagonalizable and the eigenvectors of A are orthogonal.

(c) We first find eigenvalues using equation $\det(A - \lambda I) = 0$. This gives the eigenvalues $\lambda = -2, -2+\sqrt{2}$, and $-2-\sqrt{2}$. For each λ, find corresponding eigenvector by computing $\vec{y}$ such that $(A - \lambda I)\vec{y} = \vec{0}$. Since A is diagonalizable, we find that each nullspace has dimension 1. We can now contruct $Q = [\vec{y}_1, \vec{y}_2, \vec{y}_3]$, where $\vec{y}_i$ is eigenvector (normalized to have norm 1) correspond to eigenvalue λ_i, and $\Lambda = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2-\sqrt{2} & 0 \\ 0 & 0 & -2+\sqrt{2} \end{bmatrix}$.
Note: QR decomposition can NOT be used to find eigenvectors or eigenvalues, and the LU factorization also has nothing to do with eigenvalues and eigenvectors.

(d) To show that $e^{At}\vec{u}_0$ gives the solution, plug it into the differential equation and show the equation holds. In particular,

$$\frac{d}{dt}\left(e^{At}\vec{u}_0\right) = A\left(e^{At}\vec{u}_0\right).$$

Also check that the initial condition is satisfied, $e^{A\times 0}\vec{u}_0 = \vec{u}_0$.

(e) From (c), we know that the eigenvalues of A for this α are all negative. We also know that A is diagonalizable. So, the solution can be written as $\vec{u}(t) = Qe^{\Lambda t}Q^{-1}\vec{u}_0$, where Λ is the diagonal matrices containing the eigenvalues. Since the eigenvalues are negative, the solution will go to 0 as $t \to \infty$.

* Exercise 10.13

(a) Show that for an $n \times n$ matrix A, the determinant of A is equal to the product of its eigenvalues. That is, show that $\det(A) = \prod_{i=1}^{n} \lambda_i$

(b) Let A be a matrix where for each column, the sum of the elements is equal to a constant, say, μ. Show that μ must be an eigenvalue of A

Answer for Exercise 10.13

Exercise 10.14

Consider the matrix

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

(a) For which values(s) of α is A singular, and what is the rank of A

(b) For $\alpha = -1$, prove, without computing the eigenvalues and eigenvectors of A, that A is diagonalizable and that the eigenvectors of A are orthogonal.

(c) For $\alpha = -1$, give the orthogonal matrix Q and diagonal matrix Λ for which $A = Q\Lambda Q^{-1}$.

(d) Prove that for any α, the general solution $\frac{d\vec{u}}{dt} = A\vec{u}$ with $\vec{u}(0) = \vec{u}_0$ is given by $\vec{u} = e^{At}\vec{u}_0$, with the exponential matrix defined as $e^{At} = I + tA + \frac{1}{2}(tA)^2 + \frac{1}{3!}(tA)^3 + \ldots$.

(e) For $\alpha = -1$, what is the behavior of the solution to $\frac{d\vec{u}}{dt} = A\vec{u}$ with $\vec{u}(0) = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ as $t \to \infty$?

Answer for Exercise 10.14

(a) We compute $\det(A) = -2(4 + \alpha) + -2\alpha = -4\alpha - 8$. Recall that A is singular iff $\det(A) = 0$. By setting $0 = \det(A) = -4\alpha - 8$, we have that A is singular iff $\alpha = -2$.

(b) For $\alpha = -1$, A is symmetric. Using Schur form, we show in Section **??** that A is diagonalizable and the eigenvectors of A are orthogonal.

(c) We first find eigenvalues using equation $\det(A-\lambda I) = 0$. This gives the eigenvalues $\lambda = -2, -2+\sqrt{2},$ and $-2-\sqrt{2}$. For each λ, find corresponding eigenvector by computing $\vec{y}$ such that $(A-\lambda I)\vec{y} = \vec{0}$. Since A is diagonalizable, we find that each nullspace has dimension 1. We can now contruct $Q = [\vec{y}_1, \vec{y}_2, \vec{y}_3]$, where $\vec{y}_i$ is eigenvector (normalized to have norm 1) correspond to eigenvalue λ_i, and $\Lambda = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2-\sqrt{2} & 0 \\ 0 & 0 & -2+\sqrt{2} \end{bmatrix}$.
Note: QR decomposition can NOT be used to find eigenvectors or eigenvalues, and the LU factorization also has nothing to do with eigenvalues and eigenvectors.

(d) To show that $e^{At}\vec{u}_0$ gives the solution, plug it into the differential equation and show the equation holds. In particular,

$$\frac{d}{dt}\left(e^{At}\vec{u}_0\right) = A\left(e^{At}\vec{u}_0\right).$$

Also check that the initial condition is satisfied, $e^{A\times 0}\vec{u}_0 = \vec{u}_0$.

(e) From (c), we know that the eigenvalues of A for this α are all negative. We also know that A is diagonalizable. So, the solution can be written as $\vec{u}(t) = Qe^{\Lambda t}Q^{-1}\vec{u}_0$, where Λ is the diagonal matrices containing the eigenvalues. Since the eigenvalues are negative, the solution will go to 0 as $t \to \infty$.

Exercise 10.15

We will consider the ordinary differential equations $\frac{d\vec{x}}{dt} = A\vec{x}$, with $A = \frac{1}{h^2}\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$. Here, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ is the vector of unknown functions. The initial condition is given by $\vec{x}(t=0) = \vec{x}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

(a) Show that A is diagonalizable, and that its eigenvalue are equal to -9 and -27. Give the matrices Y and Λ for which $A = Y\Lambda Y^{-1}$.

(b) Find the solution to the differential equation at time $t = 1$.

(c) Show that the solution to the differential equation converges as $t \to \infty$. What is this converged solution, and what is the associated convergence rate?

Answer for Exercise 10.15

---

* Exercise 10.16

Suppose that the $n \times n$ matrix A is symmetric, has only positive eigenvalues, and that its two largest eigenvalues are identical. Can the Power method still be used to compute the value of these two largest eigenvalues, and if so, what would the convergence rate be? Can the Power method be used to find the corresponding eigenvectors?

Answer for Exercise 10.16

Set

$$\vec{x}^{(0)} = \alpha_1\vec{y}_1 + \alpha_2\vec{y}_2 + \ldots + \alpha_n\vec{y}_n$$

with $A\vec{y}_i = \lambda_i\vec{y}_i$ and $\lambda_1 = \lambda_2 > \lambda_3 \geq \lambda_4 \geq \ldots \geq \lambda_n > 0$. Since $A = A^T$, A is diagonalizable and all $\vec{y}_i$'s are independent. Then

$$A^k\vec{x}^{(0)} = \alpha_1\lambda_1^k\vec{y}_1 + \alpha_2\lambda_2^k\vec{y}_2 + \alpha_3\lambda_3^k\vec{y}_3 + \ldots + \alpha_n\lambda_n^k\vec{y}_n$$

So as $k \to \infty$,

$$A^k\vec{x}^{(0)} \to \alpha_1\lambda_1^k\vec{y}_1 + \alpha_2\lambda_2^k\vec{y}_2 = \lambda_1^k\left(\alpha_1\vec{y}_1 + \alpha_2\vec{y}_2\right)$$

This means that in the Power method, we converge to a linear combination of $\vec{y}_1$ and $\vec{y}_2$, which depends on α_1 and α_2 only (which in turn depends on choice of $\vec{x}^{(0)}$). In fact, it is $\frac{\alpha_1\vec{y}_1+\alpha_2\vec{y}_2}{\|\alpha_1\vec{y}_1+\alpha_2\vec{y}_2\|}$. Call this $\vec{z}$.

We will be able to find λ_1 $(= \lambda_2)$ since

$$\begin{aligned}
\vec{z}^T A\vec{z} &= \frac{1}{\|\alpha_1\vec{y}_1 + \alpha_2\vec{y}_2\|^2}\left(\alpha_1\vec{y}_1 + \alpha_2\vec{y}_2\right)^T\left(\alpha_1\lambda_1\vec{y}_1 + \alpha_2\lambda_2\vec{y}_2\right) \\
&= \lambda_1 \cdot \frac{1}{\|\alpha_1\vec{y}_1 + \alpha_2\vec{y}_2\|^2}\left(\alpha_1\vec{y}_1 + \alpha_2\vec{y}_2\right)^T\left(\alpha_1\vec{y}_1 + \alpha_2\vec{y}_2\right) \\
&= \lambda_1
\end{aligned}$$

The convergence rate is now given by $\left|\frac{\lambda_3}{\lambda_1}\right|$.

For each choice of $\vec{x}^{(0)}$, we converge to some linear combination of $\vec{y}_1$ and $\vec{y}_2$. For $A = A^T$, we want to find these eigenvectors such that they are orthogonal.

If we apply the Power method for two $\vec{x}^{(0)}$'s, and check to see the converged linear combinations of $\vec{y}_1$ and $\vec{y}_2$ are independent, then we have found the space that $\vec{y}_1$ and $\vec{y}_2$ span.

Finding them then means finding an orthonormal basis for this space that is also orthogonal to the other eigenvectors.

Exercise 10.17

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement is false it is sufficient to give one counter example. To show a statement is true, provide a general proof.

(a) If the $n \times n$ matrices A and B are similar, in other words, if there is a $n \times n$ invertible matrix T such that $A = TBT^{-1}$, then A and B have the same eigenvalues.

(b) If Q is an orthogonal matrix, then $Q + \frac{1}{2}I$ is invertible.

(c) If the $n \times n$ matrices A and B have exactly the same eigenvalues and eigenvectors, then $A = B$.

*(d) The eigenvalues of a symmetric and real $n \times n$ matrix A are always real.

Answer for Exercise 10.17

Exercise 10.18

Consider the matrix

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

(a) For which value of α is pivoting required when computing the LU decomposition of A? For this value, give the permutation matrix P and the final L and U in the equation $PA = LU$.

(b) Take $\alpha = 0$. Use Gram-Schmidt orthogonalization to find an orthonormal basis for the column space of A.

For part (c), (d), and (e), take $\alpha = 1$.

(c) Explain clearly, why A is diagonalizable and give its canonical transform (in other words, give the matrix Y and diagonal matrix Λ for which $A = Y\Lambda Y^{-1}$). Show that the eigenvectors are orthogonal. Normalize them so they have unit length and $Y^{-1} = Y^T$.

(d) Which of the eigenvectors found in part (c) form a basis for the nullspace of A, and which a basis for the row space of A? Explain your answers clearly.

*(e) Use the Cayley-Hamilton theorem to find e^A. Compare the result to e^A obtained with the formula for the matrix exponential we derived in class.

Answer for Exercise 10.18

(a) Pivoting is required when we encounter a zero pivot during GE process. This can happen here if in the first step of GE (while zeroing out a_{21}), a_{22} also becomes 0. So, we have

$$-2 - \left(\frac{\alpha}{-1}\right) \times 1 = 0, \quad \text{or } \alpha = 2$$

For $\alpha = 2$, we know that we have to, at some point, swap rows 2 and 3. So, let's do this from the start. We look at PA with $P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$. This is

$$PA = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & -1 \\ 2 & -1 & 1 \end{bmatrix}.$$

Do GE

$$\begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & -1 \\ 2 & -1 & 1 \end{bmatrix} \xrightarrow{GE} \begin{bmatrix} -1 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} = U, \quad \text{with } L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$$

Note: if we pivot during GE process and not before, we have

$$U = PCA, \text{ with } C = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

or $A = C^{-1}P^{-1}U$. And we do not get $PA = LU$ as required.

(b) $A = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -2 & 1 \\ 0 & 1 & -1 \end{bmatrix}$. Then the Gram-Schmidt process is the following.

$$\begin{aligned}
\underline{\textit{Step 1:}} \qquad & \vec{u}_1 = \frac{\vec{a}_1}{\|\vec{a}_1\|} = \vec{a}_1. \\
\underline{\textit{Step 2:}} \qquad & \vec{w}_2 = \vec{a}_2 - \left(\vec{a}_2^T \vec{u}_1\right) \vec{u}_1 \\
& \quad = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} - (-1) \begin{bmatrix} -1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \\
& \vec{u}_2 = \frac{\vec{w}_2}{\|\vec{w}_2\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \\
\underline{\textit{Step 3:}} \qquad & \vec{w}_3 = \vec{a}_3 - \left(\vec{a}_3^T \vec{u}_1\right) \vec{u}_1 - \left(\vec{a}_3^T \vec{u}_3\right) \vec{u}_2 \\
& \quad = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} - 0 \cdot \vec{u}_1 - \left(\frac{-3}{\sqrt{5}}\right) \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} \\
& \quad = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} + \frac{3}{5} \begin{bmatrix} 0 \\ -2 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1/5 \\ -2/5 \end{bmatrix} \\
& \vec{u}_3 = \frac{\vec{w}_3}{\|\vec{w}_3\|} = \frac{1}{\sqrt{5}} \begin{bmatrix} 0 \\ -1 \\ -2 \end{bmatrix}
\end{aligned}$$

Orthonormal basis is given by $\vec{u}_1, \vec{u}_2,$ and $\vec{u}_3$.

(c) For $\alpha = 1$, $A = A^T$. From Schur, $A = TRT^{-1}$ for unitary T with R upper triangular. Then $A = A^T$ gives

$$TRT^{-1} = TR^T T^{-1}, \quad \text{or } R = R^T.$$

Thus R is diagonal. Hence, $A = TRT^{-1}$ is the canonical transform of A with R the diagonal matrix containing eigenvalues of A, and T is the matrix containing eigenvectors. Since T is unitary, we also know that A has orthogonal eigenvectors.
To find T and R (or in other notation Y and Λ), first we find eigenvalues by solving

$$\det(A - \lambda I) = 0$$

This gives $\lambda_i = -3, -1, 0$. For eigenvectors, we solve $(A - \lambda_i)\vec{y}_i = \vec{0}$ for $i = 1, 2, 3$.

Then the eigenvectors are $\vec{y}_i = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\-2\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\-1\end{bmatrix}, \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$. So

$$Y = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ -2/\sqrt{6} & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}, \Lambda = \begin{bmatrix} -3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

(d) $\mathcal{N}(A)$ is spanned by $\vec{y}_3$ (since $\lambda_3 = 0$). We know that $A\vec{y}_3 = \vec{0}$, so $\vec{y}_3$ orthogonal to all rows of A. We know that $\mathcal{R}(A)$ and $\mathcal{N}(A)$ are orthogonal complements. And finally we know that $\vec{y}_1$ and $\vec{y}_2$ are both orthogonal to $\vec{y}_3$. So, $\vec{y}_1$ and $\vec{y}_2$ form an orthogonal basis for $\mathcal{R}(A)$.

*(e) Cayley-Hamilton:

$$e^A = \alpha_0 I + \alpha_1 A + \alpha_2 A^2$$

for $\alpha_1, \alpha_2, \alpha_3$ constant because A^k for $k \leq n = 3$ can be expressed as linear combinations of $I, A,$ and A^2. We also know that the eigenvalues must satisfy same equation:

$$\begin{aligned} e^{-3} &= \alpha_0 - 3\alpha_1 + 3\alpha_2 \\ e^{-1} &= \alpha_0 - \alpha_1 + \alpha_2 \\ e^0 &= \alpha_0 \end{aligned}$$

Solving the system above, we get $\alpha_0 = 1, \alpha_1 = \frac{4}{3} - \frac{3}{2}e^{-1} + \frac{1}{6}e^{-3}, \alpha_2 = \frac{1}{3} - \frac{1}{2}e^{-1} + \frac{1}{6}e^{-3}$. So we expect

$$e^A = I + \left(\frac{4}{3} - \frac{3}{2}e^{-1} + \frac{1}{6}e^{-3}\right)\begin{bmatrix} -1 & 1 & 0 \\ 1 & -2 & 1 \\ 0 & 1 & -1 \end{bmatrix} + \left(\frac{1}{3} - \frac{1}{2}e^{-1} + \frac{1}{6}e^{-3}\right)\begin{bmatrix} 2 & -3 & 1 \\ -3 & 5 & -3 \\ 1 & -3 & 2 \end{bmatrix}$$

Using the matrix exponential, we have

$$e^A = Ye^{\Lambda}Y^{-1} = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \\ -2/\sqrt{6} & 0 & 1/\sqrt{3} \\ 1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix}\begin{bmatrix} e^{-3} & 0 & 0 \\ 0 & e^{-1} & 0 \\ 0 & 0 & e^0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{6} & -2/\sqrt{6} & 1/\sqrt{6} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix}$$

which is equal to the result from Cayley-Halmilton process.

Exercise 10.19

(a) Consider the matrix $A = \begin{bmatrix} 3 & \alpha \\ \alpha & 2 \end{bmatrix}$. We are interested in solving $A\vec{x} = \vec{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. We do this with the Gauss-Seidel iteration. Let $\rho(C)$ be the spectral radius of the amplification matrix C. The spectral radius is defined by the largest eigenvalue (in absolute value) of C. Here, $C = M^{-1}N$, with $A = M - N$. From the General Convergence Theorem, we know that the Gauss-Seidel iteration will converge if $\rho(C) < 1$. Find the values of α for which this is so.

(b) We are now interested in using a new stationary method for solving $A\vec{x} = vb$ for some $n \times n$ matrix A. As suggested by one of the students in class, an interesting stationary method could be formulated with splitting $A = M - N$, where $M = I$ and $N = I - A$. This splitting leads to the iterative scheme

$$\vec{x}^{(k+1)} = (I - A)\vec{x}^{(k)} + \vec{b}$$

Why is this iterative method attractive from a computational point of view, provided of course that the scheme converges? What are the restrictions on the eigenvalues of A for the scheme to converge? Motivate your answers clearly.

Answer for Exercise 10.19

Chapter 11

Exercises & Solutions

Exercise 11.1

The unfixed page rank equation is $P\vec{x} = \vec{x}$. We reasoned that this indicates that $\vec{x}$ is the eigenvector of P corresponding to $\lambda = 1$. We also know that $\lambda = 1$ is the largest eigenvalue. For the P in the wee internet example, implement the Power iteration to find $\vec{x}$. Run it for various initial guesses. What do you observe? Explain the behavior. Hint: compute the eigenvalues of P.

Answer for Exercise 11.1

---

* Exercise 11.2

We use the Jacobi iteration **??** to solve the page rank system $A\vec{x} = \vec{v}$, with $A = I - \alpha P$.

(a) The converged solution is $\vec{x}^{\infty}$. Analyze the distance between $\vec{x}^{\infty}$ and successive iterates of the Jacobi iteration, that is, find the relationship between $\|\vec{x}^{(k+1)} - \vec{x}^{\infty}\|_1$ and $\|\vec{x}^{(k)} - \vec{x}^{\infty}\|_1$. Note that we measure distance in terms of the 1-norm here. Your analysis must hold for any $n \times n$ page rank matrix P. Form your analysis, show that

α must be chosen to be smaller than 1.
(Hint: first compute the 1-norm of P.)

(b) Now, show that the matrix $I - \alpha P$ is invertible for any $n \times n$ page rank matrix P.
(Hint: don't forget that we have $0 < \alpha < 1$.)

Answer for Exercise 11.2

(a) We split the matrix $I - \alpha P = M - N$, where M is the diagonal to $I - \alpha P$. Remember that the diagonal of P has zeros only, so the diagonal of $I - \alpha P$ is exactly $M = I$. And hence $N = M - (I - \alpha P) = \alpha P$. The Jacobi iteration is the following:

Algorithm 2 Jacobi iteration for PageRank computations

$k = 0$
Choose $\vec{x}^{(0)}$ and compute $\vec{r}^{(0)} = \vec{v} - (I - \alpha P)\vec{x}^{(0)}$
while $k < step\_limit$ and $\|\vec{r}^{(k)}\| \geq \epsilon \left(\|I - \alpha P\|\|\vec{x}^{(k)}\| + \|\vec{v}\|\right)$ **do**
  Compute $\vec{x}^{(k+1)} = \alpha P\vec{x}^{(k)} + \vec{v}$
  Compute the residual $\vec{r}^{(k+1)} = \vec{v} - (I - \alpha P)\vec{x}^{(k+1)}$
  Compute $\|\vec{x}^{(k+1)}\|$ and $\|\vec{r}^{(k+1)}\|$
  $k = k + 1$
end while

Note that every element of the new iterate can be computed independently of the others (the algorithm is parallelizable).

(b)

$$\begin{aligned}\|\vec{x}^{(k+1)} - \vec{x}^*\|_1 &= \|\alpha P\vec{x}^{(k)} + \vec{v} - (\alpha P\vec{x}^* + \vec{v})\|_1 \\ &= |\alpha|\|P(\vec{x}^{(k)} - \vec{x}^*)\|_1 \\ &\leq |\alpha|\|P\|_1\|\vec{x}^{(k)} - \vec{x}^*\|_1\end{aligned}$$

Remember that P has non-negative components, and the elements on each column of P add up to 1, indicating $\|P\|_1 = 1$. Also note that $\alpha > 0$.

$$\|\vec{x}^{(k+1)} - \vec{x}^*\|_1 \leq \alpha\|\vec{x}^{(k)} - \vec{x}^*\|_1$$

Note that $0 < \alpha < 1$, the distance between $\vec{x}^*$ and successive iterates decrease to 0.

(c) Consider any vector $\vec{x}$ such that $(I - \alpha P)\vec{x} = \vec{0}$, so $\vec{x} = \alpha P\vec{x}$. By taking the 1-norm on either side, we find

$$\|\vec{x}\|_1 = \|\alpha P\vec{x}\|_1 \leq |\alpha|\|P\|_1\|\vec{x}\|_1 = \alpha\|\vec{x}\|_1 \implies (1 - \alpha)\|\vec{x}\|_1 \leq 0$$

Note that $\|P\|_1 = 1$ and $\alpha > 0$. But with $\alpha < 1$, so $1 - \alpha > 0$. This implies $\|\vec{x}\|_1 \leq 0$. Since $\|\vec{y}\|_1 \geq 0$ for any $\vec{y}$, this means that $\|\vec{x}\|_1 = 0$. This can only be true if $\vec{x} = \vec{0}$. We showed that $\mathcal{N}(I - \alpha P) = \{\vec{0}\}$. So, the matrix $I - \alpha P$ is non-singular.

Exercise 11.3

For this problem, please refer back to Exercise 11.2. The solutions will help you in coding your PageRank algorithm.

Recall the linear system for Page Rank:

$$(I - \alpha P)\vec{x} = \vec{v}$$

where α is the fraction of a pages rank that it propagates to neighbors at each step and $\vec{v}$ the initial rank we give to each page. In our problem, we set $\alpha = 0.85$ and the entries of $\vec{v}$ to be $\vec{v}_i = \frac{1}{n}$, with n the total number of pages in our network.

The PageRank calculation need not only be used for the internet! In fact, PageRank can be calculated for nodes on any connected graph representing any abstraction. For this problem, we will consider a graph of movies connected by links if they share at least one common actor. For this purpose, we provide a matlab date file `movies.mat` that can be downloaded from the Materials $->$ Assignments2013 folder on Coursework. Place this file in a local directory accessible in matlab and type `clear all; load movies.mat`. If you look in the workspace, you should have a collection of new variables, defined as follows:

- `links` : rows are (movie, person) pairs (e.g., for `links(1,:)` equal to [779,20278] means that actor personName20278 was in movie movieName779) (James Jimmy Stewart in the movie Rope)
- `movieIMDbID` : the IMDb IDs of each movie
- `movieName` : the name of each movie
- `movieRating` : the average IMDb rating of people who have rated this movie online
- `movieVotes` : the number of people who have rated this movie on IMDb
- `movieYear` : the year in which this movie was released

- `personIMDBID` : the IMDB IDs of each actor/actress
- `personName` : the name of each actor/actress

None of these are the proper PageRank matrix P.

(a) Let C be the $m \times n$ matrix defined by $C_{ij} = 1$ if movie i contains actor/actress j. Let B be the $m \times m$ matrix where B_{ij} is the number of actors starring in both movie i and movie j.

(i) Show how to calculate B from C.

(ii) Show how to calculate the PageRank matrix P from B: (Hint: it may help to construct a small graph of movies and actors (need not be based on real data) and to actually construct these individual matrices). Remember that movie i and movie j sharing at least one actor counts as one link from movie i to movie j.

(b) Using matlab, construct the PageRank matrix P. DO NOT PRINT THIS MATRIX OUT. Instead, submit the sparsity plot of P using the command `spy(P)`. Use sparse matrix commands to assist you; otherwise, matlab may be temperamental.

(c) Compute the PageRank vector $\vec{x}$ of the movies in this graph and normalize this quantity so $||\vec{x}||_1 = 1$. List the PageRank values and titles of the movies with the five highest PageRank values.

(d) Compute the PageRank vector $\vec{x}$ of the movies via a Jacobi iteration that you write yourself and normalize this quantity so $||\vec{x}||_1 = 1$ after each iteration. Decide on an intelligent measure for convergence (assume you do not know the actual PageRank vector $\vec{x}$ because your system is too large for a simple backslash calculation.) Explain your choice of convergence criterion. Next, plot this convergence measure over the steps in your iteration. How many iterations does it take your implementation to get to a tolerance of 10^{-4}.

(e) Plot IMDb movie rating vs. PageRank and IMDb movie votes vs. PageRank. Is PageRank a good predictor of movie rating or movie votes?

Answer for Exercise 11.3

Chapter 12

Exercises & Solutions

* Exercise 12.1

The singular value decomposition of the $m \times n$ matrix A is given by $A = U\Sigma V^T$, where the $m \times m$ matrix U and the $n \times n$ matrix V are orthogonal. The matrix Σ is an $m \times n$ matrix with the singular values σ_i on its diagonal. For simplicity, assume that $m \geq n$, so that A has n singular values.

(a) Prove that r, the rank of A, is equal to the number of nonzero singular values.

(b) Prove that the column space of A is spanned by the first r columns of U, where r is the rank of A.

Answer for Exercise 12.1

Exercise 12.2

Let A be an $m \times n$ matrix. Assume $n > m$ and the rank of A is m. In this case, $A\vec{x} = \vec{b}$ has an infinite number of solutions. We are interested in finding the solution to $A\vec{x} = \vec{b}$ that has minimum norm., We will use 2-norm here. The solution we after is called the "minimal 2-norm solution".

Show that the minimal 2-norm solution is given by $\vec{x} = V \begin{bmatrix} S^{-1}U^T\vec{b} \\ 0 \end{bmatrix}$, where S is the $m \times m$ matrix containing the first m columns of Σ.

Answer for Exercise 12.2

We note $\Sigma = \begin{bmatrix} S & 0 \end{bmatrix}$.

$$A\vec{x} = \vec{b} \rightarrow U\Sigma V^T\vec{x} = \vec{b} \rightarrow \Sigma\vec{y} = U^T\vec{b},\ \vec{y} = V^T\vec{x}$$

Since V is orthogonal, if we create the minimum 2-norm solution $\vec{y}$ to $\Sigma\vec{y} = U^T\vec{b}$, then $\vec{x} = V\vec{y}$ will be the minimum 2-norm solution to $A\vec{x} = \vec{b}$. Using the structure of Σ, we see

$$\Sigma\vec{y} = \begin{bmatrix} S & 0 \end{bmatrix} \begin{bmatrix} \vec{y}_1 \\ \vec{y}_2 \end{bmatrix} = S\vec{y}_1 = U^T\vec{b}$$

We conclude that $\vec{y}_1 = S^{-1}U^T\vec{b}$ and $\vec{y}_2$ is free to choose. But

$$\|\vec{y}\|_2^2 = \vec{y}^T\vec{y} = \vec{y}_1^T\vec{y}_1 + \vec{y}_2^T\vec{y}_2$$

so if we want to minimize $\|\vec{y}\|_2^2$, we set $\vec{y}_2 = \vec{0}$. Therefore, we get

$$\vec{x} = V \begin{bmatrix} S^{-1}U^T\vec{b} \\ 0 \end{bmatrix}$$

* Exercise 12.3

Indicate whether the following statement is TRUE or FALSE and motivate your answer clearly. To show a statement is false, it is sufficient to give one counter example. To show a statement is true, provide a general proof.

If an $n \times n$ matrix A is symmetric, its singular values are the same as its eigenvalues.

Answer for Exercise 12.3

* Exercise 12.4

In the book, we discussed that the number of nonzero singular values of a matrix A gives the rank of A. This is not generally true for eigenvalues. That is, if $n \times n$ matrix A has p nonzero eigenvalues, we cannot conclude that the rank of A equals p, *unless* A belongs to a special class of matrices. Explain why this must be so, and identify this special class of matrices.

Answer for Exercise 12.4

The key is to observe that the rank is simply dependent on n and the dimension of the nullspace ($r = n - \dim(\mathcal{N}(A))$). If we have p nonzero eigenvalues, it means that the zero eigenvalue is repeated $n - p$ times. Then, the dimension of the nullspace can be anything between 1 and $n - p$. Only if the zero eigenvalue has $n - p$ eigenvectors will the dimension be $n - p$ and the rank of A be p. We can assure this is the case for diagonalizable matrices.

Chapter 13

Exercises & Solutions

Exercise 13.1

Power method

In one of the exercises, we looked at the Power method for a matrix that has $\lambda_1 = \lambda_2$. What would happen if $\lambda_1 = -\lambda_2$, in other words, what would happen if the eigenvalues are the same in absolute value, but have a different sign?

Answer for Exercise 13.1

Exercise 13.2

Canonical transform and SVDs

We know that a symmetric matrix has a canonical transform of the type $A = Y\Lambda Y^{-1}$, with $Y^{-1} = Y^T$. We also know that a symmetric matrix has real eigenvalues, but the eigenvalues could of course be negative. Every matrix, so also a symmetric matrix, has an SVD of the form $A = U\Sigma V^T$. Here, the singular values that appear on the diagonal of Σ are real and always nonnegative. Can you find this singular value decomposition (so the matrices U, V and Σ for a symmetric matrix from its canonical transform?

Answer for Exercise 13.2

Our first observation is that for symmetric A we can always find a canonical transform because A is diagonalizable. So, $A = Y\Lambda Y^{-1}$. We also know that for symmetric matrices, the Y matrix can be chosen orthogonal, so $Y^{-1} = Y^T$. Finally, we know that all eigenvalues are real. However, we do not know anything about the sign of the eigenvalues. Now, $A = Y\Lambda Y^T$ does look a bit like $A = U\Sigma V^T$, as Λ and Σ are both diagonal, and U, V, Y are all orthogonal. But, singular values are always non-negative and eigenvalues may be negative. So, instead we write $A = Y\text{abs}\Lambda\text{sign}(\Lambda)Y^T$, where sign$\Lambda$ is a diagonal matrix with the signs of the eigenvalues (if the $i-$th eigenvalue is positive, the i-th diagonal element is $+1$, if the eigenvalue is negative, the diagonal element is -1 and 0 otherwise). Here, abs just denotes absolute value. Now, we can set, for example, $U = Y$, $V = sign(\Lambda)Y$ and $\Sigma = |\Lambda|$.

Exercise 13.3

Symmetric matrices

(a) Use the Schur form to prove that the eigenvectors of a symmetric matrix can always be chosen to be orthogonal.

(b) Now show the same as in question (a), that is $\vec{y}_i^T\vec{y}_j = 0$ when $i \neq j$, but only by using the relations $A = A^T$ and $A\vec{y}_i = \lambda_i\vec{y}_i$. You may assume that the eigenvalues are distinct in this case.

(c) The symmetric matrix A is given in MATLAB notation as

```
A = [3, 1, 0; 1, 3, 0; 0, 0, 2].
```

Compute the eigenvectors and eigenvalues of A. According to problems (a) and (b), we should be able to get an orthogonal set of eigenvectors. Check this.

(d) Suppose that A is a symmetric matrix with eigenvalues $\lambda_1 = 0, \lambda_2 = 0, \lambda_3 = 2$. The corresponding eigenvectors are $\vec{y}_1, \vec{y}_2,$ and $\vec{y}_3$. Show that both the column and the row spaces of A are spanned by $\vec{y}_2$ and $\vec{y}_3$.

Answer for Exercise 13.3

Exercise 13.4

True/false questions jumping around the material

Indicate whether the following statements are TRUE or FALSE and motivate your answers clearly. To show a statement is false, it is sufficient to give on counter example. To show a statement is true, provide a general proof.

(a) If A has the same nullspace as A^T, then A must be a square matrix.

(b) If the triangular matrix T is similar to diagonal matrix, T must be diagonal itself.

(c) If all eigenvaluse of a $n \times n$ matrix A are positive, then $A\vec{x} = \vec{b}$ always has a solution for any $\vec{b}$.

(d) If the $n \times n$ matrix A is symmetric, then the matrix R in the QR decomposition of A is diagonal.

(e) If A is symmetric and positive definite, and Q is an orthogonal matrix then QAQ^T is symmetric, but not necessarily positive definite.

(f) The eigenvalues of A are the same as the eigenvalues of U. Here, U is the matrix after Gaussian Elimination. You may assume that pivoting is not required to get U, so $A = LU$.

Answer for Exercise 13.4

(a) True.

If these nullspaces are the same, they must live in the same $\mathbb{R}^p$. For A the nullspace lives in $\mathbb{R}^n$ and for A^T in $\mathbb{R}^m$, assuming that A is $m \times n$. Hence $p = m = n$.

(b) False.

A counter example is easy to find. Take any diagonalizable triangular matrix T (for example any triangular matrix with distinct diagonal elements). Because it is diagonalizable, we know that it is similar to Λ, but it is not diagonal itself.

(c) True.

If there is a solution for any $\vec{b}$, then A must be nonsingular. If all eigenvalues are positive, there is no zero eigenvalue and hence A is indeed nonsingular.

(d) False.

Easiest to find a counter example. Take a $2\times$ matrix that is symmetric but not diagonal, for example

$$A = \begin{bmatrix} 3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}.$$

Q and R are then very easily found as

$$Q = \begin{bmatrix} 3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}, \quad \text{and } R = \begin{bmatrix} 1 & 24/25 \\ 0 & 7/25 \end{bmatrix}$$

(e) True.

A symmetric positive definite matrix A satisfies $\vec{x}^T A\vec{x} > 0$ for all $\vec{x} \neq \vec{0}$. Now,

$$\vec{y}^T Q^T AQ\vec{y} = (Q\vec{y})^T A(Q\vec{y}) = \vec{x}^T A\vec{x}, \text{ with} \vec{x} = Q\vec{y}.$$

Since for $\vec{y} \neq \vec{0}$, we cannot have $\vec{x} = Q\vec{y} = \vec{0}$ (Q is nonsingular after all), we must have that $\vec{y}^T Q^T AQ\vec{y} > 0$, for all$\vec{y} \neq \vec{0}$, and so $Q^T AQ$ is certainly also positive definite. Note that the matrix is symmetric too, which you can easily check by comparing it to its transpose: $(Q^T AQ)^T = Q^T A^T Q = Q^T AQ$.

(f) False.

This seems far fetched, so we look for a counter example. Note that the eigenvalues of the triangular matrix U are the diagonal elements of U, so the pivots in Gaussian Elimination. A simple counter example would be the matrix (in MATLAB notation) $A = [2, -1; 1, 0]$, with eigenvalues 1. Gaussian Elimination leads to $U = [2, -1; 0, 1/2]$ with eigenvalues 2 and $1/2$.

---

* Exercise 13.5

Eigenvalues and Gaussian Elimination

Show that the product of the eigenvalues is equal to the product of the pivots in Gaussian elimination. You may assume that pivoting is not required.

Answer for Exercise 13.5

Exercise 13.6

Positive definite matrices

Positive-definite matrices occur on a regular basis in science and engineering. A matrix B is *positive definite* if for *all* $\vec{x} \neq \vec{0}$, we have that $\vec{x}^T B\vec{x} > 0$. Here, we assume that all matrices and vectors we work with are real.

(a) Show that positive definite matrices have only positive eigenvalues.

(b) Is it true that if the $m \times n$ matrix A has rank n, then the matrix $A^T A$ is symmetric and positive definite?

Answer for Exercise 13.6

(a) Let $A\vec{x} = \lambda\vec{x}$, where $\vec{x}$ is an eigenvector which is of course not equal to the zero vector. Then

$$\vec{x}^T A\vec{x} = \lambda\vec{x}^T\vec{x} = \lambda\|\vec{x}\|_2^2.$$

From the definition of positive definiteness we know that the left hand side is larger than zero. So $\lambda\|\vec{x}\|_2^2 > 0$, which means that all $\lambda > 0$.

(b) Yes it is. First observe that $A^T A$ is symmetric. That is simply shown: $\left(A^T A\right)^T = A^T A$. Now we need to show that $\vec{x}^T A^T A\vec{x} > 0$ for all $\vec{x} \neq \vec{0}$. That is not so hard since $\vec{x}^T A^T A\vec{x} = \|A\vec{x}\|_2^2$ and this is by definition bigger than zero except when $A\vec{x} = \vec{0}$. Since A is full column rank (rank of A is n), we know that the dimension of its nullspace is 0. See this? Hence, we have $A\vec{x} = \vec{0}$ only for $\vec{x} = \vec{0}$ and we are done.

Exercise 13.7

Orthogonalization

(a) Suppose that we are doing Gram-schmidt on a matrix A that is 10×10. Suppose that the third column of A is linearly dependent on the other nine columns Explain what happens when you get to this column in QR, and what you could do to resolve the conundrum.

(b) QR can be used to solve systems of equations $A\vec{x} = \vec{b}$, with A inonsingular. In the notes it is explained how. Try this method for a few nonsingular matrices in MATLAB. You really have two direct (non-iterative) methods to solve $A\vec{x} = \vec{b}$: LU and QR. LU is often well understood, but very few remember that QR is also possible and equally attractive in most cases. Hence, this little play to consolidate it in your mind. Can you think of matrices for which LU is more attractive than QR, and perhaps vice versa?

Answer for Exercise 13.7

Exercise 13.8

Diagonalizability

In the notes, it is shown that symmetric matrices are diagonalizable using the Schur form. We can do the same for an Hermitian matrix A: $A = USU^*$ and $A = A^*$ gives $USU^* = US^*U^*$ so that $S = S^*$ and therefore we see that A is diagonalizable. Fill in all the gaps, that is, explain how we transition from one statement to the next in this proof.

Answer for Exercise 13.8

$USU^* = US^*U^*$ comes simply from the statement $A = A^*$ combined with $A = USU^*$. The rules that apply to the transpose (eg. $(AB)^T = B^T A^T$) apply to * as well.
$S = S^*$ follows from $USU^* = US^*U^*$ by premultiplying both sides of this equation with U^* and postmultiplying by U. We then use the fact that U is unitary, so that $U^*U = I$.
The last observation (that $S = S^*$ must mean that S is diagonal) comes from the knowledge that S is upper triangular. Then, S^* is lower triangular and clearly, an upper and lower triangular matrix cannot be equal unless they are both diagonal.
Here, we can go a step further. The diagonal elements must all be equal to their conjugate (because $S = S^*$), which means that all diagonal elements must be real. We see that a Hermitian matrix has real eigenvalues only (because the eigenvalues of A are on the diagonal of S).

Exercise 13.9

Decompositions

Collect all decompositions of a matrix you have seen thus far in the book, and for each of them list/discuss:

- uses
- computational complexity
- special cases (eg. does it simplify if the matrix is symmetric? any other special cases we discussed?)

If uses for decompositions overlap (eg. both are used to solve some system) then discuss pros and cons as well.

Answer for Exercise 13.9

Exercise 13.10

Complex matrices

(a) Find two complex Hermitian matrices that are 2×2 and 3×3. Are these matrices also symmetric? Repeat the same for unitary matrices.

(b) For orthogonal matrices Q, we found that $\|Q\vec{x}\|_2 = \|\vec{x}\|_2$. Convince yourself of this again. Does a similar relation hold for unitary matrices U? In other words, if U is unitary, is it true that $\|U\vec{x}\|_2 = \|\vec{x}\|_2$, where $\vec{x}$ is a complex vector?

Answer for Exercise 13.10

(a) Note that Hermitian is the complex equivalent of symmetric but is not equal to symmetric: *Hermitian = symmetric + complex conjugate*. Recall that an $n \times n$ matrix A is Hermitian if $A = A^*$. Two examples of such matrices are

$$A = \begin{bmatrix} 2 & 1-i \\ 1+i & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & i & 3+2i \\ -i & 2 & 3i \\ 3-2i & -3i & 1 \end{bmatrix}$$

(b) Yes, this is true. Same relation holds using the definition of the 2-norm given in Section

$$\|U\vec{x}\|_2^2 = (U\vec{x})^* (U\vec{x}) = \vec{x}^* U^* U \vec{x} = \vec{x}^* \vec{x}.$$

as $U^*U = I$ by definition of unitary matrix.

Some of you may have seen the superscript ^{H} instead of * in textbooks. In computational mathematics * is preferred as ^{H} is often use for things related to Hilbert spaces.

Exercise 13.11

Systems of ODEs

(a) Find a 2×2 matrix A that has 2 distinct eigenvalues. Compute its canonical form.

(b) For the matrix you choose, compute the solution to $\frac{d\vec{x}}{dt} = A\vec{x}$, with initial condition $\vec{x}(0) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$.

(c) Find a 2×2 matrix B that has a repeated eigenvalue but is still diagonalizable. Again, compute the solution to $\frac{d\vec{x}}{dt} = B\vec{x}$, with initial condition $\vec{x}(0) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. What is $\vec{x}$ at time $t = 10$?

(d) Find a 2×2 matrix C that has a repeated eigenvalue and is not diagonalizable. Again, compute the solution to $\frac{d\vec{x}}{dt} = C\vec{x}$, with initial condition $\vec{x}(0) = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. What is $\vec{x}$ at time $t = 10$?

Answer for Exercise 13.11

Exercise 13.12

Determinants

(a) Why is the determinant of a triangular matrix equal to the product of the diagonal elements? Try to really understand this yourself. Do a few examples for 2×2 case, and then argue that it holds for any $m \times m$ case.

(b) Explain in your own words why it makes sense that matrices that are singular have a 0 determinant.

(*c) Can you show that $\det(AB) = \det(A)\det(B)$? This is not as easy as it looks.

(c) Explain why the determinant is equal to the product of the pivots in LU decomposition (assume $A = LU$ so no row swapping was needed in the Gaussian elimination process). You may use the result of (c).

(d) A matrix A is singular if $|A| = 0$ but if the determinant is not zero, we cannot say anything about the matrix being close to singular or not. Give an example of a matrix that is far from being singular, yet has a tiny determinant, and an example of a matrix that is very close to being singular, but has a large determinant.

Answer for Exercise 13.12

(a) There are various approaches to this problem. I like the following argument. Assume that the matrix is lower triangular. I refer to the general determinant formula we discussed in class. To find the determinant, we loop over the first row. The only nonzero in this first row is a_{11}. All other elements are zero. This means, following the formula, that we are left with $\det(A) = a_{11}\det(M_{11})$, where M_{11} is the submatrix left after removing the first row and column. However, this submatrix is in itself lower triangular. So, when computing the determinant, we go through exactly the same process as before, now with a_{22} as only nonzero on the first row. This process repeats itself until we find the last submatrix, which is simply the element a_{nn}. Therefore $\det(A) = a_{11}a_{22}\cdots a_{nn}$.

(b) If a matrix is singular, U will have a zero row, and according to the properties of the determinant discussed in the notes, the determinant will be zero.

(*c) This is proven in Strang.

(c) Using (c), we see $\det(A) = \det(L)\det(U)$ and using part (a), we know $\det(L) = 1$ and $\det(U)$ is product of the pivots.

(d) An example of the first is a the identity times a very small constant. An example of the latter is the matrix (in MATLAB notation) $[10^8, 10^8; 0, 0.1]$.

Exercise 13.13

Stationary iterative methods

We know that both Jacobi and Gauss-Seidel converge for diagonally dominant matrices. Find a diagonally dominant 3×3 matrix (or larger if you can use MATLAB) and compare the convergence rates (here, the norms of the amplification matrices). Do you expect GS to converge faster than Jacobi? Does the answer make sense? Now, increase the diagonal dominance by making the diagonal fatter. Check again. Do you see a faster convergence overall? Is that expected?

Answer for Exercise 13.13

Exercise 13.14

Random questions

(a) Suppose that for the matrices C and D, it holds that $CD = -DC$. Is there a flaw in the following argument:
"Taking determinants gives $\det(CD) = \det(C)\det(D) = -\det(DC) = -\det(D)\det(C)$ *so we end up with* $\det(C)\det(D) = -\det(C)\det(D)$*, which means that eigher C or D must have a zero determinant. Therefore $CD = -DC$ is only possible if C or D is singular."*

(b) A matrix A is given by $A = \vec{u}\vec{v}^T$, where $\vec{u}$ and $\vec{v}$ are in $\mathbb{R}^n$. Show that this matrix always has either rank 0 or rank 1, and give an example for both cases.

(c) We have seen in the least squares discussions that $P = A\left(A^TA\right)^{-1}A^T$ is the projection matrix that projects onto the column space of A provided A is full rank. What is the project matrix that projects onto the *row* space of A?

(d) Is the following statement true of false? For any $n \times n$ matrix A, the eigenvalues of A and its transpose are always the same.

Answer for Exercise 13.14

(a) Yes, definitely. We are not allowed to say $\det(-DC) = -\det(DC)$, which we do in the above argument. Remember from the determinant properties that $det(-A) = (-1)^n \det(A)$. Now, if n is odd, the result is still correct. If n is even, we cannot conclude that C or D is singular.

(b) This is such an interesting matrix. Now, clearly if either $\vec{u}$ or $\vec{v}$ are zero, the matrix is simply 0 and has rank 0. What if $\vec{u}$ and $\vec{v}$ are both not equal to $\vec{0}$?
We can do a few things here. The simplest is to find the actual elements of the matrix. Can you see that $a_{ij} = u_iv_j$? In other words, row i of the matrix is simply $u_i[v_1, v_2, \ldots, v_n]$. Each row is just a multiple of the row vector $\vec{v}^T$. Naturally then, the rank of this matrix is at most 1, and if $\vec{v} \neq \vec{0}$ and $\vec{u} \neq \vec{0}$ then the rank is 1.

(c) The simplest way to do this is to observe that the row space of A is just the column space of A^T. We know how to find projection matrices onto column spaces, so we can do this for A^T. We would get $P = A^T\left(AA^T\right)^{-1}A$. We have to be a little careful of course to make sure that the inverse in this expression actually exists! For that, we'd need A^T to be full column rank, just as we asked full column rank for A (well, we asked full rank for A but since for A we had $m > n$, this meant full column rank). In this case then, with A^T instead of A, we want $n > m$.

(d) Yep. For this, we can use the property of the determinant that $\det(A) = \det(A^T)$. We know that $\det\left(A - \lambda I\right) = \det((A - \lambda I)^T) = \det\left(A^T - \lambda I\right)$ and so each time $\det(A - \lambda I) = 0$,we get the same for A^T. This means the characteristic equations are the same, and so the eigenvalues are the same.

Exercise 13.15

Spaces and Dimensions

(a) Prove that for any $m \times n$ matrix A, the dimensions of the row space and the column space of A are the same.

(b) The dimension of the nullspace of the $m \times n$ matrix A is $n - r$, where r is the rank of the matrix. Take $m > n$.

(i) Create a matrix that is full rank (so has rank n). If $\vec{b}$ is in the column space of A, is the solution to $A\vec{x} = \vec{b}$ unique in this case? Is every $\vec{b}$ in the column space of A? Explain.

(ii) Create a matrix that is not full rank (so has rank $< n$). Repeat the questions in (i).

Answer for Exercise 13.15

Exercise 13.16

QR decomposition

(a) Derive the full QR decomposition for the matrix

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

(b) We have an $m \times n$ matrix A. Is it true that if A is full rank, the skinny QR decomposition and the full QR decomposition are always the same?

(c) In Gram-Schmidt we used projections onto lines. What is the corresponding projection matrix?

Answer for Exercise 13.16

(a) Here the QR is relatively straightforward. The first vector $\vec{q}_1$ will be as is in the matrix (first unit vector), the second ends up being the second unit vector $\vec{q}_2 = [0, 1, 0]^T$ (simply the second unit vector) and then we have to supplement the matrix to form a third orthogonal vector. Thankfully, the $\vec{q}$'s so far as so nice that we can readily suggest $q_3 = [0, 0, 1]^T$, the third unit vector. R is also very easy. Since the full Q is the identity, R is just equal to A, as $A = QR$!

(b) Nope. Mind here that we do not say anything about the size of m and n. For example, m could be larger than n. If A if full rank for a rectangular matrix, it just means that A has the maximum rank possible. For a rectangular matrix, the maximum rank possible is simply $\min(m, n)$. For $m > n$, the max rank possible is therefore n. This means that when going through skinny QR we will obtain n orthogonal $\vec{q}$'s. In the skinny QR, Q is an $m \times n$ matrix. For the full QR, we need Q to be $m \times m$. We therefore need to add $m - n$ additional columns to Q to create the full Q, and the skinny and the full are clearly not the same.

(c) In GS we take a vector $\vec{a}$ and project it on the line spanned by a $\vec{q}$. We showed that this projection is given by $\left(\vec{a}^T\vec{q}\right)\vec{q}$. The first term in brackets is an innerproduct, so we can write $\left(\vec{q}^T\vec{a}\right)\vec{q}$. The innerproduct is just a scalar and so we can put this scalar also behind $\vec{q}$ instead of in front. So we can write $\vec{q}\left(\vec{q}^T\vec{a}\right)$. Now, we can write this as $\left(\vec{q}\vec{q}^T\right)\vec{a}$. The first term in this expression is a matrix! It takes $\vec{a}$ and creates the projection of a on the line through $\vec{q}$. This is therefore the projection matrix we are after.

Exercise 13.17

Symmetric matrices

Go through all your notes and collect everything you know about symmetric matrices. In particular, answer the following questions:

- How can you take advantage of symmetry in the LU decomposition?
- Can you take advantage of symmetry in the QR decomposition?
- Is there anything special you can do for least squares when you have a symmetric matrix?

- What do you know about the various spaces of a symmetric matrix (row, column, null)?
- Can you say anything about singularity of a symmetric matrix?
- Can you say anything about the determinant of a symmetric matrix?

Answer for Exercise 13.17