Lecture notes on advanced linear algebra

Benjamin J. Clark

Fall 2025

Contents

Chapter 0

Review from linear algebra

0.1 Linear equations and systems

Definition 0.1.1. A **system of linear equations** is a collection of linear equations in the same variables $x_1, \ldots, x_n$, that is

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

A system of linear equations is called **consistent** if it has either one of infinitely many solutions. If a system has no solutions it is called **inconsistent**.

When a system has more equations than variables we call it **overdetermined**. If there are more variables than equations we call it **underdetermined**.

Definition 0.1.2. The **coefficient matrix**, A, is a rectangular array having m rows and n columns where the coefficients of the system are recorded. The right–hand side vector, b, is a column that records the right–hand sides of the equations. That is,

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}, b = \begin{bmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{bmatrix}.$$

For any $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, we refer to a_{ij} as the (i, j)th entry of the matrix A and to b_j as the jth entry of the vector b.

We can combine the above information into one matrix, called the augmented matrix of the linear system, as follows:

$$[A \mid b] = \left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array}\right].$$

Definition 0.1.3 (Operations that preserve solution set)**.** The following three operations preserve the solution set of a system of linear equations

1. Add a multiple of one equation to another equation.

2. Multiply any equation by a nonzero number.

3. Permute the equations as needed.

Definition 0.1.4. A matrix is said to be in echelon form (EF) if the following conditions are satisfied:

1. Any zero rows (i.e., all row entries being equal to 0) are at the bottom.

2. Every nonzero row has a leading (leftmost) nonzero entry, called the pivot, which is in a column to the right of the pivot in any row above it. We refer to a column that contains a pivot as a pivot column and the location of the pivot as the pivot position in that column.

3. All the column entries below a pivot are equal to 0.

We say the matrix is in reduced echelon form (REF) if it is in EF and in addition

1. All pivots are equal to 1.

2. All the column entries above a pivot are equal 0.

With the above set of definitions we can systematically solve a system of equations. This is generally taught using Gaussian Elimination. This algorithm is effective and easy to follow, however it is computationally inefficient and generally not used by computers. Matlab for example uses a form of LU decomposition which we will see later this semester. Solving a system of equations can be done with a myriad of algorithms in $O(n^3)$ time, this means that even massive matrices (millions by millions) can generally be solved in a short period of time.

0.2 Matrix notation and basic properties

A matrix is denoted using square brackets and with a capital letter. If a matrix is said to be m by n, then we mean m rows by n columns.

The notation $A = [a_{ij}]$ for $i = 1, \ldots, m$ and $j = 1, \ldots, n$. Means A is an m by n matrix whose ith jth entry is a_{ij}. This is useful when we want a generic matrix where we can reference specific entries. Another common notation is

$$A = [a_1 \mid a_2 \mid \cdots \mid a_n]$$

where a_i is a column vector of A.

A matrix is called square if and only if $m = n$. Two matrices are equal if their dimensions (m, n) are equal and all their entries are equal. The diagonal entries of a matrix are when the row and columns are equal. The off-diagonal entries are the rest.

Denote $\mathsf{M}_{n\times m}(S)$ to be the set of $n \times m$ matrices whose entries are pulled from the set S. If $n = m$, then we will write $\mathsf{M}_n(S)$. Another frequent notation for a set of matrices is $S^{n\times n}$, this is most common with $\mathbb{R}^{m\times n}$.

A matrix is called **upper triangular** if all the entries below the main diagonal are zero. A matrix is called diagonal if all off-diagonal entries are zero.

Definition 0.2.1 (Matrix addition and scalar multiplication)**.** Let $A = [a_{ij}]$ and$B = [b_{ij}]$ be two $m \times n$ matrices, and let r be a scalar.

The matrix sum of A and B is an $m \times n$ matrix $C = A + B = [c_{ij}]$ such that

$$c_{ij} = a_{ij} + b_{ij} \text{ for all } i = 1, 2, \ldots, m \text{ and all} j = 1, 2, \ldots, n.$$

The scalar product of A by r is an $m \times n$ matrix $C = rA = [c_{ij}]$ such that

$$c_{ij} = ra_{ij} \text{ for all } i = 1, 2, \ldots, m \text{ and all } j = 1, 2, \ldots, n.$$

Theorem 0.2.2 (Basic properties of matrix addition)**.** *Let A, B, C be matrices and r be a scalar.*

1. $A + B = B + A$ *(commutative property).*
2. $(A + B) + C = A + (B + C)$ *(associative property).*
3. $A + 0 = A = 0 + A$ *(we say 0 is the additive identity or the neutral of addition).*
4. $A + (-A) = 0 = (-A) + A$ *(we say $-A$ is the additive inverse of A).*
5. $(rs)A = (sr)A = s(rA) = r(sA)$.
6. $r(A + B) = rA + rB$ *and* $(r + s)A = rA + sA$ *(distributive property).*
7. $1A = A$.
8. *If* $A + B = C$*, then* $A = C - B$ *and* $B = C - A$.
9. *If* $A + C = B + C$*, then* $A = B$ *(cancellation law).*

With the above properties we shall see that matrices form a vector space under addition and scalar multiplication.

Definition 0.2.3 (Matrix vector multiplication)**.** Given an $m \times n$ matrix A whose columns are $a_1, a_2, \ldots, a_n$, and given a column vector x whose entries are $x_1, x_2, \ldots, x_n$, the product Ax is defined as the linear combination

$$Ax = [a_1 \mid a_2 \mid \cdots \mid a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n.$$

Definition 0.2.4 (Matrix Multiplication)**.** Let A be an $m \times k$ matrix and B a $k \times n$ matrix whose columns are $b_1, b_2, \ldots, b_n$. Then, the product of matrices AB is defined as the $m \times n$ matrix whose columns are Ab_1, Ab_2, $\ldots$, Ab_n, that is,

$$AB = A[b_1 \mid b_2 \mid \cdots \mid b_n] = [Ab_1 \mid Ab_2 \mid \cdots \mid Ab_n].$$

It is important to think about the above theoretical importance of what matrix multiplication is, each column of AB is a linear combination of the columns of A using the entries of the corresponding column of B. So one way to think about this is that the matrix B is the weights of how we want the columns of A to be mixed and combined.

A more practical way of actually computing the entries of the products is for $A = [a_{ij}]$ and $B = [b_{ij}]$, then the ith jth entry of AB is

$$a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ik}b_{kj} = \sum_{h=1}^{k} a_{ih}b_{hj}.$$

Definition 0.2.5 (Commutative matrices)**.** We say that two matrices A and B commute if AB and BA are both well-defined products and $AB = BA$.

Definition 0.2.6 (Identity and zero matrix)**.** We call the matrix I which is a diagonal matrix of all 1's the identity matrix. That is $I = \operatorname{diag}(1, 1, \ldots, 1)$.

We denote the zero matrix 0 or 0_n to be the matrix of all zero entries.

Definition 0.2.7 (Trace)**.** The trace of a matrix is the sum of the diagonal entries. Symbolically if $A = [a_{ij}]$ is a matrix, then

$$\operatorname{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}.$$

Definition 0.2.8 (Transpose)**.** The transpose of a matrix swaps its rows and columns. Symbolically if $A = [a_{ij}]$, then $A^\top = [a_{ji}]$.

Theorem 0.2.9 (Matrix multiplication properties)**.** *Let A, B, C be matrices, then*

1. $AI = IA = A$.
2. $A0 = 0A = 0$.
3. $A(B + C) = AB + AC$ *and* $(B + C)A = BA + CA$.

4. *$A(BC) = (AB)C$.*

5. *$tr(AB) = tr(BA)$.*

6. *$(AB)^\top = B^\top A^\top$.*

7. *$AB = 0$ does **not** imply that either A or B is the zero matrix.*

Note that the last property of matrix multiplication means that generally division of matrices is not well defined.

Definition 0.2.10. Let A be a square matrix, then $A^k = AAA\ldots A$, so powers are the standard repeated multiplication. By convention $A^0 = I$. We will talk later about negative powers.

Definition 0.2.11. A matrix A is called **symmetric** if $A = A^\top$. The matrix A is instead called **skew-symmetric** if $A^\top = -A$.

0.2.1 Exercises

1. For matrices
$$A = \begin{bmatrix} 1 & 1 & 5 \\ 7 & 3 & -2 \\ 15 & 1 & -3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 2 & 0 \\ 4 & 0 & 3 \\ -7 & 1 & 2 \end{bmatrix}.$$
Find the trace of the following
$$A, A^\top, B, \text{ and } AB.$$

2. For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, prove that $A^2 - (a+d)A + (ad - bc)I = 0$.

3. Let $A = \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$. Find all matrices $X = \begin{bmatrix} 1 & x \\ y & z \end{bmatrix}$ such that $AX = XA$.

4. Find three 2×2 matrices $A(\neq 0)$, B, and C such that $AB = AC$ and $B \neq C$.

5. If A, B are both n by n matrices with real entries such that $\text{tr}(A) = 3$ and $\text{tr}(B) = 5$, then what can we say, if anything, about $\text{tr}(AB)$?

6. Find two matrices A, B such that $AB = 0$, but $A \neq 0$ and $B \neq 0$.

7. Let A and B be two symmetric matrices such that $A^2 + B^2 = 0$. Prove that $A = B = 0$. Hint: Use the trace of A^2.

8. Let A and B be two $n \times n$ matrices such that $A^2 = A$, $B^2 = B$, and $(A+B)^2 = A + B$. Prove that $AB = BA = 0$.

0.2.2 Solutions

1. For matrices
$$A = \begin{bmatrix} 1 & 1 & 5 \\ 7 & 3 & -2 \\ 15 & 1 & -3 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 3 & 2 & 0 \\ 4 & 0 & 3 \\ -7 & 1 & 2 \end{bmatrix}.$$
Find the trace of the following
$$A, A^\top, B, \text{ and } AB.$$
The trace is the sum of the diagonal entries, so the trace of A is $1+3+-3=1$. The trace of $A^\top$ is still 1 since taking the transpose does not affect the main diagonal. The trace of B is $3+0+2=5$. Finally to find the trace of AB we can solve for the three diagonal entries which are $1(3)+1(4)+5(-7)=-23$, $7(2)+3(0)+-2(1)=12$, and $15(0)+1(3)+-3(2)=-3$. So the trace is -14.

2. For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, prove that $A^2-(a+d)A+(ad-bc)I=0$.

To prove this we can take each piece of the sum on the left individually, giving
$$A^2 = \begin{bmatrix} a^2+bc & ab+bd \\ ac+cd & bc+d^2 \end{bmatrix}.$$
$$(a+d)A = \begin{bmatrix} a^2+ad & ab+bd \\ ac+cd & ad+d^2 \end{bmatrix}$$
$$(ad-bc)I = \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix}.$$
Now taking the sum we can see that each entry gives 0.

3. Let $A = \begin{bmatrix} 4 & 2 \\ 3 & 1 \end{bmatrix}$. Find all matrices $X = \begin{bmatrix} 1 & x \\ y & z \end{bmatrix}$ such that $AX = XA$.

To start we can find AX and XA which are
$$AX = \begin{bmatrix} 4+2y & 4x+2z \\ 3+y & 3x+z \end{bmatrix},$$
$$XA = \begin{bmatrix} 4+3x & 2+x \\ 4y+3z & 2y+z \end{bmatrix}.$$
Now we can subtract these two to get the following 4 equations
$$\begin{aligned} -3x+2y &= 0 \\ 3x+2z-2 &= 0 \\ -3y-3z+3 &= 0 \\ 3x-2y &= 0. \end{aligned}$$

Notice that the last equation is just the negative of the first so we can ignore it. For the rest we can solve this with a standard coefficient matrix giving

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

Solving this gives the constraint that $3x+2z=2$. Thus, there are infinitely many solutions.

4. Find three 2×2 matrices $A(\neq 0)$, B, and C such that $AB = AC$ and $B \neq C$.

   There are a lot of ways to approach this problem. One option is to just pick a random nice matrix A, then solve for B and C. For this we can pick

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \ B = \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}, \text{ and } C = \begin{bmatrix} c_1 & c_2 \\ c_3 & c_4 \end{bmatrix}.$$

   Taking both sides gives

$$AB = \begin{bmatrix} b_1 + b_3 & b_2 + b_4 \\ b_1 + b_3 & b_2 + b_4 \end{bmatrix} = \begin{bmatrix} c_1 + c_3 & c_2 + c_4 \\ c_1 + c_3 & c_2 + c_4 \end{bmatrix} = AC.$$

   This gives that $b_1 + b_3 = c_1 + c_3$ and $b_2 + b_4 = c_2 + c_4$. Now we can just pick some values such that B and C are not equal. In this case let $b_1 = b_2 = 1$, $b_3 = b_4 = -1$, $c_1 = c_2 = 2$, and $c_3 = c_4 = -2$. Then we have satisfied the sums, we clearly have that $B \neq C$ and $AB = AC = 0$.

5. If A, B are both n by n matrices with real entries such that $\mathrm{tr}(A) = 3$ and $\mathrm{tr}(B) = 5$, then what can we say, if anything, about $\mathrm{tr}(AB)$?

   We can't say anything about $\mathrm{tr}(AB)$. Consider

$$A = \begin{bmatrix} x & & \\ & -x+3 & \\ & & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} y & & \\ & 0 & \\ & & -y+5 \end{bmatrix},$$

   then we nicely have the trace requirements for A and B. However $\mathrm{tr}(AB) = xy$ and since x and y are arbitrary numbers we can make this trace whatever we want. Thus there is nothing nice we can say about the trace of them together.

6. Find two matrices A, B such that $AB = 0$, but $A \neq 0$ and $B \neq 0$.

   We actually did this as part of the solution for problem 4. Take

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{ and } B = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

   Then we have $AB = 0$.

7. Let A and B be two symmetric matrices such that $A^2 + B^2 = 0$. Prove that $A = B = 0$.

Proof. Let $A = [a_{ij}]$ and $B = [b_{ij}]$ for $i, j \in \{1, \ldots, n\}$. Now if $A^2+B^2 = 0$ then we have $\operatorname{tr}(A^2 + B^2) = 0$ which gives that $\operatorname{tr}(A^2) + \operatorname{tr}(B^2) = 0$. Lets look at each trace individually

$$\operatorname{tr}(A^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}a_{ji}.$$

Since A is symmetric we have $a_{ij} = a_{ji}$ which simplifies the equation to

$$\operatorname{tr}(A^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2.$$

A similar calculation with B gives the following

$$\operatorname{tr}(A^2) + \operatorname{tr}(B^2) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 + b_{ij}^2 = 0.$$

Looking at what we have here all of the entries in the above sum are being squared and are thus nonnegative. If we sum a bunch of nonnegative terms together and get zero we must have that all of the terms are zero. Therefore $A = 0$ and $B = 0$. □

8. Let A and B be two $n \times n$ matrices such that $A^2 = A$, $B^2 = B$, and $(A + B)^2 = A + B$. Prove that $AB = BA = 0$.

Proof. Let $A, B \in \mathsf{M}_n$ such that $A^2 = A$, $B^2 = 0$, and $(A+B)^2 = A+B$, then expand the binomial normally giving

$$(A + B)^2 = A^2 + AB + BA + B^2 = A + AB + BA + B = A + B.$$

Thus we have that $AB + BA = 0$ or that $AB = -BA$. Now multiply this by A on the left giving $A^2B = -ABA$ which simplifies to $AB = -ABA$. Additionally multiply this by A on the right giving $ABA = -BA^2$ which becomes $BA = -ABA$. Thus $BA = -ABA$ and $AB = -ABA$ so $AB = BA$. However we have $AB = -BA$ and $AB = BA$ thus $AB = 0$ and $BA = 0$. □

Chapter 1

Vector spaces

1.1 Definition

Definition 1.1.1 (Field)**.** A *field* is a set with two binary operations, addition (denoted by $a+b$) and multiplication (denoted by ab), such that for all $a, b, c \in F$:

1. $a + b = b + a$.
2. $(a + b) + c = a + (b + c)$.
3. There is an additive identity 0; that is $0 + a = a$ for all $a \in R$.
4. There is an element $-a$ in R such that $a + (-a) = 0$.
5. $a(bc) = (ab)c$.
6. $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$.
7. $ab = ba$
8. $\exists 1 \in F$ such that $1a = a1 = a$.
9. If $a \neq 0$, then there exists $a^{-1} \in F$ such that $aa^{-1} = 1$.

Example 1.1.2. The rational ($\mathbb{Q}$), real ($\mathbb{R}$), and complex numbers ($\mathbb{C}$) form fields.

The integers ($\mathbb{Z}$) however, do not since they have no multiplicative inverses. △

Example 1.1.3. Let

$$\begin{aligned} \mathbb{Z}_3[i] &= \{a + bi \mid a, b \in \mathbb{Z}_3\} \\ &= \{0, i, 2i, 1, 1 + i, 1 + 2i, 2, 2 + i, 2 + 2i\}. \end{aligned}$$

Where the addition and multiplication operations are as they are for complex numbers mod 3. In particular note that $-1 = 2$. With these operations $\mathbb{Z}_3[i]$ forms a field with 9 elements. △

Definition 1.1.4 (Vector space)**.** A set V is said to be a vector space over a field F if V has two operations addition and scalar multiplication such that for all $u, v \in V$ and $a \in F$ they satisfy the following

1. $u + v = v + u$
2. $(u + v) + w = u + (v + w)$
3. There exists a $0 \in V$ such that $0 + u = u + 0 = u$.
4. There exists a $-u$ such that $u + (-u) = 0$.
5. $a(v + u) = av + au$
6. $(a + b)v = av + bv$
7. $a(bv) = (ab)v$
8. $1v = v$.

The elements in V are called vectors while the elements from F are called scalars.

If the field for a vector space is $\mathbb{R}$ we call it a real vector space and if the field is $\mathbb{C}$ we call it a complex vector space. These will be by far our most common vector spaces.

Example 1.1.5. The set

$$\mathbb{R}^n = \{(a_1, a_2, \ldots, a_n) \mid a_i \in \mathbb{R}\}$$

is a vector space over $\mathbb{R}$. Where the operations are elementwise addition and scalar multiplication. △

Example 1.1.6. The set $\mathsf{M}_2(\mathbb{Q})$ of 2 by 2 matrices with entries from $\mathbb{Q}$ is a vector space over $\mathbb{Q}$. The operations are

$$\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} + \begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix} = \begin{bmatrix} a_1 + b_1 & a_2 + b_2 \\ a_3 + b_3 & a_4 + b_4 \end{bmatrix}$$

and

$$b \begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix} = \begin{bmatrix} ba_1 & ba_2 \\ ba_3 & ba_4 \end{bmatrix}.$$

△

Example 1.1.7. The set $\mathbb{Z}_p[x]$ of polynomials with coefficients from $\mathbb{Z}_p$ is a vector space over $\mathbb{Z}_p$, where p is a prime.

More generally, if F is a field and $F[x]$ are the polynomials with coefficients from F, then $F[x]$ is a vector space. △

Theorem 1.1.8. *Let V be a vector space.*

1. *The zero vector of V is unique.*
2. *The negative of a vector $v \in V$ is unique.*
3. $-v = (-1)v$ *for all* $v \in V$.
4. *If $rv = 0$ for some $v \in V$ and $r \in F$, then either $v = 0$ or $r = 0$.*

Proof. Assume there is another zero element $u \in V$, then by the properties of being a vector space $u + 0 = 0$ and $u + 0 = u$. Thus $u = 0$ giving that there was only one zero element.

Let $v \in V$ and assume that u_1 and u_2 in V such that $v + u_1 = 0$ and $v + u_2 = 0$, then

$$u_1 = u_1 + 0 = u_1 + (v + u_2) = (u_1 + v) + u_2 = 0 + u_2 = u_2.$$

For any $v \in V$ we have

$$[1 + (-1)]v = 0 \implies 1v + (-1)v = 0 \implies (-1)v = -v.$$

Suppose $rv = 0$ for some $r \in F$ and $v \in V$, then proceed by cases. If $r = 0$, then we are done. If $r \neq 0$, then

$$v = 1v = \frac{1}{r}rv = \frac{1}{r}0 = 0.$$

□

1.1.1 Exercises

1. Prove that $V = \{(x, y) \in \mathbb{R}^2 \mid y = 5x^2\}$ is not a vector space.
2. Prove that $V = \{(x, y, z) \in \mathbb{C}^3 \mid z = 2x + 3y + 1\}$ is not a vector space.
3. Prove that the set of all nonnegative infinite sequences $V = \{(x_1, x_2, \dots) \mid x_i \geq 0\}$ is not a vector space.

1.1.2 Solutions

1. Prove that $V = \{(x, y) \in \mathbb{R}^2 \mid y = 5x^2\}$ is not a vector space.

   Consider the vector $a = (1, 5) \in V$, then $a + a = (2, 10)$ is not in V. Thus V is not a vector space since it is not closed under addition.

2. Prove that $V = \{(x, y, z) \in \mathbb{C}^3 \mid z = 2x + 3y + 1\}$ is not a vector space.

   This set V is missing the zero vector.

3. Prove that the set of all nonnegative infinite sequences $V = \{(x_1, x_2, \dots) \mid x_i \geq 0\}$ is not a vector space.

   Let $a = (1, 1, \dots) \in V$, then the additive inverse is $-a = (-1)a = (-1, -1, \dots)$ which is not in V.

1.2 Subspace

Definition 1.2.1 (Subspace)**.** Let V be a vector space over a field F and let U be a nonempty subset of V. We say that U is a *subspace* of V if U is also a vector space over F under the operations of V. We denote this by $U \leq V$.

Theorem 1.2.2. *Let U be a nonempty subset of a vector space V. Then, U is a subspace of V if and only if*

$$ru + sw \in U$$

for every $u, w \in U$ and $r, s \in \mathbb{R}$.

Proof. Let U be a subset of a vector space V and suppose that $ru + sw \in U$ for every $u, w \in U$ and $r, s \in \mathbb{R}$. For $r = s = 1$ we get that U is closed under vector addition. For $s = 0$ we get that U is closed under scalar multiplication. Hence U is a subspace of V.

Conversely, suppose that $U \leq V$, then U is closed under scalar multiplication and vector addition. Thus for any $r, s \in \mathbb{R}$ and $u, w \in U$ we have that ru and sw are in U which means their sum $ru + sw \in U$. $\square$

Definition 1.2.3. Let V be a vector space and let U and W be two subspaces of V. The **intersection of subspaces** U and W is the set

$$U \cap W = \{v \in V : v \in U \text{ and } v \in W\}$$

and the **sum of subspaces** U and W is the set

$$U + W = \{v \in V : v = u + w,\ u \in U, \text{ and } w \in W\}.$$

If $U \cap W = \{0\}$, then the sum $U + W$ is called the **direct sum of subspaces** and is denoted $U \oplus W$.

In the case where $V = U \oplus W$, W is known as the **complementary subspace** of U.

Theorem 1.2.4. *Let V be a vector space, and let U and W be two subspaces of V. The intersection $U \cap W$ and the sum $U + W$ are also subspaces of V.*

Proof. To do as exercise. $\square$

Theorem 1.2.5. *Let V be a vector space, and let U and W be two subspaces of V. Then $V = U \oplus W$ if and only if every $v \in V$ can be uniquely written as a sum $v = u + w$ with $u \in U$ and $w \in W$.*

Proof. Suppose that $V = U \oplus W$. If there exists $u', u \in U$ and $w', w \in W$ such that $v = u' + w'$ and $v = u + w$, then

$$\begin{aligned} v = u + w = u' + w' &\implies u - u' = w' - w \in U \cap W = \{0\} \\ &\implies u' = u \text{ and } w' = w. \end{aligned}$$

Hence the decomposition $v = u + w$ is unique.

For the converse, suppose that any vector $v \in V$ can be uniquely written as a sum $v = u + w$ with $u \in U$ and $w \in W$. It is obvious that $V = U + W$ and therefore we must only prove that $U \cap W = \{0\}$. Assume for contradiction that this is not true, that is $U \cap W$ contains some vector x. Then by the uniqueness of the decomposition of x we can write $x = x + 0$ where $x \in U$ and $x = 0 + x$ where $x \in V$, but this implies $x = 0$. Thus we have a contradiction, giving that $U \cap W = \{0\}$. □

Example 1.2.6. Consider the vector space $\mathbb{R}^3$ and its subsets

$$\begin{aligned} V &= \{(x, y, z) \in \mathbb{R}^3 : z = 0\} \\ U &= \{(x, y, z) \in \mathbb{R}^3 : x = 0\} \\ W &= \{(x, y, z) \in \mathbb{R}^3 : x = y = 0\}. \end{aligned}$$

We can see that each of these subsets are in fact subspaces of $\mathbb{R}^3$ since they are closed under the two operations.

Looking at the intersections we get

$$\begin{aligned} V \cap U &= \{(x, y, z) \in \mathbb{R}^3 : x = z = 0\} \\ V \cap W &= \{(x, y, z) \in \mathbb{R}^3 : x = y = z = 0\} = \{0\} \\ U \cap W &= \{(x, y, z) \in \mathbb{R}^3 : x = y = 0\} = W. \end{aligned}$$

Following the above and that $V + W = \mathbb{R}^3$ we see that $V \oplus W = \mathbb{R}^3$. Furthermore, as you will prove since $U \cap W = W$ we know that U is a subspace of W. △

1.2.1 Exercises

1. Explain why the following subsets of $\mathbb{R}^2$ are not subspaces of $\mathbb{R}^2$:

   (a) $V_1 = \{(x, y) \in \mathbb{R}^2 : x \geq y\}$,

   (b) $V_2 = \{(x, y) \in \mathbb{R}^2 : y^2 = x\}$,

   (c) $V_3 = \{(x, y) \in \mathbb{R}^2 : x + y = 1\}$.

2. Prove Theorem 1.2.4.

3. Prove that the sets

$$\begin{aligned} V &= \{(x, y, z) \in \mathbb{R}^3 \mid x = 2y\}, \\ U &= \{(x, y, z) \in \mathbb{R}^3 \mid x = -3z\} \end{aligned}$$

   are subspaces of $\mathbb{R}^3$. Then, describe their intersection $V \cap U$.

1.2.2 Solutions

1. Explain why the following subsets of $\mathbb{R}^2$ are not subspaces of $\mathbb{R}^2$:

   (a) $V_1 = \{(x, y) \in \mathbb{R}^2 : x \geq y\}$,
   Let $v_1 = (2, 1)$ and $v_2 = (2, -1)$ which are both in V_1. Now $v_1 - v_2 = (0, 2)$ which is not in V_1 so it is not closed under vector addition.
   Another way to solve this is just to take v_2 and consider its additive inverse.

   (b) $V_2 = \{(x, y) \in \mathbb{R}^2 : y^2 = x\}$,
   Let $v_1 = v_2 = (1, 1)$ which are both in V_2, then $v_1 + v_2 = (2, 2) \notin V_2$. Thus V_2 is not closed under addition.

   (c) $V_3 = \{(x, y) \in \mathbb{R}^2 : x + y = 1\}$.
   Let $v = (2, -1) \in V_3$, then $0v = (0, 0) \in V_3$. Thus V_3 is not closed under scalar multiplication.
   You can also argue that V_3 is missing the zero vector.

2. Prove Theorem 1.2.4.

   Proof. Given a vector space V with field F and with subspaces U and W we want to show that the intersection and sum of U and W are subspaces of V. First consider the intersection, let $u, v \in U \cap W$ and $r, s \in F$. Now $ru + sv \in U$ since $u, v \in U$ and U is a subspace. By the same logic $ru + sv \in W$. Thus $ru + sv \in U \cap W$ giving that $U \cap W$ is a subspace.

   For the sum let $v, u \in U + W$, then we can write $v = v_1 + v_2$ and $u = u_1 + u_2$ where $v_1, u_1 \in U$ and $v_2, u_2 \in W$. Now we have

   $$rv + su = rv_1 + rv_2 + su_1 + su_2.$$

   Since U and W are both subspaces we have $rv_1 + su_1 \in U$ and $rv_2 + su_2 \in W$ giving that $rv_1 + rv_2 + su_1 + su_2 \in U + W$. $\square$

3. Prove that the sets

   $$V = \{(x, y, z) \in \mathbb{R}^3 \mid x = 2y\},$$
   $$U = \{(x, y, z) \in \mathbb{R}^3 \mid x = -3z\}$$

   are subspaces of $\mathbb{R}^3$. Then, describe their intersection $V \cap U$.

   Let $(x_1, y_1, z_1), (x_2, y_2, z_2) \in V$ and $r, s \in \mathbb{R}$, then we want to show that $r(x_1, y_1, z_1) + s(x_2, y_2, z_2) \in V$. To do this it must satisfy the restriction, which we can see from

   $$rx_1 + sx_2 = r(2y_1) + s(2y_2) = 2(ry_1 + sy_2).$$

   Thus V is a subspace of $\mathbb{R}^3$.

Similar for U let $(x_1, y_1, z_1), (x_2, y_2, z_2) \in U$ and $r, s \in \mathbb{R}$, then from the linear combination we get the condition

$$rx_1 + sx_2 = r(-3z_1) + s(-3z_2) = -3(rz_1 + sz_2).$$

When we take the intersection of subspaces defined by linear constraints we get both of their constraints. So in this case

$$V \cap U = \{(x, y, z) \in \mathbb{R}^3 \mid x = 2y \text{ and } x = -3z\}.$$

Note also that because we don't have that $V \subseteq U$ or $U \subseteq V$ the intersection is a proper subset.

1.3 Linear combinations

Definition 1.3.1. Let V be a vector space and let $v_1, v_2, \dots, v_k \in V$. For any k scalars $r_1, r_2, \dots, r_k \in \mathbb{R}$, the vector

$$v = r_1v_1 + r_2v_2 + \cdots + r_kv_k \in V$$

is called a **linear combination** of $v_1, v_2, \dots, v_k$ with **coefficients** $r_1, r_2, \dots, r_k$. The set of all linear combinations of $v_1, v_2, \dots v_k$ is denoted by

$$\text{span}\{v_1, v_2, \dots, v_k\} = \{v = r_1v_1 + r_2v_2 + \cdots + r_kv_k \in V : r_1, r_2, \dots, r_k \in \mathbb{R}\},$$

and it is called the **span** (or linear hull) of $v_1, \dots, v_k$. The vectors $v_1, \dots, v_k$ are called the **generators** of this span.

Theorem 1.3.2. *Let V be a vector space, and let $v_1, \dots, v_k \in V$. Then span$\{v_1, \dots, v_k\}$ is a subspace of V.*

Not only is the span of vectors a subspace of the original vector space, but it represents the smallest subspace that contains those vectors.

Theorem 1.3.3. *Let U and W be two subspaces of a vector space V, then their union*

$$U \cup W = \{v \in V : v \in U \text{ or } v \in W\}$$

is a subspace of V if and only if $U \leq W$ or $W \leq U$.

Proof. To do on homework. □

1.4 Linear Independence

Definition 1.4.1. Let V be a vector space, and let $v_1, v_2, \dots, v_k \in V$. The set $\{v_1, \dots, v_k\}$ is called **linearly dependent** if at least one of its vectors can be written as a linear combination of the others. Otherwise the set is called **linearly independent**.

Theorem 1.4.2. *Let V be a vector space over a field F, and let $v_1, \dots, v_k \in V$. The set $\{v_1, \dots, v_k\}$ is linearly independent if and only if it satisfies the following condition*

$$r_1 v_1 + \cdots + r_k v_k = 0 \quad \textit{for} \quad r_1, \dots, r_k \in F \implies r_1 = \cdots = r_k = 0.$$

Example 1.4.3. Lets prove that the vectors $(1,1,1), (1,1,-1), (1,-1,-1) \in \mathbb{R}^3$ are linearly independent. First we have that

$$r_1(1,1,1) + r_2(1,1,-1) + r_3(1,-1,-1) = 0.$$

This gives the equations

$$\begin{aligned} r_1 + r_2 + r_3 &= 0 \\ r_1 + r_2 - r_3 &= 0 \\ r_1 - r_2 - r_3 &= 0. \end{aligned}$$

From the first equation $r_3 = -r_1 - r_2$, substituting this into the other two gives

$$\begin{aligned} 2r_1 + 2r_2 &= 0 \\ 2r_1 &= 0. \end{aligned}$$

From the second equation we see that $r_1 = 0$, but this implies from the first equation that $r_2 = 0$. Going back to the original we have that $r_3 = 0$. Thus the only linear combination that gives zero vector is with zero coefficients giving that these vectors are linearly independent. △

Theorem 1.4.4. *Consider the vector space $\mathbb{R}^n$ and vectors $v_1, \dots, v_k \in \mathbb{R}^n$. Form the $n \times k$ matrix A whose columns are the vectors $v_1, \dots, v_k$ and compute an echelon form E of A. Then, $\{v_1, \dots, v_k\}$ is linearly independent if and only if every column of E contains a pivot.*

1.4.1 Exercises

1. Prove that the vectors $v_1 = (1,-1)$ and $v_2 = (1,2)$ span the vector space $\mathbb{R}^2$.
2. Prove that the vectors $v_1 = (1,1,1)$, $v_2 = (0,2,1)$, and $v_3 = (0,0,3)$ span the vector space $\mathbb{R}^3$.
3. Prove that $\operatorname{span}\{(1,2,3), (-1,3,2)\} = \operatorname{span}\{(1,0,1), (1,1,2)\}$.
4. Let V be a vector space, and let $u, v_1, v_2, \dots, v_k \in V$. Prove that
$$\operatorname{span}\{u, v_1, \dots, v_k\} = \operatorname{span}\{v_1, \dots, v_k\}$$
if and only if $u \in \operatorname{span}\{v_1, \dots, v_k\}$.
5. Prove that if V is a vector space with vectors $v_1, \dots, v_k$ and if $v_i = 0$ for some $i \in \{1, \dots, k\}$, then the set $\{v_1, \dots, v_k\}$ is linearly dependent.
6. Let V be a vector space and let $u, v, w \in V$. Prove that u, v, and w are linearly independent if and only if u, $u+v$, $u+v+w$ are linearly independent.

1.4.2 Solutions

1. Prove that the vectors $v_1 = (1, -1)$ and $v_2 = (1, 2)$ span the vector space $\mathbb{R}^2$.

   To show we span the vector space $\mathbb{R}^2$ we want to show that every vector in $\mathbb{R}^2$ can be achieved as a linear combination of v_1 and v_2. To do this let $(x, y) \in \mathbb{R}^2$, then for $r_1v_1 + r_2v_2 = (x, y)$ we need $r_1 + r_2 = x$ and $-r_1 + 2r_2 = y$. Solving gives that $r_1 = 2x - y$ and $r_2 = y - x$. Since these systems are consistent we know the vectors span.

2. Prove that the vectors $v_1 = (1, 1, 1)$, $v_2 = (0, 2, 1)$, and $v_3 = (0, 0, 3)$ span the vector space $\mathbb{R}^3$.

   For this problem we will take a different approach than the first one. We can instead make a matrix from the vectors which is

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

   Row reducing this gives the identity matrix which has a pivot in every row. Since the solution space stays the same through row reduction we know that these vectors span $\mathbb{R}^3$.

3. Prove that $\text{span}\{(1, 2, 3), (-1, 3, 2)\} = \text{span}\{(1, 0, 1), (1, 1, 2)\}$.

   A direct way to show that these sets are equal is to show you can make the generators from one set using linear combinations from the other. That is we can take

   $$\frac{3}{5}(1, 2, 3) - \frac{2}{5}(-1, 3, 2) = (1, 0, 1)$$
   $$\frac{4}{5}(1, 2, 3) - \frac{1}{5}(-1, 3, 2) = (1, 1, 2).$$

   Similarly for the other direction we get

   $$-(1, 0, 1) + 2(1, 1, 2) = (1, 2, 3)$$
   $$-4(1, 0, 1) + 3(1, 1, 2) = (1, 1, 2).$$

4. Let V be a vector space, and let $u, v_1, v_2, \ldots, v_k \in V$. Prove that

   $$\text{span}\{u, v_1, \ldots, v_k\} = \text{span}\{v_1, \ldots, v_k\}$$

   if and only if $u \in \text{span}\{v_1, \ldots, v_k\}$.

   Proof. If $\text{span}\{u, v_1, \ldots, v_k\} = \text{span}\{v_1, \ldots, v_k\}$, then we can write any vector in $\{u, v_1, \ldots, v_k\}$ as a linear combination of the vectors in $\{v_1, \ldots, v_k\}$. In particular, this means that

   $$u = r_1v_1 + r_2v_2 + \cdots + r_kv_k$$

which implies that $u \in \text{span}\{v_1, \dots, v_k\}$.

Conversely if, $u \in \text{span}\{v_1, \dots, v_k\}$, then we can add u to the spanning set without changing it. Thus $\text{span}\{u, v_1, \dots, v_k\} = \text{span}\{v_1, \dots, v_k\}$. $\square$

5. Prove that if V is a vector space with vectors $v_1, \dots, v_k$ and if $v_i = 0$ for some $i \in \{1, \dots, k\}$, then the set $\{v_1, \dots, v_k\}$ is linearly dependent.

   Proof. Consider the vectors $v_1, \dots, v_k$ and without loss of generality assume $v_k = 0$, then for the set $\{v_1, \dots, v_k\}$ to be linearly dependent we need to find a non-trivial linear combination to get the zero vector. To do this consider
$$0v_1 + 0v_2 + \cdots + 1v_k = 0.$$
$\square$

6. Let V be a vector space and let $u, v, w \in V$. Prove that u, v, and w are linearly independent if and only if u, $u + v$, $u + v + w$ are linearly independent.

   Proof. If u, v, and w are linearly independent, then consider the linear combination
$$r_1 u + r_2(u + v) + r_3(u + v + w) = (r_1 + r_2 + r_3)u + (r_2 + r_3)v + r_3 w = 0.$$
   Letting $s_1 = r_1 + r_2 + r_3$, $s_2 = r_2 + r_3$, and $s_3 = r_3$, then we have
$$s_1 u + s_2 v + s_3 w = 0.$$
   Following u, v, w are linearly independent we know the only solution is $s_1 = s_2 = s_3 = 0$ which gives that $r_1 = r_2 = r_3 = 0$.

   Conversely if u, $u + v$, $u + v + w$ are linearly independent, then by the same algebra from above we have that $r_1 = r_2 = r_3 = 0$ which after rearrangement gives $s_1 = s_2 = s_3 = 0$. $\square$

1.5 Basis and dimension

Definition 1.5.1. Let $V \neq \{0\}$ be a vector space, and let $\{v_1, v_2, \dots, v_n\}$ be a set of vectors in V. The set $\{v_1, \dots, v_n\}$ is called a **basis** of the vector space V if

1. the set $\{v_1, \dots, v_n\}$ is linearly independent, and
2. the vectors $v_1, \dots, v_n$ span V.

Note that for infinite vector spaces, their are infinite fields since we can scale the basis to get a new one.

Theorem 1.5.2. *Let V be a vector space, and let $v_1, \dots, v_n \in V$ be nonzero. The set $\{v_1, \dots, v_n\}$ is a basis of V if and only if every vector $v \in V$ is written as a linear combination of $v_1, \dots, v_n$ in a unique way.*

Theorem 1.5.3. *Let V be a vector space with a basis $\{v_1, v_2, \dots, v_n\}$, and let $u_1, \dots, u_m$ be linearly independent vectors of V with $m \leq n$. Then we can replace m vectors of the basis $\{v_1, \dots, v_n\}$ by the vectors $u_1, \dots, u_m$ such that the resulting set of n vectors is also a basis.*

Definition 1.5.4. Let $V \neq \{0\}$ be a vector space with a basis that has n elements, then the **dimension** of V is n. We denote this with

$$\dim(V) = n.$$

In this case V is referred to as a finite vector space.

If a basis of V requires an infinite number of vectors, then V is an infinite vector space.

By convention if $V = \{0\}$, then we say that V is zero dimensional.

Theorem 1.5.5. *Let V be a vector space with $\dim(V) = n$. The following hold:*

1. *If $m > n$, then the set $\{v_1, \dots, v_m\}$ is linearly dependent.*
2. *If $\{v_1, \dots, v_m\}$ is linearly independent, then $m \leq n$.*
3. *If $v_1, \dots, v_m \in V$ span V, then $m \geq n$.*

Theorem 1.5.6. *Let V be a vector space with $\dim(V) = n$, and let U be a subspace of V. If $\dim(U) = n$, then $U = V$.*

Theorem 1.5.7. *Let V be a finite dimensional vector space, and let U and W be two subspaces of V. Then*

$$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W).$$

Note that by the above theorem if U and V are complementary, then we can drop the difference. That is $\dim(U \oplus W) = \dim(U) + \dim(W)$.

1.6 Linear transformations

Definition 1.6.1. Given two vector spaces V and W over fields F_V and F_W a **transformation** is a function

$$T : V \to W$$

that assigns to every $v \in V$ to a vector $u \in W$. We denote this with $T(v) = u$.

A transformation is called a **linear transformation** if it satisfies

1. $T(v_1 + v_2) = T(v_1) + T(v_2)$ for every $v_1, v_2 \in V$

2. $T(cv) = cT(v)$ for every $v \in V$ and $c \in F_V$.

Definition 1.6.2. Let $T : V \to U$ be a transformation (function), then

1. The vector space V is called the **domain** and U is called the **codomain**.
2. The vector $u = T(v)$ is the **image** of v.
3. The set of all images is called the **range**, denoted $R(T)$.
4. The set of all vectors $v \in V$ such that $T(v) = 0$ is called the **nullspace** (or kernel), denoted $\text{null}(T)$.

1.6.1 Exercises

1. Are the following linear transformations
   (a) $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(u_1, u_2, u_3) = (u_1, u_2 + u_3, 0)$.
   (b) $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(u_1, u_2, u_3) = (1 + u_1, u_2, u_3)$.
   (c) $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T(u_1, u_2, u_3) = (2u_1, u_2 + u_3)$.
2. Find a basis and the dimension of the subspace of $\mathbb{R}^3$ that is spanned by the vectors $(1, -1, 0), (2, 1, 3), (1, 1, 2)$.
3. Find a basis and the dimension of the subspace of $\mathbb{R}^3$ that is spanned by the vectors $(1, 2, 1), (3, 3, 1), (1, 5, 3)$.
4. Consider the subspaces of $\mathbb{R}^3$

$$U = \{(x, y, z) \in \mathbb{R}^3 : x - 2y + z = 0\}$$
$$W = \{(x, y, z) \in \mathbb{R}^3 : 2x + 3y - z = 0\}$$

   Find a basis, and the dimension for U, W, $U + W$, and $U \cap W$.

1.6.2 Solutions

1. Are the following linear transformations
   (a) $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(u_1, u_2, u_3) = (u_1, u_2 + u_3, 0)$.
   Yes. To see this, we need to show that the transformation satisfies the two properties. First let $(v_1, v_2, v_3), (u_1, u_2, u_3) \in \mathbb{R}^3$ and $r \in \mathbb{R}$, then

$$T((v_1, v_2, v_3)+(u_1, u_2, u_3)) = T((v_1+u_1, v_2+u_2, v_3+u_3)) = (v_1+u_1, v_2+u_2+v_3+u_3, 0)$$

   which is equal to

$$T((v_1, v_2, v_3))+T((u_1, u_2, u_3)) = (v_1, v_2+v_3, 0)+(u_1, u_2+u_3, 0) = (v_1+u_1, v_2+u_2+v_3+u_3, 0).$$

Similarly for scalar multiplication

$$T(r(v_1, v_2, v_3)) = T((rv_1, rv_2, rv_3)) = (rv_1, rv_2 + rv_3)$$

which is equal to

$$rT((v_1, v_2, v_3)) = (rv_1, rv_2 + rv_3).$$

(b) $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(u_1, u_2, u_3) = (1 + u_1, u_2, u_3)$.
No, consider $(1, 1, 1)$ and the scalar 0, then $T(0(1, 1, 1)) = T((0, 0, 0)) = (1, 0, 0)$ while $0T((1, 1, 1)) = 0(2, 1, 1) = (0, 0, 0)$. Since these are not equal, we have that T is not linear.
Note, you can also say that for a transformation to be linear it must map the additive identity to the additive identity, which in this case it does not. But be careful with this, the output space does have the identity it just comes from $(-1, 0, 0)$.

(c) $T : \mathbb{R}^3 \to \mathbb{R}^2$ given by $T(u_1, u_2, u_3) = (2u_1, u_2 + u_3)$.
Let $u = (u_1, u_2, u_3)$ and $v = (v_1, v_2, v_3)$ be in $\mathbb{R}^3$, then

$$T(u+v) = T((u_1+v_1, u_2+v_2, u_3+v_3)) = (2u_1+2v_1, u_2+v_2+u_3+v_3)$$

which is equal to

$$T(u) + T(v) = (2u_1 + 2v_1, u_2 + v_2 + u_3 + v_3).$$

For scalar multiplication we have

$$T(ru) = (2ru_1, ru_2 + ru_3)$$

which is equal to

$$rT(u) = (2ru_1, ru_2 + ru_3).$$

2. Find a basis and the dimension of the subspace of $\mathbb{R}^3$ that is spanned by the vectors $(1, -1, 0), (2, 1, 3), (1, 1, 2)$.

One way to approach this problem is to place the vectors into a matrix,

$$A = \begin{bmatrix} 1 & 2 & 1 \\ -1 & 1 & 1 \\ 0 & 3 & 2 \end{bmatrix},$$

then we row reduce the matrix to echelon form yielding

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

Since we have pivots in 2 row we know that the dimension is 2. To find the basis, we now need to pick two vectors from the list that are linearly independent. In this case, any 2 will do. So a basis is $(1, -1, 0), (2, 1, 3)$.

3. Find a basis and the dimension of the subspace of $\mathbb{R}^3$ that is spanned by the vectors $(1, 2, 1), (3, 3, 1), (1, 5, 3)$.

   Approaching this problem just like the last gives the matrix

$$\begin{bmatrix} 1 & 3 & 1 \\ 2 & 3 & 5 \\ 1 & 1 & 3 \end{bmatrix}$$

   which reduces to

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

   Again we have the dimension is 2, and we can pick any two vectors for the basis. One option would be $(3, 3, 1), (1, 5, 3)$.

4. Consider the subspaces of $\mathbb{R}^3$

$$\begin{aligned} U &= \{(x, y, z) \in \mathbb{R}^3 : x - 2y + z = 0\} \\ W &= \{(x, y, z) \in \mathbb{R}^3 : 2x + 3y - z = 0\} \end{aligned}$$

   Find a basis, and the dimension for U, W, $U + W$, and $U \cap W$.

   To find a basis we can rewrite the constraint equations. So for U we have

$$x - 2y + z = 0 \implies x - 2y = -z.$$

   Now we can find the basis vectors $(1, 0, -1)$ and $(0, -2, -1)$ by setting either $x = 1, y = 0$ or $x = 0, y = 1$. Similar for W gives the vectors $(2, 0, 1)$ and $(0, 3, 1)$. Thus, the dimension of U and W are both 2.

   To find a basis for $U + W$ we can simply take all the basis vectors together and find the linear independent subset. That is consider

$$\operatorname{span}((1, 0, -1), (0, -2, -1), (2, 0, 1), (0, 3, 1)).$$

   From this we can see that we can reach all of $\mathbb{R}^3$, thus the dimension is 3 and any 3 linearly independent vectors will work for a basis like $(1, 0, 0), (0, 1, 0), (0, 0, 1)$.

   Finally, for $U \cap W$ we are taking both constraints simultaneously. So to find our basis vectors we can substitute $z = 2x + 3y$ from W into $x - 2y + z = 0$ from U giving

$$x - 2y + (2x + 3y) = 0 \implies 3x + y = 0 \implies 3x = -y.$$

   Now setting $x = 3$ forces $y = -1$ and $z = -5$. So the only basis vector is $(3, -1, -5)$. Giving the dimension is 1.

1.7 Isomorphisms

Definition 1.7.1. Let V and W be vector spaces with $v_1, v_2 \in V$. Then a linear transformation $T : V \to W$ is called **one-to-one** if whenever $v_1 \neq v_2$ we have that $T(v_1) \neq T(v_2)$.

The linear transformation is called **onto** if for all $w \in W$ there exists a $v \in V$ such that $T(v) = w$.

Theorem 1.7.2. *Let V and W be vector spaces. Then a linear transformation $T : V \to W$ is one to one if and only $Null(T) = \{0\}$.*

Definition 1.7.3. A linear transformation is called a **bijection** if it is one-to-one and onto.

Definition 1.7.4. Let V and W be vector spaces, then V and W are called **isomorphic** if there exists a bijective linear transformation T between them. This is denoted with $V \cong W$. This linear transformation T is called an **isomorphism** from V to W.

Theorem 1.7.5. *Let $T : V \to W$ be an isomorphism. Then $T^{-1} : W \to V$ is an isomorphism.*

Theorem 1.7.6. *Let V and W be vector spaces, let $T : V \to W$ be an isomorphism, and let $v_1, \ldots, v_n$ be a basis for V. Then $T(v_1), \ldots, T(v_n)$ is a basis for W.*

Theorem 1.7.7. *Suppose V and W are vector spaces over the same field. Then $V \cong W$ if and only if $\dim(V) = \dim(W)$.*

Furthermore if $\dim(V) = \dim(W)$ and $T : V \to W$ is a linear transformation, then the following are equivalent

1. *T is one-to-one,*
2. *T is onto, and*
3. *T is an isomorphism.*

This result allows us to always work over F^n for some field F as our vector space.

1.7.1 Exercises

1. Prove that polynomials of degree less than or equal to 3 with real coefficients are isomorphic to $\mathbb{R}^4$.

2. Can you have $\mathbb{R}^3$ and $\mathbb{R}^4$ be isomorphic?

1.7.2 Solutions

1. Prove that polynomials of degree less than or equal to 3 with real coefficients are isomorphic to $\mathbb{R}^4$.

   Proof. Let $p(x) = ax^3 + bx^2 + cx + d$ and define the transformation

   $$T(p) = (a, b, c, d).$$

   Following polynomial addition and scalar multiplication is done elementwise and T is the identity mapping between these two spaces, we have that it is a linear transformation.

   Let p be defined above and $q(x) = ux^3 + vx^2 + sx + t$, then if $T(p) \neq T(q)$ we have that $(a, b, c, d) \neq (u, v, s, t)$ which gives that $p \neq q$. Thus T is one-to-one.

   Let (a, b, c, d) be some vector in $\mathbb{R}^4$, then we can build $p(x) = ax^3 + bx^2 + cx + d$ such that $T(p) = (a, b, c, d)$. Thus T is onto.

   Since T is a linear transformation that is a bijection, we have that the two spaces are isomorphic. $\square$

2. Can you have $\mathbb{R}^3$ and $\mathbb{R}^4$ be isomorphic?

   No, this would break Theorem 1.7.7.

Chapter 2

Eigenvalues

2.1 Determinants

Definition 2.1.1 (Laplace expansion of determinant)**.** Let $A = [a_{ij}]$ be an $n \times n$ matrix. The **determinant** of A, denoted by $\det(A)$, is defined recursively as follows. For $n = 1$ $\det(A) = a_{11}$. For $n > 1$

$$\det(A) = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \ det(A_{1j}).$$

where A_{1j} denotes the $(n-1) \times (n-1)$ matrix obtained from A by deleting row 1 and column j.

When writing out the determinant for computation, it is common to change the square brackets to straight lines. That is

$$\det \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}.$$

Example 2.1.2. For a general two by two matrix,

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

we have that

$$\det(A) = a \det([d]) - b \det([c]) = ad - bc.$$

△

Example 2.1.3. Consider

$$A = \begin{bmatrix} 1 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 3 \end{bmatrix},$$

then

$$\det(A) = 1\begin{vmatrix} 2 & 2 \\ 2 & 3 \end{vmatrix} - 2\begin{vmatrix} 2 & 2 \\ 2 & 3 \end{vmatrix} + 2\begin{vmatrix} 2 & 2 \\ 2 & 2 \end{vmatrix} = -\begin{vmatrix} 2 & 2 \\ 2 & 3 \end{vmatrix} = -2$$

△

Definition 2.1.4. Let $A \in \mathsf{M}_n$, then

1. Define A_{ij} as the $(n-1)\times(n-1)$ matrix obtained from A by deleting its ith row and the jth column, we refer to A_{ij} as a **submatrix** of A.
2. Define $M_{ij} = \det(A_{ij})$, we refer to M_{ij} as a **minor** of A.
3. Define $C_{ij} = (-1)^{i+j}M_{ij}$, we refer to C_{ij} as the (i,j)th **cofactor** of A.

Theorem 2.1.5. *Let $A = [a_{ij}]$ be an $n \times n$ matrix. For any $i = 1, 2, \ldots, n$, the determinant of A is given by its Laplace (cofactor) expansion across the ith row*

$$\det(A) = \sum_{k=1}^{n} a_{ik}C_{ik}.$$

For any $j = 1, 2, \ldots, n$, the determinant of A is given by its Laplace (cofactor) expansion down the jth column

$$\det(A) = \sum_{k=1}^{n} a_{kj}C_{kj}.$$

While theoretically useful, the Laplace expansion is computationally terrible. Generally it requires $n!$ multiplications. Using Gaussian elimination, it is possible to reduce this to only n^3 making the determinant a fairly computable quantity even for massive matrices.

When using Gaussian elimination, elementary row operations change the determinant in controlled ways. These are

1. Permute rows i, j $(i \neq j)$, then negate $\det(A)$.
2. Scale row i by r, then scale $\det(A)$ by r.
3. Add multiple of row i to row j, then no changes to $\det(A)$.

A common question (or complaint) from students is dealing with the $(-1)^{j+1}$ power. If we drop this, we get a well-defined quantity called the permanent of a matrix. However, the permanent is generally computationally infeasible as it does not share the same Gaussian elimination trick to solve it. Thus, the -1 power is actually a simplification.

Theorem 2.1.6. *Let A, B be two $n \times n$ matrices and r be a scalar. Then,*

1. $\det(AB) = \det(A)\det(B)$
2. $\det(rA) = r^n \det(A)$
3. $\det(A) = \det(A^\top)$

2.2 Invertible matrices

Definition 2.2.1. Let A be an $n \times n$ matrix. If there exists an $n \times n$ matrix B such that

$$AB = BA = I,$$

then A is called **invertible** (or **nonsingular**). The matrix B is called the **inverse** of A. If a matrix does not have an inverse then it is called **noninvertible** (or **singular**).

Theorem 2.2.2. *The inverse of an invertible matrix A is unique and denoted by A^{-1}.*

Theorem 2.2.3. *Let A, B be invertible matrices, then*

1. $(AB)^{-1} = B^{-1}A^{-1}$.
2. $(A^\top)^{-1} = (A^{-1})^\top$.
3. $\det(A^{-1}) = \frac{1}{\det(A)}$.

Theorem 2.2.4 (Invertible matrix theorem)**.** *Let $A \in \mathsf{M}_n(\mathbb{R})$, then the following statements are equivalent.*

1. *A is invertible.*
2. *The linear system $Ax = b$ has a unique solution for every $b \in \mathbb{R}^n$.*
3. *The homogeneous system $Ax = 0$ has only the trivial solution.*
4. *The reduced echelon form of A equals the identity matrix.*
5. *The columns of A form a linearly independent set.*
6. *The columns of A span $\mathbb{R}^n$.*
7. *The linear transformations $T(x) = Ax$ is a bijection.*
8. *The transpose $A^\top$ is invertible.*

Procedure 2.2.5. Computing the inverse can be done with Gaussian elimination. This procedure will fail if the matrix is not invertible. Let $A \in \mathsf{M}_n$,

1. Set up the augmented $n \times 2n$ matrix $[A \mid I]$.
2. Fully row reduce A into reduced echelon form.
3. If A does not have a pivot in every row/column, then we are left with an inconsistent system of equations and A is not invertible.
4. If A has a pivot in every row/column, then the resulting augmented matrix should be $[I \mid A^{-1}]$ giving the inverse.

△

2.2.1 Exercises

1. Compute the determinant of the following matrices

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

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 6 & 2 & 0 \end{bmatrix}, \text{ and}$$

$$C = \begin{bmatrix} 1 & 2 & 34 & 4.67 & 52 \\ 0 & 2 & 7 & 86 & 9 \\ 0 & 0 & 3 & -15 & \pi \\ 0 & 0 & 0 & 2 & -\sqrt{3} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

2. Let A and B be 2×2 matrices such that $\det(A) = -4$ and $\det(B) = 3$. Find the following quantities or explain why they cannot be found with the given information

   (a) $\det(B^3)$

   (b) $\det(A^{-1})$

   (c) $\det(AB)$

   (d) $\det(A + B)$

3. Find 2×2 matrices A, B such that

   (a) A and B are invertible, but $A + B$ is not invertible.

   (b) A and B are not invertible, but $A + B$ is invertible.

4. Find the inverse of

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

2.2.2 Solutions

1. Compute the determinant of the following matrices

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

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 3 & 0 \\ 6 & 2 & 0 \end{bmatrix}, \text{ and}$$

$$C = \begin{bmatrix} 1 & 2 & 34 & 4.67 & 52 \\ 0 & 2 & 7 & 86 & 9 \\ 0 & 0 & 3 & -15 & \pi \\ 0 & 0 & 0 & 2 & -\sqrt{3} \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

For the determinant of A note that the second row has the most zeros so it would be easiest to go across that row. To use this row we need to swap rows 1 and 2. Doing this causes us to need to take the negative of the determinant of A. This gives

$$\det(A) = 1 \begin{vmatrix} 2 & 1 \\ -3 & 7 \end{vmatrix} = 14 + 3 = 17.$$

So the answer is -17.

For B, we can see that there is a column of all zeros. If we take the transpose and use the Laplace expansion across that row, we can see that the determinant will be zero.

For C, this is a upper triangular matrix. So we will take the Laplace expansion across the first column. This gives

$$\begin{aligned} \det(C) &= 1 \begin{vmatrix} 2 & 7 & 86 & 9 \\ 0 & 3 & -15 & \pi \\ 0 & 0 & 2 & -\sqrt{3} \\ 0 & 0 & 0 & 1 \end{vmatrix} \\ &= 1(2) \begin{vmatrix} 3 & -15 & \pi \\ 0 & 2 & -\sqrt{3} \\ 0 & 0 & 1 \end{vmatrix} \\ &= 1(2)(3) \begin{vmatrix} 2 & -\sqrt{3} \\ 0 & 1 \end{vmatrix} \\ &= 1(2)(3)(2)(1) = 12. \end{aligned}$$

Note that this process works in general. So for arbitrary upper or lower triangular matrices, the determinant is just the product of the main diagonal.

2. Let A and B be 2×2 matrices such that $\det(A) = -4$ and $\det(B) = 3$. Find the following quantities or explain why they cannot be found with the given information

   (a) $\det(B^3)$
   Following, $\det(AB) = \det(A)\det(B)$ we can turn this into $\det(B^3) = \det(B)^3 = 3^3 = 27$.

   (b) $\det(A^{-1})$
   We can use one of the properties of the matrix inverse to know that $\det(A^{-1}) = \frac{1}{\det(A)} = -1/4$.

   (c) $\det(AB)$
   We have $\det(AB) = \det(A)\det(B) = -4(3) = -12$.

   (d) $\det(A + B)$
   There is no nice formula for this. To see this, consider
   $$A = \begin{bmatrix} -4 & 0 \\ 0 & 1 \end{bmatrix}, \ B = \begin{bmatrix} 4 & 0 \\ 0 & \frac{3}{4} \end{bmatrix}, \ \text{and } C = \begin{bmatrix} 9 & 3 \\ 8 & 3 \end{bmatrix}.$$
   Note that $\det(B) = \det(C) = 3$ If we take $\det(A+B) = \det \begin{bmatrix} 0 & 0 \\ 0 & \frac{7}{4} \end{bmatrix} = 0$. If instead we take $\det(A + C) = 97/3$. So we can make the determinant of the sum arbitrary.

3. Find 2×2 matrices A, B such that

   (a) A and B are invertible, but $A + B$ is not invertible.
   Let $A = I$ and $B = -I$, then $A + B = 0$ is clearly not invertible.

   (b) A and B are not invertible, but $A + B$ is invertible.
   Take
   $$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix},$$
   then we can see both A and B are not invertible since they have rows of all zeros. The sum $A + B = I$ which is invertible.

4. Find the inverse of
   $$A = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 2 \\ 1 & 0 & -1 \end{bmatrix}.$$
   Following the inverse procedure, take
   $$[A \mid I] = \begin{bmatrix} 1 & 0 & -2 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

Row reducing this augmented matrix gives

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

Now taking the right half gives the inverse as

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

2.3 Eigenvalues

Definition 2.3.1. Let $A \in \mathsf{M}_n$ and let x be a nonzero vector such that

$$Ax = \lambda x,$$

where λ is some scalar. We then call λ the **eigenvalue** of A corresponding to the **eigenvector** x. The ordered pair (λ, x) is called an **eigenpair**.

The set of all eigenvalues of A is referred to as the **spectrum** of A.

Note that the zero vector is never considered an eigenvector.

Example 2.3.2. Consider

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

then we have

$$Ax = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = 3x.$$

Thus x is an eigenvector of A with eigenvalue 3. △

For the eigenvalue equation, we can write

$$Ax = \lambda x \implies (A - \lambda I)x = 0 \implies x \in \mathrm{null}(A - \lambda I).$$

This gives the following definition.

Definition 2.3.3. Let A be an $n \times n$ matrix and suppose $Ax = \lambda x$ for some scalar λ and some nonzero vector x. Then, we refer to $\mathrm{null}(A - \lambda I)$ as the **eigenspace** of A corresponding to λ.

From the above definition, we have the following

1. All the eigenvectors of A corresponding to its eigenvalue λ belong to $\mathrm{null}(A - \lambda I)$.

2. Since the null space of a matrix is a subspace, any nonzero linear combination of eigenvectors of A corresponding to λ is also an eigenvector of A corresponding to λ.

3. We can find a basis of the eigenspace $\mathrm{null}(A - \lambda I)$ by solving the homogeneous linear system $(A - \lambda I)x = 0$ and expressing the solution set as the span of linearly independent vectors.

Example 2.3.4. Let
$$A = \begin{bmatrix} 2 & 1 & 2 & -2 \\ -1 & 4 & 2 & -2 \\ -3 & 1 & 7 & -2 \\ 1 & 1 & 2 & -1 \end{bmatrix},$$
then A has distinct eigenvalues $1, 3, 5$, where the eigenvectors for them are
$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \text{ and } \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix}.$$
Note that for this matrix there are only 3 distinct eigenvalues and the sum of the dimension of the eigenspaces is 3. △

Example 2.3.5. Let
$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$
Then A has eigenvalues $1, -1$ where -1 has eigenvectors
$$\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$
and 1 has eigenvectors
$$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$
Here we have 2 distinct eigenvalues and 2 eigenspaces each of dimension 2. △

Theorem 2.3.6. *Let $A \in \mathsf{M}_n(\mathbb{C})$ and let $\lambda \in \mathbb{C}$. The following statements are equivalent:*

1. *λ is an eigenvalue of A.*

2. *$Ax = \lambda x$ for some nonzero $x \in \mathbb{C}^n$.*

3. *$(A - \lambda I)x = 0$ has a nontrivial solution.*

4. *$A - \lambda I$ is not invertible.*

5. *$A^\top - \lambda I$ is not invertible.*

6. *λ is an eigenvalue of $A^\top$.*

Corollary 2.3.7. *Let $A \in \mathsf{M}_n$. Then A is invertible if and only if 0 is not an eigenvalue of A.*

Procedure 2.3.8 (Finding the eigenspace)**.** Let $A \in \mathsf{M}_n$

1. To find the eigenvalues of A we setup the polynomial $\det(tI - A)$. The roots of this polynomial are then the eigenvalues for the matrix.

2. For each distinct eigenvalue, λ, found in step one, solve the linear system $(A - \lambda I)x = 0$ to compute a basis of the eigenspace corresponding to λ.

$\triangle$

From the above procedure, you may ask why $\det(tI - A)$ instead of $\det(A - tI)$? From our properties of determinants we have

$$\det(A - tI) = (-1)^n \det(tI - A),$$

so they are equivalent in terms of the roots of the polynomial. However, $\det(tI - A)$ is always **monic**, meaning the coefficient of the largest degree term is 1.

From Theorem 2.3.6 we know that the eigenvalues of A and $A^\top$ are equivalent. However, their eigenvectors are generally **not** equivalent. This leads to the following definition.

Definition 2.3.9. Let λ be an eigenvalue of an $n \times n$ matrix A. Let y be an eigenvector of $A^\top$ corresponding to λ,

$$A^\top y = \lambda y \implies y^\top A = \lambda y^\top.$$

We refer to the eigenvector y of $A^\top$ as a **left eigenvector** of A corresponding to λ. We will also refer to its eigenspace as the **left eigenspace** of A.

2.3.1 Exercises

1. Confirm that $\lambda = 2$ is an eigenvalue of

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

and compute a basis of its eigenspace.

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

 (a) Compute the eigenvalues of A.

 (b) Compute bases for each of the eigenspaces of A.

3. Let A be an $n \times n$ **nilpotent matrix**, which means that $A^m = 0$ for some positive integer m.

 (a) Show that every eigenvalue of A is equal to 0.

 (b) Find, if possible, a 2×2 nilpotent matrix with exactly one nonzero entry.

 (c) Find, if possible, a 2×2 nilpotent matrix all of whose entries are nonzero.

4. Let A be an $n \times n$ **idempotent matrix**, which means that $A^2 = A$.

 (a) Show that every eigenvalue of A is either equal to 0 or 1.

 (b) Find, if possible, a 2×2 idempotent matrix all of whose entries are nonzero.

2.3.2 Solutions

1. Confirm that $\lambda = 2$ is an eigenvalue of
$$A = \begin{bmatrix} 1 & -1 & 5 \\ 1 & 3 & -4 \\ 3 & 3 & 6 \end{bmatrix}$$
and compute a basis of its eigenspace.

To confirm that 2 is an eigenvalue of A we want to find a vector x such that $Ax = 2x$. To do this we compute $(Ax - 2I)x = 0$ this gives the following system of equations
$$\begin{bmatrix} -1 & -1 & 5 \\ 1 & 1 & -4 \\ 3 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \implies \begin{aligned} -x_1 - x_2 + 5x_3 &= 0 \\ x_1 + x_2 - 4x_3 &= 0 \\ 3x_1 + 3x_2 + 4x_3 &= 0 \end{aligned}$$
Solving these gives the vector $x = \begin{bmatrix} 1 & -1 & 0 \end{bmatrix}^\top$.

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

(a) Compute the eigenvalues of A.

To compute the eigenvalues, we compute $\det(tI - A)$. To do this note that a Laplace expansion across the bottom row has the most zeros. This gives

$$\det(tI-A) = \begin{vmatrix} t+1 & 0 & 1 \\ -3 & t-4 & 1 \\ 0 & 0 & t-2 \end{vmatrix} = (t-2)\begin{vmatrix} t+1 & 0 \\ -3 & t-4 \end{vmatrix} = (t-2)(t+1)(t-4)$$

Since this is already factored we can see the eigenvalues are $t = 2$, $t = -1$, and $t = 4$.

(b) Compute bases for each of the eigenspaces of A.

Since we have three distinct eigenvalues, we will have that each eigenvalue has a dimension 1 eigenspace. To see this take $(A - 2I)x = 0$ which gives

$$\begin{bmatrix} -3 & 0 & 1 \\ -3 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \implies \begin{aligned} -3x_1 + x_3 &= 0 \\ -3x_1 + 2x_2 + x_3 &= 0 \\ 0 &= 0. \end{aligned}$$

From this we get $x_1 = x_3/3$ giving the eigenvector $x = \begin{bmatrix} 1 & 0 & 3 \end{bmatrix}^\top$.

Doing the same for $t = -1$ gives

$$\begin{bmatrix} 0 & 0 & 1 \\ -3 & 5 & 1 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \implies \begin{aligned} x_3 &= 0 \\ -3x_1 + 5x_2 + x_3 &= 0 \\ 3x_3 &= 0. \end{aligned}$$

This gives $x_3 = 0$ and $x_1 = 5x_2/3$.

Finally for $t = 4$ we have

$$\begin{bmatrix} -5 & 0 & 1 \\ -3 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \implies \begin{aligned} -5x_1 + x_3 &= 0 \\ -3x_1 + x_3 &= 0 \\ 3x_3 &= 0. \end{aligned}$$

This gives the vector $x = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^\top$.

3. Let A be an $n \times n$ **nilpotent matrix**, which means that $A^m = 0$ for some positive integer m.

(a) Show that every eigenvalue of A is equal to 0.

Let (λ, x) be a eigenpair for the matrix A, then

$$Ax = \lambda x \implies A^2 x = \lambda^2 x \implies A^m x = \lambda^m x \implies 0 = \lambda^m x.$$

Following $x \neq 0$ we must have that $\lambda = 0$.

(b) Find, if possible, a 2×2 nilpotent matrix with exactly one nonzero entry.

Consider $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. We know that the eigenvalues of a diagonal matrix are the entries along the main diagonal, so the eigenvalues of A are zero and we have a non-zero element. To see that A is really nilpotent take $A^2 = 0$.

(c) Find, if possible, a 2×2 nilpotent matrix all of whose entries are nonzero.

Consider
$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$
Then
$$p_A(t) = t^2 - (ad - bc)t + (a + d).$$
For the roots of this to both be zero we need $a+d = 0$ and $ad-bc = 0$. This gives that $a = -d$ and $-a^2 = bc$. Thus, one option is
$$A = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}.$$

4. Let A be an $n \times n$ **idempotent matrix**, which means that $A^2 = A$.

(a) Show that every eigenvalue of A is either equal to 0 or 1.

Let (λ, x) be an eigenpair of A, then
$$Ax = \lambda x \implies A^2 x = \lambda^2 x \implies Ax = \lambda^2 x \implies \lambda = \lambda^2.$$
The only numbers that equal their own square are 0 and 1.

(b) Find, if possible, a 2×2 idempotent matrix all of whose entries are nonzero.

One way to approach constructing this is to start by a matrix all of whose entries are equal. That is
$$A = \begin{bmatrix} a & a \\ a & a \end{bmatrix},$$
then
$$A^2 = \begin{bmatrix} 2a^2 & 2a^2 \\ 2a^2 & 2a^2 \end{bmatrix}.$$
This gives we need to pick a such that $a = 2a^2 \implies a = \frac{1}{2}$. Thus the matrix
$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$
works.

2.4 Characteristic polynomial

Definition 2.4.1. Let A be an $n \times n$ matrix. The n-degree polynomial $p_A(t) = \det(tI - A)$ is called the **characteristic polynomial** of A. The equation $p_A(t) = 0$ is called the **characteristic equation** of A.

Lemma 2.4.2. *Let (λ, x) be an eigenpair of $A \in \mathsf{M}_n$ and let p be a polynomial. Then,*

$$p(A)x = p(\lambda)x,$$

that is, $(p(\lambda), x)$ is an eigenpair of $p(A)$.

Theorem 2.4.3. *Let $A \in \mathsf{M}_n$ and let p be a polynomial such that $p(A) = 0$, then*

1. *Every eigenvalue of A is a root of $p(z) = 0$.*
2. *A has at most n different eigenvalues.*

Definition 2.4.4. Let $A \in \mathsf{M}_n$. The **geometric multiplicity** of λ as an eigenvalue of A is the dimension of the eigenspace for λ.

The multiplicity of λ as a root of the characteristic polynomial $p_A(t)$ is called the **algebraic multiplicity** of λ.

Example 2.4.5. Let

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

We can see that the two eigenvalues are -1 and 1. The characteristic polynomial is

$$p_A(\lambda) = (1 - \lambda)^2(\lambda - 1)^2,$$

so we can see that the algebraic multiplicities of both eigenvalues is 2.

The geometric multiplicity of -1 is 2, since we can have the eigenvectors

$$\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}.$$

The geometric multiplicity of 1 is 1. To see this consider

$$(A - I)x = 0 \implies \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 0.$$

This gives $x_2 = 0$, $x_3 = 0$, and $x_4 = 0$ with x_1 being a free variable. Thus, the only eigenvector is $\begin{bmatrix} 1 & 0 & 0 & 0 \end{bmatrix}^\top$. $\triangle$

Since finding eigenvalues of a matrix involves finding the roots of a polynomial, it follows that there is no general analytic formula past $n = 4$. However, in practice numerical methods for root finding are excellent, so it is computationally not bad to find the eigenvalues even for large matrices accurately and in a short time.

2.4.1 Exercises

1. Let
$$A = \begin{bmatrix} 3 & 2 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & -1 & 5 \end{bmatrix}.$$

   (a) Compute the eigenvalues of A.

   (b) What are the algebraic and geometric multiplicities of each eigenvalue?

2. Let
$$A = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 7 & 2 \\ -2 & 3 & 4 \end{bmatrix}$$

   (a) Find the fully expanded characteristic polynomial of A.

   (b) Find the trace and determinant of A.

   (c) What do you notice between parts (a) and (b)?

2.4.2 Solutions

1. Let
$$A = \begin{bmatrix} 3 & 2 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & -1 & 5 \end{bmatrix}.$$

   (a) Compute the eigenvalues of A.

   The eigenvalues are 3 and 5. To see this, take the determinant down the first column.

$$\begin{aligned}
\det(tI-A) &= \begin{vmatrix} t-3 & -2 & 0 & 0 \\ 0 & t-3 & 0 & 0 \\ 0 & 0 & t-5 & 0 \\ 0 & 0 & -1 & t-5 \end{vmatrix} \\
&= (t-3)\begin{vmatrix} t-3 & 0 & 0 \\ 0 & t-5 & 0 \\ 0 & 1 & t-5 \end{vmatrix} \\
&= (t-3)^2 \begin{vmatrix} t-5 & 0 \\ -1 & t-5 \end{vmatrix} \\
&= (t-3)^2(t-5)^2.
\end{aligned}$$

(b) What are the algebraic and geometric multiplicities of each eigenvalue?

The algebraic multiplicity for each is 2 since that is the power in the factored characteristic polynomial.

For each of the two eigenvalues, our options are 1 or 2 for the geometric multiplicity. For $t = 3$ consider

$$\operatorname{rank}(3I-A) = \operatorname{rank}\left(\begin{bmatrix} 0 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & -1 & -2 \end{bmatrix}\right) = 3$$

Thus by rank nullity theorem we have $\operatorname{null}(3I-A) = 1$.

A similar computation shows that $\operatorname{null}(5I-A) = 1$. Thus, the geometric multiplicity is 1 for both eigenvalues.

2. Let

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

(a) Find the fully expanded characteristic polynomial of A.

The characteristic polynomial is

$$p_A(t) = \det(tI-A) = \begin{vmatrix} t-1 & -3 & 0 \\ 0 & t-7 & -2 \\ 2 & -3 & t-4 \end{vmatrix} = t^3 - 12t^2 + 33t - 10.$$

(b) Find the trace and determinant of A.

The trace of A is $1+7+4 = 12$ and the determinant is 10.

(c) What do you notice?

The trace of 12 and determinant of 10 are both coefficients in the characteristic polynomial (with a negative). This holds in general, that the trace is the first (non-monic) coefficient of the characteristic polynomial while the determinant is the last.

2.5 Similarity transformations

Definition 2.5.1. An $n \times n$ matrix A is **similar** to an $n \times n$ matrix B if there exists an invertible matrix S such that

$$A = SBS^{-1}.$$

Similarity between matrices forms an equivalence relation. That is

1. Every $n \times n$ matrix is similar to itself using $S = I$.
2. If A is similar to B, then B is similar to A since $A = SBS^{-1}$ implies $B = S^{-1}AS$.
3. If A is similar to B and B is similar to C, then A is similar to C since $A = SBS^{-1}$ and $B = RCR^{-1}$ implies $A = (SR)C(SR)^{-1}$.

Theorem 2.5.2. *If $A, B \in \mathsf{M}_n$ are similar, then*

1. *The characteristic polynomials of A and B are the same.*
2. *The eigenvalues of A and B are the same.*

Note that the converse of the above statement is not true. That is, we can find two matrices with the same eigenvalues that are not similar.

Example 2.5.3. Let

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

We can see that the eigenvalues of A and B are both 0. However, they are not similar. To show this, assume that they are that is

$$A = SBS^{-1}$$

for some invertible S, then $SBS^{-1} = S0S^{-1} = 0$ implying that $A = 0$ which is a contradiction. △

2.6 Diagonalization

Definition 2.6.1. An $n \times n$ matrix A is called **diagonalizable** if A is similar to a diagonal matrix.

Note that not all matrices are diagonalizable.

Example 2.6.2. Consider

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

Then following A is upper triangular, we have that the eigenvalues are all zero. However, A is not triangular since that would imply it is similar to 0, which we have shown in a previous example can't happen. △

Theorem 2.6.3. *Let $A \in \mathsf{M}_n$. Then A is diagonalizable if and only if A has n linearly independent eigenvectors $s_1, s_2, \ldots, s_n$. If this is the case, then the matrix*

$$S = [s_1 \mid s_2 \mid \cdots \mid s_n],$$

is invertible and $S^{-1}AS = D$, where

$$D = diag(\lambda_1, \ldots, \lambda_n)$$

with $\lambda_1, \ldots, \lambda_n$ being the eigenvalues of A corresponding to $s_1, \ldots, s_n$.

Corollary 2.6.4. *Let $A \in \mathsf{M}_n$. Then A is diagonalizable if and only if for any eigenvalue of A the algebraic and geometric multiplicities are equal.*

Theorem 2.6.5. *Let A be an $n \times n$ matrix that has n distinct eigenvalues. Then A is diagonalizable.*

Note that the above theorem is sufficient, but not necessary. To see this, the identity matrix is trivially diagonalizable and has all its eigenvalues being 1.

Example 2.6.6. Let

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

then the eigenpairs of A are

$$(10, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}), \ (-2-2i, \begin{bmatrix} 1 \\ i \\ -1 \\ -i \end{bmatrix}), \ (-2, \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}), \quad \text{and} \quad (-2+2i, \begin{bmatrix} 1 \\ -i \\ -1 \\ i \end{bmatrix}).$$

With this we can then diagonalize A by using the eigenvectors as the columns of S giving

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} 10 & & & \\ & -2-2i & & \\ & & -2 & \\ & & & -2+2i \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$$

$\triangle$

2.6.1 Exercises

1. Find a matrix similar to
$$A = \begin{bmatrix} 2 & 3 & 4 \\ 10 & 2 & -1 \\ -5 & 2 & 2 \end{bmatrix}.$$

2. Give the diagonalization of
$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

3. If A is diagonalizable, $A = SDS^{-1}$, and p a polynomial, then prove
$$p(A) = Sp(D)S^{-1}.$$

2.6.2 Solutions

1. Find a matrix similar to
$$A = \begin{bmatrix} 2 & 3 & 4 \\ 10 & 2 & -1 \\ -5 & 2 & 2 \end{bmatrix}.$$

   This problem is very open, but one nice class of similarity transforms is that of permutation matrices. Consider
$$P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$
   then first note that $P^2 = I$ giving that $P^{-1} = P$. Our similarity transform is then
$$PAP = \begin{bmatrix} 10 & 2 & -1 \\ 2 & 3 & 4 \\ -5 & 2 & 2 \end{bmatrix} P = \begin{bmatrix} 2 & 10 & -1 \\ 3 & 2 & 4 \\ 2 & -5 & 2 \end{bmatrix}.$$

2. Give the diagonalization of

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

Given the matrix:

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

We seek an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

We solve the characteristic equation $\det(A - \lambda I) = 0$

$$\begin{aligned} \det(A - \lambda I) &= \det \begin{bmatrix} -\lambda & 1 & 1 \\ 1 & -\lambda & 1 \\ 1 & 1 & -\lambda \end{bmatrix} \\ &= -\lambda(\lambda^2 - 1) + (\lambda + 1) + (1 + \lambda) \\ &= -\lambda^3 + \lambda + 2\lambda + 2 \\ &= -(\lambda^3 - 3\lambda - 2) \end{aligned}$$

Factoring the polynomial, we observe that $\lambda = -1$ is a root

$$-(\lambda + 1)^2(\lambda - 2) = 0$$

The eigenvalues are $\lambda_1 = -1$ (algebraic multiplicity 2) and $\lambda_2 = 2$ (algebraic multiplicity 1).

For $\lambda_1 = -1$ We calculate the null space of $(A - (-1)I) = (A + I)$:

$$A + I = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

This corresponds to the equation $x_1 + x_2 + x_3 = 0$. We select two linearly independent vectors satisfying this condition:

$$\mathbf{v}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

For $\lambda_2 = 2$ We calculate the null space of $(A - 2I)$

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

Row reducing this matrix leads to $x_1 = x_2 = x_3$. We select the vector:

$$\mathbf{v}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

We construct matrix P using the eigenvectors as columns, and matrix D using the corresponding eigenvalues:

$$P = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

Thus, the diagonalization is given by $A = PDP^{-1}$.

3. If A is diagonalizable, $A = SDS^{-1}$, and p a polynomial, then prove

$$p(A) = Sp(D)S^{-1}.$$

Proof. Let $A = SDS^{-1}$ and $p(x) = \sum_{k=0}^{m} a_k x^k$, then

$$\begin{aligned} p(A) &= p(SDS^{-1}) \\ &= \sum_{k=0}^{m} a_k (SDS^{-1})^k \\ &= \sum_{k=0}^{m} a_k SD^k S^{-1} \\ &= S\left(\sum_{k=0}^{m} a_k D^k\right) S^{-1} \\ &= Sp(D)S^{-1}. \end{aligned}$$

□

2.7 Caley-Hamilton theorem

Definition 2.7.1. Given an $n \times n$ matrix A, the **adjoint** (or **adjugate**) of A is the $n \times n$ matrix

$$\operatorname{adj}(A) = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix},$$

where $C_{ij} = (-1)^{i+j} \det(A_{ij})$ us the (i, j)th cofactor of A.

Lemma 2.7.2. *Let $A \in \mathsf{M}_n$, then*

$$A \; adj(A) = adj(A)A = \det(A)I_n.$$

Theorem 2.7.3 (Caley-Hamilton Theorem)**.** *Let $A \in \mathsf{M}_n$ with characteristic polynomial p_A, then $p_A(A) = 0$.*

Proof. Apply Lemma 2.7.2 to the matrix $tI_n - A$, it follows that

$$(tI_n - A)\mathrm{adj}(tI_n - A) = \det(tI_n - A) = p_A(t)I_n. \tag{2.1}$$

The entries of the adjoint matrix $\mathrm{adj}(tI_n - A)$ are polynomial in t of degree smaller than or equal to $n-1$, which we will denote $p_{ij}(t)$. That is we can write

$$\mathrm{adj}(tI_n - A) = \begin{bmatrix} p_{11}(t) & p_{21}(t) & \cdots & p_{n1}(t) \\ p_{12}(t) & p_{22}(t) & \cdots & p_{n2}(t) \\ \vdots & \vdots & \ddots & \vdots \\ p_{1n}(t) & p_{2n}(t) & \cdots & p_{nn}(t) \end{bmatrix}.$$

Now we can collect the coefficients matrices, B_i, in front of each power of t giving that

$$\mathrm{adj}(tI_n - A) = B_{n-1}t^{n-1} + B_{n-2}t^{n-2} + \cdots + B_1 t + B_0.$$

This allows us to rewrite equation 2.1 as

$$(tI_n - A)(B_{n-1}t^{n-1} + B_{n-2}t^{n-2} + \cdots + B_1 t + B_0) = (t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0)I_n.$$

Matching the powers of t on both sides gives the following equations

$$\begin{aligned} B_{n-1} &= I_n \\ B_{n-2} - AB_{n-1} &= a_{n-1}I_n \\ B_{n-3} - AB_{n-2} &= a_{n-2}I_n \\ &\vdots \\ B_0 - AB_1 &= a_1 I_n \\ -AB_0 &= a_0 I_n. \end{aligned}$$

Now we can multiply each of these equations by a decreasing power of A^n, so the first one by A^n, second by A^{n-1}, and so on. This is equivalent to plugging in A for t. This gives

$$\begin{aligned} A^n B_{n-1} &= A^n \\ A^{n-1}B_{n-2} - A^n B_{n-1} &= a_{n-1}A^{n-1} \\ A^{n-2}B_{n-3} - A^{n-1}B_{n-2} &= a_{n-2}A^{n-2} \\ &\vdots \\ AB_0 - A^2 B_1 &= a_1 A \\ -AB_0 &= a_0 I_n. \end{aligned}$$

Adding all of these equations back together, we can see that the left side becomes 0 and the right becomes $p_A(A)$. ☐

Corollary 2.7.4. *Let $A \in \mathsf{M}_n$, then for any $k \geq n$ we can write A^k as a linear combination of $A^0, A, \dots, A^{n-1}$.*

Proof. By Caley-Hamilton we have

$$\sum_{k=0}^{n} a_k A^k = 0 \iff A^n = \sum_{k=0}^{n-1} -a_k A^k.$$

Multiply both sides by A giving

$$\begin{aligned} A^{n+1} &= \sum_{k=0}^{n-1} -a_k A^{k+1} \\ &= -a_{n-1}A^n + \sum_{k=0}^{n-2} -a_k A^{k+1} \\ &= -a_{n-1}\sum_{k=0}^{n-1} -a_k A^k + \sum_{k=0}^{n-2} -a_k A^{k+1} \end{aligned}$$

Rearranging gives A^{n+1} as a linear combination of $A^0, A^1, \dots, A^{n-1}$. Repeating this process gives A^k for any $k \geq n+1$. ☐

Example 2.7.5. For the matrix

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

compute the matrix A^{200}.

First, we calculate the characteristic polynomial of A giving

$$p_A(t) = \det(tI_3 - A) = (t-1)^2(t-2) = t^3 - 4t^2 + 5t - 2.$$

Now we want to know if we can diagonalize A. To do this, we compute the geometric multiplicities of the eigenvalue $t = 1$ (since $t = 2$ is gauranteed to be 1), we can take

$$\operatorname{rank}(A - I_3) = \operatorname{rank}\left(\begin{bmatrix} -2 & 0 & 1 \\ -2 & 1 & 1 \\ -4 & 0 & 2 \end{bmatrix}\right) = \operatorname{rank}\left(\begin{bmatrix} -2 & 0 & 1 \\ -2 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}\right) = 2.$$

Thus the dimension of the null space is 1 giving that the algebraic and geometric multiplicities do not match. Since the algebraic and geometric multiplicities are not equal for $t = 1$ we can't diagonalize this matrix. Thus, we need to use the characteristic polynomial.

Consider the long division of t^{200} by $p_A(t)$:

$$t^{200} = p_A(t)q(t) + r(t) = p_A(t)q(t) + (at^2 + bt + c).$$

This gives the linear system

$$\begin{cases} 1^{200} = p_A(1)q(1) + a + b + c \\ 2^{200} = p_A(2)q(2) + (4a + 2b + c) \\ 200 \cdot 1^{199} = p'_A(1)q(1) + p_A(1)q'(1) + (2a + b) \end{cases} \iff \begin{cases} a + b + c = 1 \\ 2a + b = 200 \\ 4a + 2b + c = 2^{200} \end{cases}$$

Solving gives $r(t) = (2^{200} - 201)t^2 + (-2^{201} + 602)t + 2^{200} - 400$. This gives that

$$\begin{aligned} A^{200} &= p_A(A)q(A) + r(A) = r(A) \\ &= (2^{200} - 201)A^2 + (-2^{201} + 602)A + 2^{200} - 400I_n \\ &= \begin{bmatrix} -399 & 0 & 200 \\ -2^{201} + 2 & 2^{200} & 2^{200} - 1 \\ -800 & 0 & 401 \end{bmatrix}. \end{aligned}$$

△

2.7.1 Exercises

1. Consider
$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
compute A^{100}

2. Consider
$$A = \begin{bmatrix} -1 & 1 & -1 \\ -4 & 3 & 4 \\ 0 & 0 & -2 \end{bmatrix},$$
compute A^{123}

2.7.2 Solutions

1. Consider
$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$
compute A^{100}

   Following the process in the section we will compute the characteristic polynomial of A giving

$$p_A(t) = (t - 2)^2(t - 1).$$

Now we can apply long division to the polynomial t^{100} with the characteristic polynomial. This gives

$$t^{100} = p_A(t)q(t) + r(t) = p_A(t)q(t) + at^2 + bt + c.$$

Now plugging in $1, 2$ and 2 into the derivative gives the equations

$$\begin{aligned} a + b + c &= 1 \\ 4a + 2b + c &= 2^{100} \\ 2a + b &= 100(2^{99}). \end{aligned}$$

Solving this gives the remainder polynomial

$$r(t) = (-49(2^{100}) - 1)t^2 + (148(2^{100}) + 2)t - 99(2^{100}).$$

Thus we have

$$A^{100} = (-49(2^{100}) - 1)A^2 + (148(2^{100}) + 2)A - 99(2^{100})I.$$

2. Consider

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

compute A^{123}

Following the process in the section we will compute the characteristic polynomial of A giving

$$p_A(t) = (t-1)^2(t+2).$$

Now we can apply long division to the polynomial t^{123} with the characteristic polynomial. This gives

$$t^{123} = p_A(t)q(t) + r(t) = p_A(t)q(t) + at^2 + bt + c.$$

Now plugging in $1, 2$ and 2 into the derivative gives the equations

$$\begin{aligned} a + b + c &= 1 \\ 4a - 2b + c &= 2^{123} \\ 2a + b &= 123. \end{aligned}$$

Solving this gives the remainder polynomial

$$r(t) = ((2^{123} + 368)t^2 + (371 - 2^{124})t - 2^{123} - 730)/9.$$

Thus we have

$$A^{123} = ((2^{123} + 368)A^2 + (371 - 2^{124})A - (2^{123} - 730)I)/9.$$

2.8 Minimal polynomial

Definition 2.8.1. Let $A \in \mathsf{M}_n$, then define

$$\mathcal{P}(A) = \{p(t) \mid p(A) = 0\}.$$

That is $\mathcal{P}(A)$ represents that the set of polynomials such that plugging A gives the zero matrix.

From Caley-Hamilton we know that $\mathcal{P}(A) \neq \emptyset$. With that we will define $m_A(t) \in \mathcal{P}(A)$ to be the monic polynomial of minimal degree called the **minimal polynomial**.

Theorem 2.8.2. *Let $A \in \mathsf{M}_n$ and $m_A(t)$ be the minimal polynomial, then the follow properties hold.*

1. *Every nonzero $p(t) \in \mathcal{P}(A)$ is divisible by $m_A(t)$.*
2. *The polynomial $m_A(t)$ is the unique monic polynomial of minimal degree that vanishes at A.*
3. *Every eigenvalue of A is a root of $m_A(t)$.*

Proof. (1) For any polynomial $p(t) \in \mathcal{P}(A)$, consider the long division

$$p(t) = m_A(t)q(t) + r(t),$$

then

$$p(A) = 0 \implies m_A(A)q(A) + r(A) \implies r(A) = 0.$$

From the definition of long division $r(t)$ is a polynomial of degree strictly less than $m_A(t)$, thus if $r(t)$ was not the zero polynomial it would be the minimal polynomial so we have $r(t) = 0$.

(2) Assume for contradiction that there exists another minimal polynomial $\hat{m}_A(t)$, then from part we know that both polynomials divide each other. This forces them to be scalar multiples of one another. However they are both assumed to be monic, so they are actually the same polynomial.

(3) Let λ be an eigenvalue of A and consider the long division of the minimal polynomial with $(t - \lambda)$ giving

$$m_A(t) = (t - \lambda)q(t) + r,$$

where r is just a scalar. Now

$$0 = m_A(A) = (A - \lambda I)q(A) + rI.$$

If $r \neq 0$, then

$$(A - \lambda I)q(A) = -rI \implies (A - \lambda I)\left(-\frac{q(A)}{r}\right) = I_n.$$

As a consequence, then matrix $A - \lambda I$ is invertible and $\det(A - \lambda I) \neq 0$ which contradicts λ being an eigenvalue of A. Thus $r = 0$. $\square$

Corollary 2.8.3. *Let $A \in \mathsf{M}_n$. If all the eigenvalues of A have algebraic multiplicity of 1, then $p_A(t) = m_A(t)$.*

Corollary 2.8.4. *If two matrices are similar, then they have the same minimal polynomial.*

Theorem 2.8.5. *Let $A \in \mathsf{M}_n$. Then A is diagonalizable if and only if all of the roots of $m_A(t)$ have multiplicity 1.*

Proof. Let $A \in \mathsf{M}_n$ be diagonalizable, then we know that A is similar to a diagonal matrix D. By Corollary 2.8.4 $m_D(t) = m_A(t)$. ☐

Example 2.8.6. Consider the upper triangular matrices

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

We can see that their characteristic polynomials are the same

$$p_A(t) = p_B(t) = (t-1)^2(t-2).$$

Now the only option for a potential minimal polynomial is $m(t) = (t-1)(t-2)$. Trying this for A and B gives

$$(A-I)(A-2I) = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq 0$$
$$(B-I)(B-2I) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0.$$

Thus their minimal polynomials are

$$m_A(t) = (t-1)^2(t-2)$$
$$m_B(t) = (t-1)(t-2).$$

△

2.8.1 Exercises

1. Find the minimal polynomial of

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 5 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

and prove that A is not diagonalizable.

2.8.2 Solutions

1. Find the minimal polynomial of

$$A = \begin{bmatrix} 3 & 0 & 0 \\ 5 & 3 & 0 \\ 0 & 0 & -2 \end{bmatrix}$$

and prove that A is not diagonalizable.

First we compute the characteristic polynomial which gives

$$p_A(t) = (t-3)^2(t+2).$$

Thus the only options for the minimal polynomial are $p_A(t)$ and $(t-3)(t+2)$. Plugging A into $(t-3)(t+2)$ shows it is not zero. Thus the minimal polynomial is the characteristic polynomial.

2.9 Elementary symmetric functions

Recall that a minor of a matrix A is $\det(A_{ij})$. We call that minor principal if $i = j$. The minor is called size k if we have deleted $n-k$ of the rows/columns.

Definition 2.9.1. Let $A \in \mathsf{M}_n$. The sum of its principal minors of size k is denoted by $E_k(A)$.

Example 2.9.2. Consider

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

then

$$\begin{aligned} E_1(A) &= \operatorname{tr}(A) = 2, \\ E_2(A) &= \det\left(\begin{bmatrix} 1 & 2 \\ 0 & 2 \end{bmatrix}\right) + \det\left(\begin{bmatrix} 1 & -3 \\ 0 & -1 \end{bmatrix}\right) + \det\left(\begin{bmatrix} 2 & 4 \\ 0 & -1 \end{bmatrix}\right) = 2 - 1 - 2 = -1, \\ E_3(A) &= \det(A) = -2. \end{aligned}$$

$\triangle$

Definition 2.9.3. The kth elementary symmetric function of n complex numbers $\lambda_1, \dots, \lambda_n$, $k \le n$, is

$$S_k(\lambda_1, \dots, \lambda_n) = \sum_{1 \le i_1 < \cdots < i_k \le n} \prod_{j=1}^{k} \lambda_{i_j}.$$

For $A \in \mathsf{M}_n$ with eigenvalues $\lambda_1, \dots, \lambda_n$ we denote $S_k(A) = S_k(\lambda_1, \dots, \lambda_n)$.

Example 2.9.4. Consider $(1,2,3,4)$, then

$$\begin{aligned} S_1(1,2,3) &= 1+2+3 = 6, \\ S_2(1,2,3) &= 1(2)+1(3)+2(3) = 11, \\ S_3(1,2,3) &= 1(2)(3) = 6. \end{aligned}$$

△

Theorem 2.9.5. *Let $A \in \mathsf{M}_n$ have characteristic polynomial*

$$p_A(t) = t^n + (-1)^1 a_1 t^{n-1} + (-1)^2 a_2 t^{n-2} + \cdots + (-1)^n a_n,$$

then $a_k = E_k(A) = S_k(A)$ for each $k = 1, \ldots, n$.

The proof of this theorem is a bit of an algebraic mess. To show $a_k = S_k(A)$ we can use Vieta's formula.

To show that $a_k = E_k(A)$ we can use induction with the Laplace expansion.

Example 2.9.6. Consider

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

then a direct computation shows

$$\begin{aligned} E_1(A) &= 20 \\ E_2(A) &= 138 \\ E_3(A) &= 396 \\ E_4(A) &= 405. \end{aligned}$$

This gives that

$$p_A(t) = t^4 - 20t^3 + 138t^2 - 396t + 405.$$

With the characteristic polynomial we know that the eigenvalues satisfy

$$\begin{aligned} \lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 &= 20 \\ \lambda_1(\lambda_2 + \lambda_3 + \lambda_4) + \lambda_2(\lambda_3 + \lambda_4) + \lambda_3\lambda_4 &= 138 \\ \lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_2\lambda_4 + \lambda_1\lambda_3\lambda_4 + \lambda_2\lambda_3\lambda_4 &= 396 \\ \lambda_1\lambda_2\lambda_3\lambda_4 &= 405. \end{aligned}$$

△

2.9.1 Exercises

1. Determine the characteristic polynomial of the matrix

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

2. Determine the characteristic polynomial of the matrix

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

3. Determine the characteristic polynomial of the matrix

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

4. Determine a recurrence form for the characteristic polynomial of the constant tridiagonal matrix

$$\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & 1 & 1 \end{bmatrix}.$$

2.9.2 Solutions

1. Determine the characteristic polynomial of the matrix

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

The characteristic polynomial is

$$p_3(t) = (t-1)((t-1)^2 + 2) - (t-1).$$

2. Determine the characteristic polynomial of the matrix

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

The characteristic polynomial is

$$p_4(t) = (t-1)p_3(t) - \det \begin{bmatrix} t-1 & 1 & 0 \\ 0 & t-1 & 1 \\ 0 & 1 & t-1 \end{bmatrix} = (t-1)p_3(t) - (t-1)p_2(t).$$

3. Determine the characteristic polynomial of the matrix

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

The characteristic polynomial is

$$p_5(t) = (t-1)p_4(t) - (t-1)p_3(t).$$

4. Determine a recurrence form for the characteristic polynomial of the constant tridiagonal matrix

$$\begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 1 \\ 0 & 0 & 0 & \ldots & 1 & 1 \end{bmatrix}.$$

The characteristic polynomial is

$$p_n(t) = (t-1)p_{n-1}(t) - (t-1)p_{n-2}(t).$$

Chapter 3

Distance in spaces

3.1 Inner product space

For this chapter a field is only going to refer to $\mathbb{R}$ or $\mathbb{C}$. To emphasize this we will use the notation $\mathbb{F}$.

Definition 3.1.1. Let V be a vector space and $\mathbb{F}$ be a field. A function whose input consists of a pair of vectors in V and output is a scalar in the field, symbolized by

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{F},$$

is called an **inner product** when, for every $u, v, w \in V$, the following conditions are satisfied:

1. Nonnegativity: $\langle v, v \rangle$ is real and nonnegative.

2. Positivity: $\langle v, v \rangle = 0$ if and only if $v = 0$.

3. Additivity: $\langle v + u, w \rangle = \langle v, w \rangle + \langle u, w \rangle$.

4. Homogeneity: $\langle rv, u \rangle = r\langle v, u \rangle$ for every $r \in \mathbb{R}$.

5. Conjugate symmetry: $\langle v, u \rangle = \overline{\langle u, v \rangle}$.

The vector space V equipped with an inner product is called an **inner product space**.

Example 3.1.2. Consider $V = \mathbb{F}^n$ as a vector space over $\mathbb{F}$. For $u = [u_i],\ v = [v_i] \in V$, let

$$\langle u, v \rangle = v^* u = \sum_{i=1}^{n} u_i \overline{v_i}.$$

This is the *standard inner product* on $\mathbb{F}^n$. △

Example 3.1.3. For the vector space $C[a,b] = \{$all the continuous functions $f : [a,b] \to \mathbb{R}\}$ of all real continuous functions, in variable $t \in [a,b]$, the function

$$\langle f(t), g(t) \rangle = \int_a^b f(t)g(t)\ dt,$$

$f(t), g(t) \in C[a,b]$ is an inner product. $\triangle$

Example 3.1.4. Let $V = \mathsf{M}_{m\times n}(\mathbb{F})$ and let $A = [a_{ij}]$ and $B = [b_{ij}]$ in V. The *Frobenius inner product*

$$\langle A, B \rangle_F = \operatorname{tr}(B^*A)$$

$\triangle$

Theorem 3.1.5 (Cauchy-Schwarz Inequality)**.** *Let V be a vector space and $\langle \cdot, \cdot \rangle$ be an inner product for V. Then, for every $v, u \in V$ we have that*

$$\langle v, u \rangle^2 \leq \langle v, v \rangle \langle u, u \rangle.$$

Moreover we have equality if the vectors are colinear, or one of them is the zero vector.

Proof. The proof will be covered after vector norms. $\square$

So why the field restriction to $\mathbb{R}$ and $\mathbb{C}$? The main issue is of the first condition, of nonnegativity. It is hard to define inner products over arbitrary fields such that when you inner product a vector with itself you get a real nonnegative value. To fight this you might say, well can we loosen that first restriction. The answer is yes and no.

For the no side, as we shall see in the next section, the nonnegativity requirement is what allows us to turn these inner products into vector norms which can be thought of as a sort of distance. This will be an amazing geometric tool for us moving forward, so we can't drop this requirement.

For the yes side, we can study a more general object called a symmetric bilinear form, see wikipedia page: Symmetric bilinear form. This allows us to work over arbitrary fields and gives us some of the properties we are after. If you are further interested in fields being bad. I would recommend two topics.

1. The first is Non-Archimedean ordered fields, see wikipedia page: Non-Archimedean ordered field. These are fields that have an ordering, but fail the Archimedean property (you can scale one number by some number of copies of itself to be larger than another).

2. The other are fields of characteristic 2. These are particularly problematic fields where if you square any element you get back the identity. These tend to be a counterexample to many theorems and are explicitly disallowed in many more.

3.1.1 Exercises

1. For vectors $x, y \in \mathbb{R}^3$, determine which of the following are inner products

   (a) $\langle x, y\rangle = x_1y_1 + x_3y_3$

   (b) $\langle x, y\rangle = x_1y_1 - x_2y_2 + x_3y_3$

   (c) $\langle x, y\rangle = 2x_1y_1 + x_2y_2 + 4x_3y_3$

   (d) $\langle x, y\rangle = x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2$

   (e)
$$\langle x, y\rangle = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

3.1.2 Solutions

1. For vectors $x, y \in \mathbb{R}^3$, determine which of the following are inner products

   (a) $\langle x, y\rangle = x_1y_1 + x_3y_3$
   This is not an inner product. Consider $x = (0, 1, 0)$, then
$$\langle x, x, \rangle = 0 + 0 = 0.$$
   This breaks property 2 since $x \neq 0$.

   (b) $\langle x, y\rangle = x_1y_1 - x_2y_2 + x_3y_3$
   This is not an inner product. Consider $x = (1, 1, 0)$, then
$$\langle x, x, \rangle = 1 - 1 = 0.$$
   This breaks property 2 since $x \neq 0$.

   (c) $\langle x, y\rangle = 2x_1y_1 + x_2y_2 + 4x_3y_3$
   This is an inner product.

   (d) $\langle x, y\rangle = x_1^2y_1^2 + x_2^2y_2^2 + x_3^2y_3^2$
   This is not an inner product. Consider $x = (1, 1, 1)$, then
$$\langle x+x, x, \rangle = \langle (2, 2, 2), (1, 1, 1)\rangle = 4+4+4 = 12 \neq \langle x, x\rangle + \langle x, x\rangle = 3+3.$$
   This breaks property 3.

   (e)
$$\langle x, y\rangle = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$
   This is an inner product.

3.2 Vector norms

Definition 3.2.1. Let V be a vector space over $\mathbb{F}$. A function whose input consists of a vector in V and outputs is a real number, symbolized by

$$||\cdot|| : V \to \mathbb{R},$$

is called a vector **norm** when, for every $v, u \in V$, the following are satisfied:

1. $||v|| \geq 0$
2. $||v|| = 0$ if and only if $v = 0$
3. $||rv|| = |r|||v||$ for every $r \in \mathbb{F}$
4. $||v + u|| \leq ||v|| + ||u||$

The norm of the difference of two vectors v and u is called the **distance** between these two vectors, and it is denoted by $\mathrm{dist}(v, u) = ||v - u||$.

Definition 3.2.2. If $||\cdot||$ is a norm in a vector space V, then a vector $v \in V$ is called a **unit vector** if $||v|| = 1$.

Note that for any vector v, we can **normalize** v by taking $v/||v||$.

Lemma 3.2.3. *Let V be a vector space. For any inner product $\langle \cdot, \cdot \rangle$ in V, the function*

$$||v|| = \sqrt{\langle v, v \rangle},$$

for $v \in V$ is a vector norm.

Proof. The proof is an exercise for this section. □

Definition 3.2.4. Consider a vector norm $||\cdot||$ defined via an inner product $\langle \cdot, \cdot \rangle$. Then we say that $||\cdot||$ is a norm induced by the inner product $\langle \cdot, \cdot \rangle$.

Note that not every vector norm is able to be induced by an inner product.

Example 3.2.5. Let $V = \mathbb{F}^n$ derived from the standard inner product is the *Euclidean norm*

$$||u||_2 = \sqrt{u^* u}.$$

This norm is also called the ℓ_2 norm. △

Theorem 3.2.6. *Let V be a vector space, and let $\langle \cdot, \cdot \rangle$ be an inner product in V. If $||\cdot||$ is the induced norm of this inner product, then for every $u, v \in V$ the following hold:*

1. *Parallelogram law* $||v + u||^2 + ||v - u||^2 = 2||v||^2 + 2||u||^2$.
2. $$\langle v, u \rangle = \frac{1}{4}\left(||v + u||^2 - ||v - u||^2\right) = \frac{1}{2}\left(||v + u||^2 - ||v||^2 - ||u||^2\right).$$

3. $||v+u||\ ||v-u|| \le ||v||^2 + ||u||^2$.

Proof. The proof is an exercise for this section. □

Corollary 3.2.7. *A norm is induced by an inner product if and only if it satisfies the parallelogram law.*

Example 3.2.8. Let $V = \mathbb{F}^n$ and let $v \in V$, then the following norms are not induced by an inner product:

1. 1-norm or ℓ_1 norm: $||v||_1 = \sum_{k=1}^n |v_k|$.
2. ∞-norm or ℓ_∞: $||v||_\infty = \max\{|v_1|, \dots, |v_n|\}$.

△

3.2.1 Exercises

1. Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$. Prove that
$$||v|| = \sqrt{\langle v, v \rangle}$$
is a norm.

2. Prove theorem 3.2.6 for a real vector space. That is for $V = \mathbb{R}^n$ a vector space with inner product $\langle \cdot, \cdot \rangle$ and $u, v \in V$, then prove
   (a) Parallelogram law $||v+u||^2 + ||v-u||^2 = 2||v||^2 + 2||u||^2$.
   (b)
$$\langle v, u \rangle = \frac{1}{4}\left(||v+u||^2 - ||v-u||^2\right) = \frac{1}{2}\left(||v+u||^2 - ||v||^2 - ||u||^2\right).$$
   (c) $||v+u||\ ||v-u|| \le ||v||^2 + ||u||^2$.

3.2.2 Solutions

1. Let V be a vector space with inner product $\langle \cdot, \cdot \rangle$. Prove that
$$||v|| = \sqrt{\langle v, v \rangle}$$
is a norm.

   To be a norm we need to satisy the 4 properties. First note that $\|v\| \ge 0$ following the inner product of v with itself is nonnegative.

   Next, we have that the inner product of v with itself is 0 iff $v = 0$. This gives us norm property 2.

   For norm property 3 we have
$$\|rv\|^2 = \langle rv, rv \rangle = r\langle v, rv \rangle = r\overline{\langle rv, v \rangle} = r\bar{r}\langle v, v \rangle = |r|^2\|v\|^2.$$

Finally, for the last property we have

$$\begin{aligned}\|v+u\|^2 &= \langle v+u, v+u\rangle \\ &= \langle v,v\rangle + \langle u,u\rangle + \langle u,v\rangle + \overline{\langle v,u\rangle} \\ &\leq \langle v,v\rangle + \langle u,u\rangle + 2\langle u,v\rangle \\ &= (\|v\|+\|u\|)^2.\end{aligned}$$

2. Prove theorem 3.2.6 for a real vector space. That is for $V=\mathbb{R}^n$ a vector space with inner product $\langle\cdot,\cdot\rangle$ and $u,v\in V$, then prove
   (a) Parallelogram law $||v+u||^2+||v-u||^2=2||v||^2+2||u||^2$.

$$\begin{aligned}||v+u||^2+||v-u||^2 &= \langle v+u, v+u\rangle + \langle v-u, v-u\rangle \\ &= \langle v,v\rangle + \langle v,u\rangle + \langle u,v\rangle + \langle v,v\rangle - \langle v,u\rangle - \langle u,v\rangle + \langle u,u\rangle \\ &= 2\langle v,v\rangle + 2\langle u,u\rangle \\ &= 2||v||^2+2||u||^2.\end{aligned}$$

   (b)

$$\langle v,u\rangle = \frac{1}{4}\left(||v+u||^2-||v-u||^2\right) = \frac{1}{2}\left(||v+u||^2-||v||^2-||u||^2\right).$$

   Moved to homework 5.

   (c) $||v+u||\ ||v-u|| \leq ||v||^2+||u||^2$.
   Moved to homework 5.

3.3 Orthonormal systems

Definition 3.3.1. In an inner product space V, we say that two vectors are **orthogonal** if their inner product is zero.

Two subspaces, $U,W\subseteq V$, are called **orthogonal subspaces** if for every $u\in U$ and $w\in W$ we have that $\langle u,w\rangle=0$.

Definition 3.3.2. Let V be an inner product space. A set of nonzero vectors, $\{v_1,\dots,v_n\}$, is called orthogonal if $\langle v_i,v_j\rangle=0$ for all $i,j\in\{1,\dots,n\}$ such that $i\neq j$.

If in addition we require $||v_j||=1$ for all the vectors in the set, then we call the set **orthonormal**.

Finally, if that set forms a basis for V, then we say that the vectors are an **orthonormal basis** of V.

Theorem 3.3.3. *Let V be a vector space, and let $\langle \cdot, \cdot \rangle$ be an inner product in V. If a set of nonzero vectors $\{v_1, \dots, v_l\}$ in V is orthogonal, then the set is linearly independent.*

Definition 3.3.4. Let v, u be in the inner product space V. If the vector u is written as

$$u = \hat{u} + u_v,$$

where the vector $\hat{u}$ is orthogonal to v and the vector u_v is colinear to v, then u_v is said to be the **orthogonal projection** of u onto v and is denoted by $\text{proj}_v(u)$.

More explicitly we can compute $\text{proj}_v(u)$ as follows

$$\text{proj}_v(u) = \frac{\langle u, v \rangle}{||v||^2} v$$

Theorem 3.3.5. *Let V be an inner product space. If $\{v_1, \dots, v_n\}$ is an orthogonal basis of V, then any vector $x \in V$ can be written uniquely as*

$$x = \sum_{k=1}^{n} \frac{\langle x, v_k \rangle}{||v_k||^2} v_k = \sum_{k=1}^{n} proj_{v_k}(x).$$

If the set is instead orthonormal, then we can write x as

$$x = \sum_{k=1}^{n} \langle x, v_k \rangle v_k.$$

Theorem 3.3.6. *Every finite dimensional inner product space has an orthogonal basis.*

Proof. This follows from the Gram-Schmidt process that we will see in the next section. ☐

3.4 Gram-Schmidt process

Procedure 3.4.1. Given linearly independent vectors $u_1, u_2, \dots, u_n$, compute the orthogonal vectors

$$\begin{aligned}
w_1 &= u_1 \\
w_2 &= u_2 - \frac{\langle u_2, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 \\
w_3 &= u_3 - \frac{\langle u_3, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle u_3, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 \\
&\vdots \\
w_n &= u_n - \frac{\langle u_n, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 - \frac{\langle u_n, w_2 \rangle}{\langle w_2, w_2 \rangle} w_2 - \cdots - \frac{\langle u_n, w_{n-1} \rangle}{\langle w_{n-1}, w_{n-1} \rangle} w_{n-1}.
\end{aligned}$$

Written in our projection form this simplifies to

$$\begin{aligned}
w_1 &= u_1 \\
w_2 &= u_2 - \operatorname{proj}_{w_1}(u_2) \\
w_3 &= u_3 - \operatorname{proj}_{w_1}(u_3) - \operatorname{proj}_{w_2}(u_3) \\
&\ \vdots \\
w_n &= u_n - \operatorname{proj}_{w_1}(u_n) - \operatorname{proj}_{w_2}(u_n) - \cdots - \operatorname{proj}_{w_{n-1}}(u_n).
\end{aligned}$$

We can further normalize these vectors by

$$v_1 = \frac{w_1}{||w_1||},\ v_2 = \frac{w_2}{||w_2||},\ \ldots,\ v_n = \frac{w_n}{||w_n||}.$$

We then have

$$\operatorname{span}\{w_1, \ldots, w_n\} = \operatorname{span}\{v_1, \ldots, v_n\} = \operatorname{span}\{u_1, \ldots, u_n\}.$$

△

Example 3.4.2. Consider the basis

$$\{u_1, u_2, u_3\} = \{(1,1,1), (0,1,2), (1,0,3)\}$$

of $\mathbb{R}^3$ with the standard inner product. Then applying Gram-Schmidt we have

$$\begin{aligned}
w_1 &= u_1 = (1,1,1), \\
w_2 &= u_2 - \operatorname{proj}_{w_1}(u_2) = (0,1,2) - \frac{\langle (0,1,2), (1,1,1) \rangle}{\langle (1,1,1), (1,1,1) \rangle}(1,1,1) = (-1,0,1), \\
w_3 &= u_3 - \operatorname{proj}_{w_1}(u_3) - \operatorname{proj}_{w_2}(u_3) \\
&= (1,0,3) - \frac{\langle (1,0,3), (1,1,1) \rangle}{\langle (1,1,1), (1,1,1) \rangle}(1,1,1) - \frac{\langle (1,0,3), (-1,0,1) \rangle}{\langle (-1,0,1), (-1,0,1) \rangle}(1,1,1) \\
&= \left(\frac{2}{3}, -\frac{4}{3}, \frac{2}{3} \right).
\end{aligned}$$

△

3.5 Orthogonal complements

Definition 3.5.1. Let V be an inner product space of finite dimension. If U is a subspace of V, then the set

$$U^{\perp} = \{v \in V : \langle v, u \rangle = 0 \ \text{ for every } \ u \in U\}$$

is called the **orthogonal complement** of U also known as the “perp space” of U.

Example 3.5.2. Let $U = \text{span}\{(1,0,0,0),(0,1,0,0)\}$ be a subspace of $\mathbb{R}^4$, then

$$U^\perp = \text{span}\{(0,0,1,0),(0,0,0,1)\}.$$

△

Theorem 3.5.3. *Let V be an inner product space of finite dimension. For any subspace U of V the following hold:*

1. $V = U \oplus U^\perp$.
2. $(U^\perp)^\perp = U$.

Proof. (1) Assume that the dimension of V is dimension n. By Theorem 3.3.6, we know that U has an orthonormal basis $\{u_1, \ldots, u_k\}$. If $k = n$, then we are done since $U = V$ and $U^\perp = \{0\}$. If $k < n$, then $U^\perp \neq \{0\}$. Let $x \in V$, then define

$$y = x - \sum_{j=1}^{k} \text{proj}_{u_j}(x).$$

With this note that $\langle y, u_i \rangle = 0$ for $i = 1, \ldots, k$. Thus $y \in U^\perp$. This gives that $V = U + U^\perp$. Following $U \cap U^\perp = \{0\}$ we have that $V = U \oplus U^\perp$.

Part (2) is a homework exercise. □

Theorem 3.5.4. *Let V be an inner product space with subspace U. Then for any vector $v \in V$ we can write v uniquely as*

$$v = v_U + v_{U^\perp}$$

where $v_U \in U$ and $v_{U^\perp} \in U^\perp$.

Theorem 3.5.5. *Let V be a inner product space. If U and W are subspaces of V, then the following hold:*

1. *If $U \leq W$, then $W^\perp \leq U^\perp$.*
2. $(U + W)^\perp = U^\perp \cap W^\perp$.
3. $(U \cap W)^\perp = U^\perp + W^\perp$.

Proof. (1) If $U \leq W$, the for any $\hat{w} \in W^\perp$, we have

$$\langle \hat{w}, w \rangle = 0 \text{ for every } w \in W \implies \langle \hat{w}, u \rangle = 0 \text{ for every } u \in U \implies \hat{w} \in U^\perp.$$

(2) and (3) are left as exercises. □

3.5.1 Exercises

Prove parts (2) and (3) of Theorem 3.5.5. That is, let V be a inner product space. If U and W are subspaces of V, then prove the following hold:

1. $(U + W)^\perp = U^\perp \cap W^\perp$.
2. $(U \cap W)^\perp = U^\perp + W^\perp$.

3.5.2 Solutions

Prove parts (2) and (3) of Theorem 3.5.5. That is, let V be a inner product space. If U and W are subspaces of V, then prove the following hold:

1. $(U+W)^{\perp} = U^{\perp} \cap W^{\perp}$.

   Let $v \in (U+W)^{\perp}$, then for any $u \in U$ and $w \in W$, we have

   $$\langle v, u+w \rangle = 0.$$

   Now we will use specific u, w to show where v must be. If $u = 0$, then we see that $v \in W^{\perp}$, while if $w = 0$, then $v \in U^{\perp}$. Thus $v \in U^{\perp} \cap W^{\perp}$ giving $(U+W)^{\perp} \leq U^{\perp} \cap W^{\perp}$.

   Conversely, if $v \in U^{\perp} \cap W^{\perp}$, then $\langle v, u \rangle = 0$ for all $u \in U$ and $\langle v, w \rangle = 0$ for all $w \in W$. Putting these together gives $\langle v, u+w \rangle = 0$ for all $u \in U$ and $w \in W$. Thus $v \in (U+W)^{\perp}$ giving $U^{\perp} \cap W^{\perp} \leq (U+W)^{\perp}$.

2. $(U \cap W)^{\perp} = U^{\perp} + W^{\perp}$.

   $$(U \cap W)^{\perp} = ((U^{\perp})^{\perp} \cap (W^{\perp})^{\perp})^{\perp} = ((U^{\perp} + W^{\perp})^{\perp})^{\perp} = U^{\perp} + W^{\perp}.$$

3.6 Unitary matrices

Definition 3.6.1. If the columns of an $n \times n$ matrix U form an orthonormal basis of a vector space $\mathbb{F}^n$, then U is called a unitary matrix.

If we are over $\mathbb{R}^n$ this matrix is also called an orthogonal matrix.

Example 3.6.2. Consider the following unitary matrices

$$U = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$T = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$$

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

$$S = \begin{bmatrix} 1/2 & 1/2 & 1/\sqrt{2} & 0 \\ 1/2 & 1/2 & -1/\sqrt{2} & 0 \\ 1/2 & -1/2 & 0 & 1/\sqrt{2} \\ 1/2 & -1/2 & 0 & -1/\sqrt{2} \end{bmatrix}$$

△

The definition of a unitary matrix is often given in many forms. The following theorem addresses this.

Theorem 3.6.3. *Given a matrix $U \in \mathsf{M}_n$, the following are equivalent:*

1. *U is a unitary matrix.*
2. *The columns of U are orthonormal.*
3. *$U^*U = UU^* = I$.*
4. *U^* is a unitary matrix.*
5. *The rows of U are orthonormal.*
6. *U is invertible and $U^{-1} = U^*$.*
7. *For each $x \in \mathbb{F}^n$, $||x|| = ||Ux||$*
8. *For each $x, y \in \mathbb{F}^n$, $(Ux)^*(Uy) = x^*y$.*

Theorem 3.6.4. *Given a unitary matrix $U \in \mathsf{M}_n$, the following hold:*

1. *Every eigenvalue, λ, of U satisfies $|\lambda| = 1$.*
2. *The determinant of U satisfies $|\det(U)| = 1$.*
3. *The product of two unitary matrices is a unitary matrix.*

Unitary matrices are used extensively in numerical linear algebra. They are well behaved with respect to error propagation and it is essentially free (computationally) to invert them. We will see them several times in the following chapter.

3.6.1 Exercises

1. Prove or give a counter example to the claim that the product of two unitary matrices is unitary.
2. Prove or give a counter example to the claim that the sum of two unitary matrices is unitary.
3. Consider a nonzero vector $u \in \mathbb{R}^n$. Prove that the matrix
$$H = I - \frac{2}{u^\top u} uu^\top$$
is symmetric and unitary.
4. Let A be an $n \times n$ matrix such that the matrix $I_n + A$ is invertible. Prove that A is skew-symmetric if and only if the matrix
$$B = (I - A)(I + A)^{-1}$$
is unitary.

3.6.2 Solutions

1. Prove or give a counter example to the claim that the product of two unitary matrices is unitary.

   Let U, R be two unitary matrices. Then we have

   $$(UR)(UR)^* = URR^*U^* = UIU^* = I.$$

   Thus the product is unitary.

2. Prove or give a counter example to the claim that the sum of two unitary matrices is unitary.

   This is false. Consider,

   $$U = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}, \quad \text{then } U + U = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

   which is not unitary since the columns are no longer unit length.

3. Consider a nonzero vector $u \in \mathbb{R}^n$. Prove that the matrix

   $$H = I - \frac{2}{u^\top u} uu^\top$$

   is symmetric and unitary.

   To show that H is symmetric we want to show that $H = H^\top$, so consider

   $$\begin{aligned} H^\top &= \left(I - \frac{2}{u^\top u} uu^\top \right)^\top \\ &= I^\top - \frac{2}{u^\top u} (uu^\top)^\top \\ &= I - \frac{2}{u^\top u} uu^\top \\ &= H. \end{aligned}$$

   Now to show unitary we will show that $HH^* = I$, but since H is real $H^* = H^\top$ and following H is symmetric $H^\top = H$. Thus, we want to show that $H^2 = I$. Consider

   $$\begin{aligned} H^2 &= \left(I - \frac{2}{u^\top u} uu^\top \right)^2 \\ &= I - \frac{4}{u^\top u} uu^\top + \frac{4}{u^\top u} uu^\top \\ &= I. \end{aligned}$$

4. Let A be an $n \times n$ matrix such that the matrix $I_n + A$ is invertible. Prove that A is skew-symmetric if and only if the matrix

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

is unitary.

Proof. Let A be skew-symmetric, then we want to show that B is unitary. To do this, consider

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

Thus B is unitary.

Conversely, let B be unitary, then from the algebra above we have

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

Now we can do the following

$$\begin{aligned}
(I + A^*)^{-1}(I - A^*)(I - A)(I + A)^{-1} = I &\implies (I - A^*)(I - A)(I + A)^{-1} = (I + A^*) \\
&\implies (I - A^*)(I - A) = (I + A^*)(I + A) \\
&\implies I - A^* - A + AA^* = I + A^* + A + A^*A \\
&\implies A = -A^*.
\end{aligned}$$

□

Therefore A is skew-symmetric.

3.7 Matrix norms

Definition 3.7.1. Let $\mathsf{M}_{m\times n}(\mathbb{F})$ be the vector space of matrices with the standard operations of matrix addition and multiplication. A function

$$||\cdot|| : \mathsf{M}_{m\times n} \to \mathbb{R},$$

is called a **matrix norm** when, for every $A, B \in \mathsf{M}_{m\times n}(\mathbb{F})$ and $C \in \mathsf{M}_{n\times k}$, the following conditions are satisfied

1. $||A|| \geq 0$.
2. $||A|| = 0 \iff A = 0$.

3. $||rA|| = |r|\ ||A||$ for every $r \in \mathbb{F}$.
4. $||A+B|| \leq ||A|| + ||B||$.
5. $||AC|| \leq ||A||\ ||C||$.

Note that the first four properties are just our standard norm properties, but we have added a condition dealing with matrix multiplication. We call this last property **submultiplicativity property**.

Example 3.7.2. Let $A \in \mathsf{M}_{m\times n}$. When dealing with matrix norms the following are by far the most common

1. The **matrix 1-norm** is
$$||A||_1 = \max_{j\in\{1,\ldots,n\}} \sum_{i=1}^{m} |a_{ij}|.$$
2. The **matrix ∞-norm** is
$$||A||_\infty = \max_{i\in\{1,\ldots,n\}} \sum_{j=1}^{m} |a_{ij}|.$$
3. The **matrix 2-norm** (also known as the **spectral norm**) is
$$||A||_2 = \max\{\sqrt{\lambda} : \lambda \in \sigma(A^\top A)\}.$$
4. The **Frobenius norm** is
$$||A||_F = \sqrt{\operatorname{trace}(A^\top A)} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m} a_{ij}^2}.$$

△

Theorem 3.7.3. *Let $||\cdot||$ be a matrix norm over $\mathsf{M}_n(\mathbb{F})$, then we have the following*

1. $1 \leq ||I_n||$.
2. $||A^k|| \leq ||A||^k$
3. $\frac{1}{||A||} \leq ||A^{-1}||$
4. $|\lambda| \leq ||A||$ *for any eigenvalue, λ, of A.*

Theorem 3.7.4. *Let $||\cdot||_a$ and $||\cdot||_b$ be two matrix norms over $\mathsf{M}_{m\times n}$, then we have that there exists positive $r, s \in \mathbb{R}$ such that $r||A||_a \leq ||A||_b \leq s||A||_a$ for all $A \in \mathsf{M}_{m\times n}$.*

The above theorem says that all matrix norms are in some sense equivalent. That is you can convert one matrix norm into another by some scalar multiple. Because of this, when doing proofs with matrix norms it is common to see authors picking whichever one is most convenient for their proof.

3.7.1 Exercises

1. Let $A \in \mathsf{M}_n(\mathbb{R})$. Prove that for any symmetric matrix $S \in \mathsf{M}_n(\mathbb{R})$, we have
$$\left\| A - \frac{A + A^\top}{2} \right\|_F \leq \|A - S\|_F .$$

2. Let $A \in \mathsf{M}_n$ and $E \in \mathsf{M}_n$ be the matrix of all ones. Consider the sequence of matrices
$$\left\{ A_k = A + \frac{1}{k} E \right\}_{k=1}^{\infty} .$$
Prove that
$$\lim_{k \to \infty} \|A_k - A\|_1 = \lim_{k \to \infty} \|A_k - A\|_\infty = 0.$$

3. Let $A \in \mathsf{M}_n(\mathbb{R})$ and let $x \in \mathbb{R}^n$. Prove that
$$\|Ax\|_1 \leq \|A\|_1 \|x\|_1.$$

4. Prove that for any $A \in \mathsf{M}_n$ with $\operatorname{rank}(A) = 1$, we have that $\|A\|_2 = \|A\|_F$.

5. Let U be a unitary matrix and λ and eigenvalue of U. Prove that $|\lambda| = 1$.

3.7.2 Solutions

1. Let $A \in \mathsf{M}_n(\mathbb{R})$. Prove that for any symmetric matrix $S \in \mathsf{M}_n(\mathbb{R})$, we have
$$\left\| A - \frac{A + A^\top}{2} \right\|_F \leq \|A - S\|_F .$$

   Let $A \in \mathsf{M}_n(\mathbb{R})$ and $S \in \mathsf{M}_n(\mathbb{R})$ be symmetric, then
$$\begin{aligned}
\left\| A - \frac{A + A^\top}{2} \right\|_F &= \left\| \frac{A - A^\top}{2} \right\|_F \\
&= \left\| \frac{A - S + S - A^\top}{2} \right\|_F \\
&= \left\| \frac{A - S}{2} + \frac{S - A^\top}{2} \right\|_F \\
&\leq \left\| \frac{A - S}{2} \right\|_F + \left\| \frac{S - A^\top}{2} \right\|_F \\
&= \frac{\|A - S\|_F}{2} + \frac{\left\|(S - A)^\top\right\|_F}{2} \\
&= \frac{\|A - S\|_F}{2} + \frac{\|S - A\|_F}{2} \\
&= \frac{\|A - S\|_F}{2} + \frac{\|A - S\|_F}{2} \\
&= \|A - S\|_F .
\end{aligned}$$

2. Let $A \in \mathsf{M}_n$ and $E \in \mathsf{M}_n$ be the matrix of all ones. Consider the sequence of matrices
$$\left\{ A_k = A + \frac{1}{k}E \right\}_{k=1}^{\infty}.$$
Prove that
$$\lim_{k\to\infty} \|A_k - A\|_1 = \lim_{k\to\infty} \|A_k - A\|_\infty = 0.$$
First lets simplify inside the norms. We have
$$\|A_k - A\| = \left\| A + \frac{1}{k}E - A \right\| = \left\| \frac{1}{k}E \right\|.$$
So we want to prove that the limit of $\left\|\frac{1}{k}E\right\|$ goes to 0 in the 1 and ∞ norms. First the 1 norm
$$\lim_{k\to\infty} \left\| \frac{1}{k}E \right\|_1 = \lim_{k\to\infty} \max_{j\in\{1,\ldots,n\}} \sum_{i=1}^{n} \frac{1}{k} = \lim_{k\to\infty} \max_{j\in\{1,\ldots,n\}} \frac{n}{k} = \lim_{k\to\infty} \frac{n}{k} = 0.$$
A similar computation for the ∞ norm is
$$\lim_{k\to\infty} \left\| \frac{1}{k}E \right\|_\infty = \lim_{k\to\infty} \max_{i\in\{1,\ldots,n\}} \sum_{j=1}^{n} \frac{1}{k} = \lim_{k\to\infty} \max_{i\in\{1,\ldots,n\}} \frac{n}{k} = \lim_{k\to\infty} \frac{n}{k} = 0.$$

3. Let $A \in \mathsf{M}_n(\mathbb{R})$ and let $x \in \mathbb{R}^n$. Prove that
$$\|Ax\|_1 \le \|A\|_1 \|x\|_1.$$
Consider first
$$\|Ax\|_1 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right|.$$
By using the triangle inequality we can bring the modulus in giving
$$\sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \le \sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ij} x_j| = \sum_{j=1}^{n} |x_j| \sum_{i=1}^{n} |a_{ij}|.$$
We can view the above sum as multiplying $|x_j|$ by the sum of the modulus of the jth row of A. This would be smaller than multiplying $|x_j|$ by the maximum modulus row sum. This gives
$$\sum_{j=1}^{n} |x_j| \sum_{i=1}^{n} |a_{ij}| \le \sum_{j=1}^{n} |x_j| \left(\max_{k\in\{1,\ldots,n\}} \sum_{i=1}^{n} |a_{ik}| \right) = \|x\|_1 \|A\|_1.$$

4. Prove that for any $A \in \mathsf{M}_n$ with $\operatorname{rank}(A) = 1$, we have that $\|A\|_2 = \|A\|_F$.

   First note that if $\operatorname{rank}(A) = 1$, then $\operatorname{rank}(A^\top A) = 1$. Next the rank of a matrix is the number of nonzero eigenvalues. Thus $A^\top A$ has 1 nonzero eigenvalue say λ. With that we know

   $$\|A\|_2^2 = \lambda.$$

   Now for the Frobenius norm we know that the trace of a matrix is the sum of the eigenvalues, thus

   $$\|A\|_F^2 = \operatorname{trace}(A^\top A) = \lambda + 0 + \cdots + 0 = \lambda.$$

   Therefore following norms are nonnegative we have $\|A\|_2 = \|A\|_F$.

5. Let U be a unitary matrix and λ and eigenvalue of U. Prove that $|\lambda| = 1$.

   Proof. Let $U \in \mathsf{M}_n$ be unitary with λ and eigenvalue with corresponding eigenvector x, then

   $$(Ux)^*(Ux) = (\lambda x)^*(\lambda x) = \overline{\lambda}\lambda x^* x = |\lambda|^2 \|x\|_2^2.$$

   Now we also know that this is

   $$(Ux)^*(Ux) = x^* U^* U x = x^* I x = x^* x = \|x\|_2^2.$$

   Thus we have

   $$|\lambda|^2 \|x\|_2^2 = \|x\|_2^2 \implies |\lambda|^2 = 1 \implies |\lambda| = 1.$$

   $\square$

Chapter 4

Matrix Factorizations

4.1 LU factorization

Consider the following three matrices. The first corresponds to adding r times row i to row k.

$$T_{ik}^r = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & r & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{array}{l} \\ \\ \\ \leftarrow i\text{th row} \\ \\ \leftarrow k\text{th row} \\ \\ \\ \end{array}$$

The second corresponds to scaling row i by r.

$$T_i^r = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & r & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} \begin{array}{l} \\ \\ \\ \leftarrow i\text{th row} \\ \\ \\ \end{array}$$

The final corresponds to permuting rows i and k.

$$T_{ik} = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{array}{l} \\ \\ \\ \leftarrow i\text{th row} \\ \\ \leftarrow k\text{th row} \\ \\ \\ \end{array}$$

Definition 4.1.1. The matrices T_{ik}^r, T_i^r, and T_{ik} defined above are the elementary (row operation) matrices.

The value in these matrices is that we can take the standard Gaussian elimination row operations and encode them as matrix multiplications.

Example 4.1.2. Let

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix},$$

then

$$A \xrightarrow{R_3 - 2R_1} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 7 & 6 & 5 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} = T_{1,3}^{-2} A,$$

$$A \xrightarrow{3R_2} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 15 & 18 & 21 & 24 \\ 9 & 10 & 11 & 12 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} = T_2^3 A,$$

$$A \xrightarrow{P_{2,3}} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 9 & 10 & 11 & 12 \\ 5 & 6 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} = T_{2,3} A.$$

△

Note that the product of elementary matrices is a lower triangular matrix.

Definition 4.1.3. Let $A \in \mathsf{M}_n$, let U be the row reduced matrix from A by Gaussian elimination, and let L be the product of all elementary matrices needed to make A, then $A = LU$ is known as the **LU factorization** of A.

Example 4.1.4. Let

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

then apply the following Gaussian elimination steps

$$\begin{bmatrix}1 & 1 & 1\\ 2 & -1 & -1\\ 4 & -2 & 1\end{bmatrix} \xrightarrow{R_2-2R_1} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -3\\ 4 & -2 & 1\end{bmatrix} \xrightarrow{R_3-4R_1} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -3\\ 0 & -6 & -3\end{bmatrix} \xrightarrow{R_3-2R_2} \begin{bmatrix}1 & 1 & 1\\ 0 & -3 & -3\\ 0 & 0 & 3\end{bmatrix}$$

Now let

$$M_1 = T_{1,3}^{-4}T_{1,2}^{-2} = \begin{bmatrix}1 & 0 & 0\\ -2 & 1 & 0\\ -4 & 0 & 1\end{bmatrix}$$

and

$$M_2 = T_{2,3}^{-2} = \begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & -2 & 1\end{bmatrix},$$

then

$$M = M_2M_1 = \begin{bmatrix}1 & 0 & 0\\ -2 & 1 & 0\\ 0 & -2 & 1\end{bmatrix}$$

is the combination of all the row operations. This gives that $L = M^{-1}$ and U is the ecelon form of A.

$$A =$$

△

Definition 4.1.5. For a $A \in \mathsf{M}_n$, and $k \in \{1, \dots, n\}$, let

$$A(1 : k, 1 : k)$$

denote the $k \times k$ submatrix of A determined by the first k rows and first k columns of A. The matrix $A(1 : k, 1 : k)$ is known as the **leading principal submatrices** of A, and their determinants are known as the **leading principal minors** of A.

Theorem 4.1.6. *Let $A \in \mathsf{M}_n$. If all the leading principal minors of A are nonzero, then the matrix A has an LU factorization, $A = LU$. Moreover, if A has an LU factorization $A = LU$ and A is invertible, then this LU factorization is unique.*

Note that the above theorem in sufficient, but not necessary. To see this consider

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

If permutations are allowed, then all matrices emit an LU factorization. This is known as an LUP factorization.

So we have this factorization of a matrix into a product of a upper and lower triangular matrix, but what does this give us?

Let $A \in \mathsf{M}_n$ such that $A = LU$, then we can solve any linear system

$$Ax = b \iff (LU)x = b \iff L(Ux) = b$$

in two steps. First solve $Ly = b$, then solve $Ux = y$. These are both easy to do given the triangular form.

The computational cost of computing the LU factorization of a matrix is equal to the cost of performing Gaussian elimination, but once this is done, solving systems of equations can be done in linear time. This makes LU very efficient when one matrix is needed to solve many systems of equations.

Two notable examples of this are inverting a matrix and computing the determinant.

4.2 Schur's triangularization

Theorem 4.2.1. *Let $A \in \mathsf{M}_n$, then*

$$A = UTU^*$$

where U is a unitary matrix and T is an upper triangular matrix.

Proof. We will proceed by induction on the dimension n of the matrix A.

For the base case of $n = 1$, $A = [a_{11}]$ is a scalar and we are trivially an upper triangular matrix. Assume that the theorem holds for all $(n-1)\times(n-1)$ matrices, where $n > 1$.

Let A be an $n \times n$ matrix and let (λ_1, v_1) be an eigenpair of A where v_1 is normalized to unit length. Using the Gram-Schmidt process, we can extend the single vector $\{v_1\}$ to an orthonormal basis $\{v_1, w_2, \dots, w_n\}$ for $\mathbb{C}^n$. Let V_1 be the $n \times n$ matrix whose columns are these basis vectors

$$V_1 = \begin{bmatrix} v_1 & | & w_2 & | & \dots & | & w_n \end{bmatrix}$$

By construction, V_1 is a unitary matrix, satisfying $V_1^* V_1 = I$.

Now, consider the matrix transformation $V_1^* A V_1$. Let's examine its first column

$$(V_1^* A V_1)e_1 = V_1^*(AV_1 e_1) = V_1^*(Av_1).$$

Since v_1 is an eigenvector of A with eigenvalue λ_1, we have $Av_1 = \lambda_1 v_1$.

$$V_1^*(Av_1) = V_1^*(\lambda_1 v_1) = \lambda_1(V_1^* v_1) = \lambda_1 e_1.$$

Thus the first column of $V_1^* A V_1$ is $\lambda_1 e_1$ and $V_1^* A V_1$ has the following block matrix form

$$V_1^* A V_1 = \begin{bmatrix} \lambda_1 & x^\top \\ 0 & A_1 \end{bmatrix}.$$

where $x^\top$ is a $1 \times (n-1)$ row vector, 0 is an $(n-1) \times 1$ zero vector, and A_1 is an $(n-1) \times (n-1)$ matrix.

By the inductive hypothesis, there exists an $(n-1)\times(n-1)$ unitary matrix U_1 such that $U_1^* A_1 U_1 = T_1$, where T_1 is an upper triangular matrix. Now, define the $n \times n$ unitary block matrix

$$V_2 = \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix}.$$

Let $U = V_1 V_2$. and compute

$$\begin{aligned}
U^* A U &= (V_1 V_2)^* A (V_1 V_2) \\
&= V_2^* (V_1^* A V_1) V_2 \\
&= \begin{bmatrix} 1 & 0 \\ 0 & U_1^* \end{bmatrix} \begin{bmatrix} \lambda_1 & x^T \\ 0 & A_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & U_1 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 \\ 0 & U_1^* \end{bmatrix} \begin{bmatrix} \lambda_1 & x^T U_1 \\ 0 & A_1 U_1 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & x^T U_1 \\ 0 & U_1^* A_1 U_1 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & x^T U_1 \\ 0 & T_1 \end{bmatrix}
\end{aligned}$$

which is an upper triangular matrix. By the principle of mathematical induction, the theorem holds for all $n \geq 1$. □

Definition 4.2.2. The factorization of a matrix $A \in \mathsf{M}_n$ into $A = UTU^*$ where U is unitary and T is upper triangular is known as the **Schur's factorization** of A or the **triangularization** of A.

Example 4.2.3. The following matrices are triangularizable, but not diagonalizable

$$\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 5 \end{bmatrix} \to \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 5 \end{bmatrix}$$
$$\begin{bmatrix} 3 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 7 \end{bmatrix} \to \begin{bmatrix} 7 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$
$$\begin{bmatrix} 4 & 1 & 0 \\ -1 & 5 & 1 \\ -1 & 0 & 6 \end{bmatrix} \to \begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}.$$

△

This factorization is an extremely powerful theoretical tool, we can't always diagonalize a matrix, but we can always triangularize one. Note that this factorization is not unique, there are lots of ways to reorder the eigenvalues and scale the column vectors.

When we get to Jordan Canonical form, we will see how we can improve this result and get a more "unique" factorization.

For practical applications, Schur's is essentially never used. The issue is to compute this factorization we need to compute all of the eigenvalues/eigenvectors and repeatedly apply processes like Gram-Schmidt to slowly reduce the matrix. This makes it computationally impractical, especially given the most sought after piece of a matrix is usually the eigenvalues/eigenvectors.

4.2.1 Exercises

1. Find a Schur's factorization for

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

2. Consider an $A \in \mathsf{M}_n(\mathbb{R})$ with all real eigenvalues, and a Schur's triangularization of $A = UTU^\top$. Prove that $AA^\top = A^\top A$ if and only if T is a diagonal matrix.

3. Prove Cayley-Hamilton theorem using Schur's triangularization.

Theorem: Let $A \in \mathsf{M}_n$ with characteristic polynomial p_A, then $p_A(A) = 0$.

4.2.2 Solutions

1. Find a Schur's factorization for

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

By solving for the eigenvalues and eigenvectors we get

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad T = \begin{bmatrix} 5 & 2 \\ 0 & -1 \end{bmatrix}.$$

2. Consider an $A \in \mathsf{M}_n(\mathbb{R})$ with all real eigenvalues, and a Schur's triangularization of $A = UTU^\top$. Prove that $AA^\top = A^\top A$ if and only if T is a diagonal matrix.

Proof. If $AA^\top = A^\top A$, then we have that $T^\top T = TT^\top$. By looking at the off diagonal entries, we can see this is only possible if T is diagonal, since otherwise there would be a transpose flipping of entries.

Conversely, if T is a diagonal matrix, then

$$AA^\top = UTU^\top UT^\top U^\top = UTT^\top U^\top = UT^\top TU^\top = A^\top A.$$

$\square$

3. Prove Cayley-Hamilton theorem using Schur's triangularization.

Theorem: Let $A \in \mathsf{M}_n$ with characteristic polynomial p_A, then $p_A(A) = 0$.

Proof. Let $A \in \mathsf{M}_n$ with characteristic polynomial p_A and triangularization $A = UTU^\top$. Note first that $p_A(A) = Up_A(T)U^\top$. Thus, we want to show that $p_A(T) = 0$. We can write p_A as the product of its eigenvalues, that is

$$p_A(t) = \prod_{k=0}^{n}(t - \lambda_k).$$

plugging in T we get

$$p_A(T) = \prod_{k=0}^{n}(T - \lambda_k I).$$

Notice that each linear factor will leave T with a zero on that main diagonal entry. Furthermore the last diagonal entry will than be zero giving that $T - \lambda_n I$ will have a zero row column. Thus when we product that in we will be left with the last row/column be zero for the rest. Continue this process backwards will give that the entire matrix is zero. □

4.3 QR factorization

"The QR algorithm for solving the nonsymmetric eigenvalue problem is one of the jewels in the crown of matrix computations." – Higham (2003)

We know that computing the eigenvalues of matrices of size larger than 5 is not analytically possible. Thus we need to resort to numerical methods. This was an open problem for a long time in numerical linear algebra, how do you efficiently approximate the eigenvalues of a matrix? This extends to an even broader question, how to numerically find all of the roots of a polynomial?

As we shall see QR effectively solves both of these problems.

Definition 4.3.1. Let $A \in \mathsf{M}_{m\times n}(\mathbb{R})$ where $m \geq n$, then A can be written as $A = QR$ where $Q \in \mathsf{M}_{m\times n}(\mathbb{R})$ is real orthogonal and $R \in \mathsf{M}_n(\mathbb{R})$ is upper triangular.

Definition 4.3.2. Let $v \in \mathbb{R}^n$ be nonzero, then the matrix

$$H = I - 2\frac{vv^T}{\|v\|^2}$$

is a Householder matrix (or reflector). Note that H is real orthogonal.

This matrix reflects any vector x across the hyperplane perpendicular to v.

Theorem 4.3.3. *Every $A \in \mathsf{M}_n(\mathbb{R})$ and rank$(A) = n$ has a QR factorization $A = QR$ where Q is $n \times n$ with orthonormal columns and R is $n \times n$ upper triangular with all diagonal entries positive.*

Proof. Proceed by induction on n. For the base case of $n = 1$ we are trivially done since $Q = [1]$ and $R = A$ works.

Now assume the result holds for some $n - 1 \geq 1$ and let $A \in \mathsf{M}_n(\mathbb{R})$ with rank$(A) = n$. Let a_1 be the first column of A. Since rank$(A) = n$, a_1 is not the zero vector, so $r_{11} = \|a_1\|_2 > 0$. Construct the $m \times m$ Householder matrix H_1 using $v_1 = a_1 - r_{11}e_1$ such that

$$H_1 a_1 = (r_{11}, 0, \dots, 0)^T$$

Applying H_1 to the full matrix A gives the following block form

$$H_1 A = \begin{bmatrix} r_{11} & * \\ 0 & A_2 \end{bmatrix}$$

where $*$ represents unknown values.

Applying our induction hypothesis we know that $A_2 = \hat{Q}_2 R_2$ where $\hat{Q}_2$ is real orthogonal and R_2 is upper triangular. We can rewrite this as $R_2 = \hat{Q}_2^\top A_2$. Define

$$H_2 = \begin{bmatrix} 1 & 0 \\ 0 & \hat{Q}_2^\top \end{bmatrix}$$

and note that H_2 is also orthogonal. Now define

$$R = H_2 H_1 A = \begin{bmatrix} 1 & 0 \\ 0 & \hat{Q}_2^\top \end{bmatrix} \begin{bmatrix} r_{11} & * \\ 0 & A_2 \end{bmatrix} = \begin{bmatrix} r_{11} & * \\ 0 & R_2 \end{bmatrix}$$

which is upper triangular. Thus we can define $Q = H_1^\top H_2^\top$ and we have $A = QR$. $\square$

The above theorem can be extended to the following theorem that we will not prove. The proof is very similar to the one above, but we need more machinery on Householder transformations.

Theorem 4.3.4. *Every $A \in \mathsf{M}_{m\times n}(\mathbb{F})$ with $m \geq n$.*

1. *There is a unitary $Q \in \mathsf{M}_{m\times n}(\mathbb{F})$ and upper triangular $R \in \mathsf{M}_n(\mathbb{F})$ with real nonnegative diagonal entries such that $A = QR$.*
2. *If rank$(A) = n$, then the factors of Q and R are unique and R has positive diagonal entries.*

Procedure 4.3.5 (QR algorithm)**.** Let $A \in \mathsf{M}_n$.

1. Find a QR factorization of A,

$$A_0 = A = Q_1 R_1,$$

   where Q_1 is unitary and R_1 is upper triangular.

2. Interchange the order of the factors of A_0 in Step 1 to form the matrix
$$A_1 = R_1Q_1.$$
3. Find a QR factorization of A_1,
$$A_1 = Q_2R_2.$$
4. Form the matrix $A_2 = R_2Q_2$ and continue in this manner.

The sequences of matrices $A_0, A_1, \ldots,$ all have the same eigenvalues as the original A, and this process will converge to an upper triangular matrix.

Thus to compute **all** of the eigenvalues of a matrix to some degree of accuracy, we simply compute R_k for some k. △

Procedure 4.3.6 (Compute all roots of a polynomial)**.** Let $p \in \mathbb{C}[t]$ (this notation just means complex coefficients with the variable of t) be a monic polynomial of degree n,
$$p = t^n + a_{n-1}t^{n-1} + \cdots + a_0$$
then construct the matrix
$$C(p) = \begin{bmatrix} 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}.$$
This is known as the companion matrix for the polynomial p. The main property is that the characteristic polynomial of $C(p)$ is p. Thus the eigenvalues of $C(p)$ are the same as the roots of p. To find the roots of p we simply apply the QR algorithm to $C(p)$. △

4.3.1 Exercises

1. Let $A \in \mathsf{M}_n(\mathbb{R})$ with a QR factorization $A = QR$. Prove that A is similar to RQ.

4.3.2 Solutions

1. Let $A \in \mathsf{M}_n(\mathbb{R})$ with a QR factorization $A = QR$. Prove that A is similar to RQ.

   Let $A = QR$, then we want to find an invertible matrix P such that $PAP^{-1} = RQ$. Let $P = Q^\top$, then we have
$$PAP^{-1} = QAQ^\top = Q^\top QRQ = RQ.$$

4.4 SVD factorization

Theorem 4.4.1. *Let $A \in \mathsf{M}_{m\times n}$, then for $r = \min\{m, n\}$ there exists a unitary matrix $U \in \mathsf{M}_m$ and unitary matrix $V \in \mathsf{M}_n$ such that*

$$A = U\Sigma V^*,$$

where the $m \times n$ matrix $\Sigma = diag(\sigma_1, \dots, \sigma_r)$ has diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$.

Definition 4.4.2. The decomposition given in Theorem 4.4.1 is known as the **singular value decomposition (SVD)** of A. The diagonal values σ_i of Σ are known as the **singular values** of A.

The columns of U, u_i, are known as the **singular vectors** of A and the columns of V, v_i, are known as the **right singular vectors** of A.

From this, another way to express A is in the form

$$A = U\Sigma V^* = \sum_{k=1}^{r} \sigma_k u_k v_k^*$$

where r is the number of non-zero singular values of A, u_k is the kth row of U and v_k is the kth column of V. We will case this the rank 1 form of SVD.

Theorem 4.4.3. *Let $A \in \mathsf{M}_{m\times n}$ with rank$(A) = r$ and an SVD of $A = U\Sigma V^*$. Then, the column space of A is made from the first r singular vectors and the null space from the last $n - r$ right singular vectors. Or symbolically*

$$Col(A) = span\{u_1, \dots, u_r\}$$

and the null space of A is

$$Null(A) = span\{v_{r+1}, \dots, v_n\}.$$

Theorem 4.4.4. *Let $A \in \mathsf{M}_{m\times n}$, then the singular values of A are equal to the square roots of the eigenvalues of A^*A.*

The **matrix 2-norm** (also known as the **spectral norm**), $\|A\|_2$, can also be defined as the largest singular value of A.

4.4.1 Psuedo-inverse of a matrix

Definition 4.4.5 (Moore-Penrose inverse)**.** Let $A \in \mathsf{M}_n$ with SVD of $A = U\Sigma V^*$, then the matrix $A^\dagger = V\Sigma^\dagger U^*$ is known as the **pseudo-inverse** of A where $\Sigma^\dagger$ is formed by taking the reciprocal of all non-zero singular values of Σ.

Theorem 4.4.6. *Let $A \in \mathsf{M}_n$, then*

1. $AA^\dagger A = A$

2. $A^\dagger A A^\dagger = A^\dagger$

3. *The matrices* $AA^\dagger$ *and* $A^\dagger A$ *are symmetric.*

Theorem 4.4.7. *Let* $A \in \mathsf{M}_n$ *be invertible, then* $A^{-1} = A^\dagger$*.*

The pseudo-inverse is a useful tool in minimization problems. It allows non-invertible matrices to act like they are and gives ways to find closest solutions to systems of equations.

Note that the Moore-Penrose inverse is one of the most common pseudo-inverses, but there are many more. They are collectively called generalized inverses.

4.4.2 Dimension reduction ideas

Definition 4.4.8. Let $A \in \mathsf{M}_{m\times n}$ where A can be written in rank 1 SVD form as

$$A = \sum_{j=1}^{r} \sigma_j u_j v_j^*$$

, then define the rank k **singular truncation matrix** of A to be

$$A_k = \sum_{j=1}^{k} \sigma_j u_j v_j^*.$$

Theorem 4.4.9 (Eckart-Young-Mirsky theorem)**.** *Let* $A \in \mathsf{M}_{m\times n}$*, then*

$$\|A - A_k\|_2 \le \|A - B\|_2$$

for all rank k *matrices* $B \in \mathsf{M}_{m\times n}$*.*

Proof. Let $A \in \mathsf{M}_{m\times n}$ with r singular values, then note that

$$\|A - A_k\|_2 = \left\| \sum_{j=1}^{r} \sigma_j u_j v_j^* - \sum_{j=1}^{k} \sigma_j u_j v_j^* \right\|_2 = \left\| \sum_{j=k+1}^{r} \sigma_j u_j v_j^* \right\|_2 = \sigma_{k+1}.$$

Consider a rank k matrix $B \in \mathsf{M}_{m\times n}$, then B can be written as $B = XY$ where $X \in \mathsf{M}_{m\times k}$ and $Y \in \mathsf{M}_{k\times n}$. Now Y has k rows which we can express as a nontrivial combination of the columns of V, that is

$$w = a_1 v_1 + a_2 v_2 + \cdots + a_n v_{k+1}$$

such that $Yw = 0$. We can then scale w such that $\|w\|_2 = 1$ which gives that $a_1^2 + a_2^2 + \cdots + a_{k+1}^2 = 1$. Thus

$$\|A - B\|_2^2 \ge \|(A - B)w\|_2^2 = \|Aw\|_2^2 = a_1^2\sigma_1^2 + \cdots + a_{k+1}^2\sigma_{k+1}^2 \ge \sigma_{k+1}^2,$$

and taking the square root of both sides gives the desired result. □

Note that the above theorem also applies to the Frobenius norm, though the proof requires more work.

We will use theorem later, but the main idea is that the closest rank k matrix to rank n $(n \geq k)$ matrix is found by taking the first k singular values. It turns out that for larger generated matrices in the real world we can often find a small set of singular values that are substantially larger than the rest. This allows us to vastly reduce the size of the matrix we are dealing with, without greatly compromising the data within. We will talk about some ways this can be done in chapter 6.

4.4.3 Other applications of interest

1. SVD can help define what is known as a polar decomposition of a matrix. This roughly generalizes the idea of polar forms of complex numbers by having a rotation, and a distance from origin.
2. SVD is used to solve total least squares minimization problems.
3. SVD solves the "Orthogonal Procrustes problem" which asks to find the closest orthogonal matrix that takes a matrix A into a matrix B.

4.4.4 Exercises

1. Prove that a square matrix is unitary if and only if all its singular values are equal to 1.
2. If it true or not that two $n \times n$ matrices are unitarily equivalent if and only if they have exactly the same singular values?

4.4.5 Solutions

1. Prove that a square matrix is unitary if and only if all its singular values are equal to 1.

   Proof. Let $A \in \mathsf{M}_n$ such that $A = U\Sigma V^*$. For the forward direction, assume A is unitary, then $AA^* = I$ which implies

   $$AA^* = (U\Sigma V^*)^*(U\Sigma V^*) = V\Sigma U^*U\Sigma V^* = V\Sigma^2V^* = I$$

   We can multiply on the right by V^* and on the right by V giving that $\Sigma^2 = I$. Thus all of the singular values squared are equal to 1. Since the singular values are real nonnegative we have that they must be equal to 1.

   Conversely if all of the singular values of a matrix are equal to 1, then our SVD becomes $A = UIV^*$. To test unitary we have

   $$A^*A = (UIV^*)^*(UIV^*) = VIU^*UIV^* = VIV^* = I.$$

   Thus A is unitary. □

2. If it true or not that two $n \times n$ matrices are unitarily equivalent if and only if they have exactly the same singular values?

   This is false. Consider the following two matrices

   $$I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

   We have the singular values of both are both 1. However if I was unitarily similar to A, then there would exist a unitary matrix U such that

   $$I = UAU^* \implies U^* I U = A \implies I = A$$

   which is false.

4.5 Jordan canonical form

As we have seen in previous sections it is not possible to diagonalize all matrices, but through Schur's triangularization we can see that we can get close. In some ways the "closest" we can get to diagonal is known as the Jordan canonical form of a matrix, which we shall discuss in this section.

Definition 4.5.1. A $k \times k$ **Jordan block with eigenvalue** λ is an upper bidiagonal matrix

$$J_k(\lambda) = \begin{bmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{bmatrix}.$$

Another way to say this is that every main diagonal entry is λ, every super diagonal entry is 1, and all other entries are zero.

Definition 4.5.2. Let $A \in \mathsf{M}_n$ and $B \in \mathsf{M}_m$, then

$$A \oplus B = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$$

is the block diagonal matrix of size $m + n$. We call the operator $\oplus$ the **direct sum** of matrices.

Definition 4.5.3. A **Jordan matrix** J is a direct sum of Jordan blocks. That is J can be written in the following form

$$J = J_{k_1}(\lambda_1) \oplus J_{k_2}(\lambda_2) \oplus \cdots \oplus J_{k_m}(\lambda_m).$$

Note that not all the λ_i eigenvalues are distinct. There might be multiple blocks that share the same eigenvalue.

Theorem 4.5.4. *Let $A \in \mathsf{M}_n$, then A is similar to a Jordan matrix $J \in \mathsf{M}_n$. That is, there exists and invertible matrix S such that*

$$A = SJS^{-1}.$$

Furthermore J is unique up to permutation of the blocks.

Definition 4.5.5. For a matrix $A \in \mathsf{M}_n$ the Jordan matrix $J \in \mathsf{M}_n$ that A is similar to is known as the **Jordan canonical form** of A and is denoted $\mathrm{JCF}(A)$.

Theorem 4.5.6. *Let J be a Jordan matrix, then we have the following properties*

1. *For a given eigenvalue, the sum of the size of all the blocks corresponding to that eigenvalue is the algebraic multiplicity.*

2. *The number of Jordan blocks for a given eigenvalue is its geometric multiplicity.*

3. *A matrix is diagonalizable if and only all the Jordan blocks are size 1×1.*

4. *Given an eigenvalue. Its power in the minimal polynomial is the size of its largest Jordan block.*

Theorem 4.5.7. *Every matrix is similar to its transpose.*

Proof. On homework. □

Lemma 4.5.8. *Let $J_m(\lambda)$ be a Jordan block and let $k > 0$, then*

$$J_n(\lambda)^k = \begin{bmatrix} \lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \cdots & \binom{k}{m-1}\lambda^{k-(m-1)} \\ 0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots & \binom{k}{m-2}\lambda^{k-(m-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \binom{k}{1}\lambda^{k-1} \\ 0 & 0 & 0 & \cdots & \lambda^k \end{bmatrix}$$

Proof. We can write $J_m(\lambda)$ as $\lambda I + J_m(0)$. Note that $J_m(0)^m = 0$ (you will prove this in the below exercise). We will now apply the binomial theorem to this sum

$$\begin{aligned}
J_m(\lambda)^k &= (\lambda I + J_m(0))^k \\
&= \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} J_m(0)^j \\
&= \sum_{j=0}^{m-1} \binom{k}{j} \lambda^{k-j} J_m(0)^j \\
&= \begin{bmatrix}
\lambda^k & \binom{k}{1}\lambda^{k-1} & \binom{k}{2}\lambda^{k-2} & \cdots & \binom{k}{m-1}\lambda^{k-(m-1)} \\
0 & \lambda^k & \binom{k}{1}\lambda^{k-1} & \cdots & \binom{k}{m-2}\lambda^{k-(m-2)} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \binom{k}{1}\lambda^{k-1} \\
0 & 0 & 0 & \cdots & \lambda^k
\end{bmatrix}.
\end{aligned}$$

□

Theorem 4.5.9. *Let $A \in \mathsf{M}_n$. Then there exists a matrix $B \in \mathsf{M}_n$ such that*

$$\lim_{k\to\infty} A^k = B$$

if and only if

1. *every eigenvalue λ of A satisfy $|\lambda| \le 1$, and*
2. *when $|\lambda| = 1$, then $\lambda = 1$ and every Jordan block corresponding to λ is 1×1.*

Moreover $B = 0$ if and only if every eigenvalue λ of A satisfies $|\lambda| < 1$.

Theorem 4.5.10. *Let f be a differentiable function, then*

$$f(J_k(\lambda)) = \begin{bmatrix}
f(\lambda) & f'(\lambda) & \frac{f''(\lambda)}{2} & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\
0 & f(\lambda) & f'(\lambda) & \cdots & \frac{f^{(n-2)}(\lambda)}{(n-2)!} \\
0 & 0 & f(\lambda) & \cdots & \frac{f^{(n-1)}(\lambda)}{(n-1)!} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & f(\lambda)
\end{bmatrix}$$

4.5.1 Exercises

1. Let $A = J_3(1) \oplus J_4(2) \oplus J_2(0)$. What is the size of A? What are the eigenvalues? Is A diagonalizable?
2. Let $A = J_2(1) \oplus J_2(2) \oplus J_3(1) \oplus J_5(0)$. Find the characteristic polynomial of A. Find the minimal polynomial of A.
3. Prove that $(J_k(0))^m = 0$ for all $m \ge k$.

4. Let $A \in \mathsf{M}_n$. Prove that all the eigenvalues of A are equal to 0 if and only if there exists an $m \leq n$ such that $A^m = 0$.

5. For the above problem if you know the JCF of a matrix, what is the largest power to reach 0?

4.5.2 Solutions

1. Let $A = J_3(1) \oplus J_4(2) \oplus J_2(0)$. What is the size of A? What are the eigenvalues? Is A diagonalizable?

   To get the size we simply add the sizes of the Jordan blocks giving $3+4+2=9$.

   The eigenvalues are the values of the Jordan blocks, so $0, 1, 2$.

   The matrix A is not diagonalizable, since we have Jordan blocks that are larger than 1x1.

2. Let $A = J_2(1) \oplus J_2(2) \oplus J_3(1) \oplus J_5(0)$. Find the characteristic polynomial of A. Find the minimal polynomial of A.

   The characteristic polynomial is
   $$p_A(t) = (t-1)^5(t-2)^2t^5.$$
   The minimal polynomial is
   $$m_A(t) = (t-1)^2(t-2)t$$

3. Prove that $(J_k(0))^m = 0$ for all $m \geq k$.

   Each power of $J_k(0)$ will shift the main diagonal up by one. Once $m \geq k$ the main diagonal will be lost of the matrix leaving the zero matrix.

4. Let $A \in \mathsf{M}_n$. Prove that all the eigenvalues of A are equal to 0 if and only if there exists an $m \leq n$ such that $A^m = 0$.

   Proof. This follows from the JCF. For the forward direction let all the eigenvalues be zero, then by the previous theorem there exists a power such that all the jordan blocks will become zero, giving that A to that power will be zero.

   For the back direction, if $A^m = 0$, then we can't have any nonzero Jordan blocks, because they will never power to be zero by Theorem 4.5.10. □

5. For the above problem if you know the JCF of a matrix what is the largest power to reach 0?

   The largest power is one less than the size of the largest Jordan block. So in the worse case, $n-1$.

Chapter 5

Special matrix types

5.1 Hermitian matrices

Definition 5.1.1. Let $A \in \mathsf{M}_n(\mathbb{F})$, then A is called a **Hermitian matrix** if $A = A^*$

When a Hermitian matrix is real, it reduces to just being a symmetric matrix.

Theorem 5.1.2. *Given Hermitian matrices A, B we have the following*

1. *$A + B$ is a Hermitian matrix.*
2. *The diagonal entries of A are real.*
3. *A^k is a Hermitian matrix.*
4. *A^{-1} is a Hermitian matrix.*
5. *The eigenvalues of A are real numbers.*
6. *A is diagonalizable by a unitary matrix.*

Definition 5.1.3. A matrix is called **skew-Hermitian** if $A = -A^*$.

Theorem 5.1.4. *Every $A \in \mathsf{M}_n$ can be written as the sum of a Hermitian and skew Hermitian matrix.*

Proof. For $A \in \mathsf{M}_n$ consider the following matrices

$$H(A) = \frac{A + A^*}{2} \quad \text{and} \quad S(A) = \frac{A - A^*}{2}.$$

Note that

$$H(A)^* = \left(\frac{A + A^*}{2}\right)^* = \frac{A + A^*}{2} = H(A)$$

and

$$S(A)^* = \left(\frac{A - A^*}{2}\right)^* = \frac{A^* - A}{2} = -S(A).$$

Thus $H(A)$ is Hermitian and $S(A)$ is skew-Hermitian. Putting this together gives

$$A = H(A) + S(A).$$

$\square$

Definition 5.1.5. Let $A \in \mathsf{M}_n$. The matrix

$$H(A) = \frac{A + A^*}{2}$$

is called the **Hermitian part of** A and the matrix

$$S(A) = \frac{A - A^*}{2}$$

is called the **skew-Hermitian part of** A.

Theorem 5.1.6. *Let $A \in \mathsf{M}_n$, let λ be an eigenvalue of A, and let $\alpha \in \mathbb{R}$ be the smallest eigenvalue of $H(A)$ and $\beta \in \mathbb{R}$ be the largest eigenvalue of $H(A)$, then*

$$\alpha \leq Re(\lambda) \leq \beta.$$

5.1.1 Exercises

1. Given Hermitian matrices A, B prove the following

   (a) $A + B$ is a Hermitian matrix.

   (b) The diagonal entries of A are real.

   (c) A^k is a Hermitian matrix.

2. Let $A \in \mathsf{M}_n$ be a Hermitian matrix such that all of its eigenvalues satisfy $|\lambda| = 1$. Prove the following:

   (a) A is invertible.

   (b) A is a unitary matrix.

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

3. Let $A \in \mathsf{M}_n$ be a Hermitian matrix all of whose eigenvalues are equal to 2. Show that $A = 2I$.

5.1.2 Solutions

1. Given Hermitian matrices A, B prove the following

   (a) $A + B$ is a Hermitian matrix.

   $$(A + B)^* = A^* + B^* = A + B.$$

   (b) The diagonal entries of A are real.

   Note that the conjugate transpose leaves the main diagonal entries in the same spot, so the only difference is that they are their conjugates. Thus if they were not real, then they would not be equal to their conjugate.

   (c) A^k is a Hermitian matrix.

   $$(A^k)^* = (A^*)^k = A^k.$$

2. Let $A \in \mathsf{M}_n$ be a Hermitian matrix such that all of its eigenvalues satisfy $|\lambda| = 1$. Prove the following:

   (a) A is invertible.

   A matrix is invertible if all of its eigenvalues are non-zero, which we have here.

   (b) A is a unitary matrix.

   A matrix is unitary if it times it's conjugate is the identity. Here we have

   $$AA^* = UDU^*UD^*U^* = UDD^*U^*.$$

   Following $|\lambda| = 1$ we know that $DD^* = I$. Thus

   $$UDD^*U^* = UU^* = I.$$

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

   From the last problem $AA^* = I$, but we know that $A = A^*$, thus $AA = I$ giving that $A^{-1} = A$.

3. Let $A \in \mathsf{M}_n$ be a Hermitian matrix all of whose eigenvalues are equal to 2. Show that $A = 2I$.

   We have

   $$A = U2IU^* = 2UU^* = 2I.$$

5.2 Positive definite matrices

Definition 5.2.1. Let $A \in \mathsf{M}_n$ be a Hermitian matrix. We say that A is **positive definite** if

$$x^*Ax > 0 \text{ for all nonzero } x \in \mathbb{C}^n.$$

If we relax the condition to $x^*Ax \geq 0$ for all $x \in \mathbb{C}^n$, then the matrix is called **positive semi-definite**.

Theorem 5.2.2. *Let $A, B \in \mathsf{M}_n$ be positive definite, then we have the following.*

1. *$A + B$ is positive definite.*
2. *The diagonal entries of A are positive.*
3. *The eigenvalues of A are positive. This is also an if and only if. That if a matrix has positive eigenvalues, then it is positive definite.*
4. *A^k is positive definite for all $k > 0$.*
5. *The inverse of A exists and it is positive definite.*

The above results generally extend to positive-semidefinite, with some minor changes. We can see this in the following theorem.

Theorem 5.2.3. *Let $A, B \in \mathsf{M}_n$ be positive semi-definite, then we have the following.*

1. *$A + B$ is positive semi-definite.*
2. *The diagonal entries of A are nonnegative.*
3. *The eigenvalues of A are nonnegative. This is also an if and only if. That if a matrix has nonnegative eigenvalues, then it is positive semi-definite.*
4. *A^k is positive semi-definite for all $k > 0$.*

Theorem 5.2.4. *A Hermitian matrix $A \in \mathsf{M}_n$ is positive definite if and only if all of the principal minors of A are positive.*

Proof. (This proof only covers the forward direction. The reverse direction is more involved).

Let $A \in \mathsf{M}_n$ be a positive definite matrix and let $S \subseteq \{1, \ldots, n\}$ be any non-empty index set with size $k = |S|$. Let $A[S]$ be the $k \times k$ principal submatrix of A formed by selecting the rows and columns indexed by S. We must show that $\det(A[S]) > 0$.

Let $y \in \mathbb{C}^k$ be any non-zero vector. We can construct a vector $x \in \mathbb{C}^n$ such that the entries of x corresponding to the indices in S are the entries of y, and all other entries are zero. Since $y \neq 0$, it follows that $x \neq 0$. By this construction, the quadratic form x^*Ax simplifies exactly to $y^*A[S]y$.

$$x^*Ax = \sum_{i,j=1}^{n} \bar{x}_i a_{ij} x_j = \sum_{i,j \in S} \bar{x}_i a_{ij} x_j = \sum_{i,j=1}^{k} \bar{y}_i (A[S])_{ij} y_j = y^*A[S]y$$

Since A is positive definite and $x \neq 0$, we have $x^*Ax > 0$. Therefore, $y^*A[S]y > 0$ for all $y \neq 0$. This proves that the principal submatrix $A[S]$ is also positive definite.

Since $A[S]$ is positive definite and Hermitian, all of its eigenvalues $\lambda_1, \ldots, \lambda_k$ are real and positive. The determinant is the product of the eigenvalues which is therefore positive. This holds for any choice of S, so all principal minors of A are positive. □

The previous theorem can be improved to the following.

Theorem 5.2.5 (Sylvester's Criteria)**.** *A Hermitian matrix $A \in \mathsf{M}_n$ is positive definite if and only if all of the leading principal minors of A are positive.*

5.2.1 Exercises

1. Show that the following matrix is positive definite

$$A = \begin{bmatrix} 5 & -1 & 3 \\ -1 & 2 & -2 \\ 3 & -2 & 3 \end{bmatrix}$$

2. Show that the leading principal minors of the symmetric matrix

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

are nonnegative, but the matrix is not positive semi-definite. Why does this not violate Sylvester's Criteria?

3. Prove that the sum of a positive definite and positive semidefinite matrix is a positive definite matrix.

4. Prove that if A is positive semidefinite, then $A + \alpha I$ is positive definite for all $\alpha > 0$.

5. Show that if a positive semi-definite matrix has a zero on the main diagonal, then that entire row/column must be zero.

5.2.2 Solutions

1. Show that the following matrix is positive definite

$$A = \begin{bmatrix} 5 & -1 & 3 \\ -1 & 2 & -2 \\ 3 & -2 & 3 \end{bmatrix}$$

   Here the principal minors are

$$5, 12, 1$$

   Since these are all positive the matrix is positive definite.

2. Show that the leading principal minors of the symmetric matrix

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

   are nonnegative, but the matrix is not positive semi-definite. Why does this not violate Sylvester's Criteria?

   The leading principal minors are $0, 0$ neither of which are positive.

   This matrix is not positive semi-definite since -1 is an eigenvalue.

   This does not violate Sylvester's Ceriteria, since that only applies to positive definite matrices.

3. Prove that the sum of a positive definite and positive semidefinite matrix is a positive definite matrix.

   Let $x \in \mathbb{C}^n$, A be positive definite and B be positive semi-definite, then

$$x(A + B)x^* = xAx^* + xBx^*.$$

   Since A is positive definite we know that $xAx^* > 0$ and $xBx^* \geq 0$. The sum of a positive and nonnegative numbers is positive.

4. Prove that if A is positive semidefinite, then $A + \alpha I$ is positive definite for all $\alpha > 0$.

   Follows from the previous exercise noting that αI is positive definite for all $\alpha > 0$.

5. Show that if a positive semi-definite matrix has a zero on the main diagonal, then that entire row/column must be zero.

   If it had a nonzero off entry value, then we could construct the principal submatrix

$$\begin{bmatrix} 0 & a \\ a & b \end{bmatrix}$$

   which would have a negative minor value.

5.3 Normal matrices

Definition 5.3.1. Let $A \in \mathsf{M}_n$. We say that A is a **normal matrix** if $A^*A = AA^*$. That is A is normal if it commutes with its conjugate transpose.

Similar to some other matrix classes there are many equivalent ways to define normal matrices. Some of these are captured in the following theorem.

Theorem 5.3.2. *Let $A \in \mathsf{M}_n$, then the following are equivalent.*

1. *A is normal.*
2. *$A^*A = AA^*$.*
3. *A is unitarily diagonalizable.*
4. *There exists an orthonormal basis of $\mathbb{F}^n$ that consists of the eigenvectors of A.*

Some notable classes of matrices are normal given in the below examples.

Example 5.3.3. All symmetric and Hermitian matrices are normal. To see this let $A \in \mathsf{M}_n$ be Hermitian, then

$$AA^* = A^2 = A^*A.$$

△

Example 5.3.4. All unitary matrices are normal. Let $U \in \mathsf{M}_n$ be unitary, then

$$U^*U = I = UU^*.$$

△

Keep in mind that normal matrices form a broader class, then either Hermitian or unitary. The following theorem gives the connection to see how normal matrices form that broader class.

Theorem 5.3.5. *Let A be a normal matrix, then*

1. *For every eigenvector x of A corresponding to an eigenvalue λ is also an eigenvector of A^* corresponding to $\overline{\lambda}$.*
2. *If A is normal and its eigenvalues are real, then A is Hermitian.*

Theorem 5.3.6. *(Schur's Inequality) Let $A \in \mathsf{M}_n$ have eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then*

$$\sum_{k=1}^{n} |\lambda_k|^2 \leq \|A\|_F^2,$$

with equality if and only if A is normal.

Proof. Let $A \in \mathsf{M}_n$ and we will use Schur's Triangularization to write $A = UTU^*$ where T is upper triangular and U is unitary. With this $A^*A = UT^*TU^*$ and we have

$$\sum_{i=1}^{n} |\lambda_i|^2 \leq \sum_{i=1}^{n} |\lambda_i|^2 + \sum_{i<j} |t_{ij}|^2 = \mathrm{tr} T^*T = \mathrm{tr} UT^*TU^* = \mathrm{tr} A^*A = \|A\|_F^2.$$

With this we also get equality if and only if $\sum_{i<j} |t_{ij}|^2 = 0$ which happens if T is diagonal. Thus we have equality if and only if A is unitarily diagonalizable which was one definition of normal. □

This gives us a way of measuring the distance a given matrix is from being normal and by extension a way of measuring how far it is from being unitarily diagonalizable. For the eigenvalues, being unitarily diagonalizable is the ideal, so we can talk about how deficient a matrix is by how large

$$\|A\|_F^2 - \sum_{k=1}^{n} |\lambda_k|^2$$

is. This is sometimes known as the matrix's defect from normality.

5.3.1 Exercises

1. Can the product of two nonzero non-normal matrices be normal?

2. Let $A \in \mathsf{M}_n$. If there is a nonzero polynomial p such that $p(A)$ is normal, does it follow that A is normal?

3. If $A \in \mathsf{M}_n$ is nilpotent and normal, prove that $A = 0$.

4. Let $A, B \in \mathsf{M}_n$ be normal. Prove that A and B are similar if and only if they are unitarily similar.

5.3.2 Solutions

1. Can the product of two nonzero non-normal matrices be normal?

   Yes, let A be normal, then $AI = A$ is normal.

2. Let $A \in \mathsf{M}_n$. If there is a nonzero polynomial p such that $p(A)$ is normal, does it follow that A is normal?

   No, consider
   $$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$
   which is not normal, and the polynomial $p(t) = t^2$. Plugging A into p gives 0, which is a normal matrix.

3. If $A \in \mathsf{M}_n$ is nilpotent and normal, prove that $A = 0$.

   From a previous exercise we know that if $A^*A = AA^*$, then A is diagonalizable. If A is diagonalizable and nilpotent, then $A = U0U^\top = 0$.

4. Let $A, B \in \mathsf{M}_n$ be normal. Prove that A and B are similar if and only if they are unitarily similar.

   If A, B are normal and similar, then

   $$AA^* = SBS^{-1}S^*B^*(S^{-1})^*$$

   and

   $$A^*A = S^*B^*(S^{-1})^*SBS^{-1}$$

   which are equal giving

   $$S^*B^*(S^{-1})^*SBS^{-1} = SBS^{-1}S^*B^*(S^{-1})^*.$$

   Thus we must have that S is unitary.

   The backwards direction following immediately.

5.4 Nonnegative matrices

Definition 5.4.1. Let $A = [a_{ij}] \in \mathsf{M}_n(\mathbb{R})$, then A is called **nonnegative** if $a_{ij} \geq 0$ for all $i, j \in \{1, \dots, n\}$ and **positive** if $a_{ij} > 0$ for all $i, j \in \{1, \dots, n\}$.
These are denoted by $A \geq 0$ and $A > 0$ respectively.

This causes a natural partial ordering upon real matrices, where we can say that $A \geq B$ if $A - B \geq 0$. Similar notation can be used for vectors.

When dealing with nonnegative matrices, there is an important distinguishing property. To define it we will need the concept of graphs.

Definition 5.4.2. Let $A \in \mathsf{M}_n$ be a nonnegative matrix. We construct a graph $G(A) = (V, E)$ which consists of n vertices, $V = \{1, 2, \dots, n\}$, and a set of directed edges E that join the vertices according to the entries of A,

$$E = \{(i, j) : \text{ if } a_{ij} > 0\}.$$

We call $G(A)$ the **directed graph** associated with A. We refer to any edges of the form (i, i) as **loops**.

Definition 5.4.3. Let $A \in \mathsf{M}_n$ be nonnegative, then A is an **irreducible matrix** if there is at-least one directed path from any vertex to any other in $G(A)$. Otherwise the matrix is called reducible.

Definition 5.4.4. Let $A \in \mathsf{M}_n$, then the **spectral radius** of A is

$$\rho(A) = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}.$$

Another way to think of the spectral radius is the radius of the smallest circle in the complex plane that contains all the eigenvalues of A.

With these definitions we are able to give the foundational result in nonnegative matrix theory. It is stated in three forms relating to the different type of matrix used.

The first is the most general only dealing with nonnegative matrices.

Theorem 5.4.5 (Perron-Frobenius 1)**.** *Let $A \in \mathsf{M}_n$ be a nonnegative matrix. Then*

1. *The spectral radius of A is an eigenvalue of A. That is the largest eigenvalue of A is real and nonnegative.*
2. *There exists a nonnegative eigenvector x of A corresponding to $\rho(A)$,*
$$Ax = \rho(A)x.$$
3. *The spectral radius is a non-decreasing function of the entries of A. That is if $B \geq A \geq 0$, then $\rho(B) \geq \rho(A) \geq 0$.*

We can strengthen this result by switching to irreducible nonnegative matrices.

Theorem 5.4.6 (Perron-Frobenius 2). *Let $A \in \mathsf{M}_n$ be a irreducible nonnegative matrix. Then*

1. *The spectral radius of A is a simple eigenvalue of A. That is the largest eigenvalue of A is real and nonnegative and it has multiplicity 1.*
2. *There exists a positive eigenvector x of A corresponding to $\rho(A)$,*
$$Ax = \rho(A)x.$$
3. *The spectral radius is a increasing function of the entries of A. That is if $B > A \geq 0$, then $\rho(B) > \rho(A) \geq 0$.*

For the strongest form we need a slight generalization of a positive matrix given in the following definition.

Definition 5.4.7. Let $A \in \mathsf{M}_n$ be irreducible nonnegative, then A is called **primitive** if $A^k > 0$ for some $k > 0$.

Theorem 5.4.8 (Perron-Frobenius 3). *Let $A \in \mathsf{M}_n$ be a primitive nonnegative matrix. Then*

1. *The spectral radius of A is a simple eigenvalue of A. That is the largest eigenvalue of A is real and nonnegative and it has multiplicity 1.*
2. *There exists a positive eigenvector x of A corresponding to $\rho(A)$,*
$$Ax = \rho(A)x.$$
3. *The spectral radius of A is the sole dominant eigenvalue, that is*
$$\rho(A) > |\lambda|$$
for all eigenvalues λ of A that are not $\rho(A)$.
4. *The spectral radius is a increasing function of the entries of A. That is if $B \geq A \geq 0$, then $\rho(B) \geq \rho(A) \geq 0$.*

Theorem 5.4.9. *Let $A \geq 0$ be an irreducible nonnegative matrix. Let the minimal and maximal row sums of A be*

$$r = \min_{i \in \{1,\ldots,n\}} \sum_{j=1}^{n} a_{ij}, \quad R = \max_{i \in \{1,\ldots,n\}} \sum_{j=1}^{n} a_{ij}.$$

Then

$$r \leq \rho(A) \leq R.$$

Theorem 5.4.10. *Let $A \in \mathsf{M}_n$ be primitive, let x be the eigenvector for $\rho(A)$, and let y be the right eigenvector of $\rho(A)$, then*

$$\lim_{k \to \infty} \left(\frac{A}{\rho(A)} \right)^k = xy^\top.$$

5.4.1 Exercises

1. Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, then is A a primitive matrix?

2. Find $\lim_{k\to\infty} A^k$, where $A = \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$

3. Let
$$A = \begin{bmatrix} 1 & 3 & 1 \\ 4 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}.$$
Without directly computing the spectral radii of A and B, show the following

   (a) $\rho(A)$ is a simple eigenvalue of A, and

   (b) $\rho(B) > \rho(A)$.

5.4.2 Solutions

1. Let $A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$, then is A a primitive matrix?

   No, taking A to any power will not change the sign pattern. The zeros will always be zero.

2. Find $\lim_{k\to\infty} A^k$, where $A = \begin{bmatrix} 1/3 & 2/3 \\ 2/3 & 1/3 \end{bmatrix}$

   The limit is
$$\lim_{k\to\infty} A^k = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}.$$

3. Let
$$A = \begin{bmatrix} 1 & 3 & 1 \\ 4 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 1 & 2 \\ 1 & 0 & 1 \end{bmatrix}.$$
Without directly computing the spectral radii of A and B, show the following

   (a) $\rho(A)$ is a simple eigenvalue of A, and
   Note that A is a primitive matrix since $A^2 > 0$. Thus $\rho(A)$ is simple by the Perron-Frobenius theorem.

   (b) $\rho(B) > \rho(A)$.
   This follows from the Perron-Frobenius theorem noting that $B \geq A$ and B,A are both primitive matrices.