MAT 217 Notes

Max Chien

January 2024

Contents

Introduction

This document contains notes taken for the class MAT 217: Honors Linear Algebra at Princeton University, taken in the Spring 2024 semester. These notes are primarily based on lectures and lecture notes by Professor Jakub Witaszek. Other references used in these notes include *Linear Algebra Done Right* by Sheldon Axler, *Linear Algebra* by Kenneth Hoffman and Ray Kunze, and *Linear Algebra* by Stephen Friedberg, Arnold Insel, and Lawrence Spence. Since these notes were primarily taken live, they may contains typos or errors.

Chapter 1

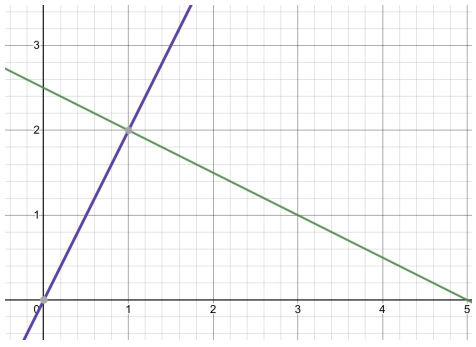
Systems of Linear Equations and Matrices

1.1 Introducton

We are often concerned with finding solutions to simultaneous equations, particular linear equations. We can solve these by performing certain operations on the various equations in the system. For instance, we may have the following system:

$$\begin{cases} x + 2y = 5 \\ 3x - y = 0 \end{cases} \iff \begin{cases} x + 2y = 5 \\ 0 - 5y = -10 \end{cases} \iff$$
$$\begin{cases} x + 2y = 5 \\ y = 2 \end{cases} \iff \begin{cases} x = 1 \\ y = 2 \end{cases}$$

We can also think of these equations as describing lines in the plane (or hyperplane), and the solutions as the set of intersections between these (see Figure 1.1).



We are often concerned about studying vector spaces, especially $\mathbb{R}^n$, of which the elements are real n-tuples:

$$\mathbb{R}^n = \left\{ \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \middle| v_1, \ldots, v_n \in \mathbb{R} \right\}$$

Vectors can have geometric interpretation as elements of n-dimensional space. However, they can also model mathematical concepts. For instance, we may encode a polynomial in a matrix:

$$\begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} \longleftrightarrow c_0 + c_1 x + \ldots + c_n x^n$$

Similarly, the coordinates of an n-tuple may encode the features of a particular entry in a database (such as data on a user).
Consider a system of three linear equations in three (real) variables. Then each equation represents a plane, and we are interested in their intersection. The solutions may take the following forms:

0) No solution: if the three planes are parallel (but not equal), or two planes intersect in a line and the third is parallel to that line.

1) A point: for instance if we have the system $\begin{cases} x = 0 \\ y = 0 \\ z = 0 \end{cases}$

2) A line: if the three planes intersect in a line.

3) A plane: if the three planes are equal.

4) $\mathbb{R}^3$: If each plane involves no variables (i.e. each is the entire space $\mathbb{R}^3$).

1.2 Gauss-Jordan Elimination

Definition 1.1

Call a rectangular array of numbers a **matrix**. We denote its size by $m \times n$, where m is the number of rows and n the number of columns:

$$A = \begin{bmatrix} * & * & * \\ * & * & * \end{bmatrix} (A \text{ is a } 2 \times 3 \text{ matrix})$$

Definition 1.2

The $n \times n$ **identity matrix** I_n has 1 along the diagonal and 0 everywhere else:

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

For a given system of linear equations, we define the **augmented matrix**:

$$\begin{cases} x + 2y = 5 \\ 2x - y = 0 \end{cases} \longleftrightarrow \left[\begin{array}{cc|c} 1 & 2 & 5 \\ 2 & -1 & 0 \end{array}\right]$$

We define the permitted operations on augmented matrices as follows:

Definition 1.3

The **elementary row operations** (EROs) on matrices are:

- Swap any two rows.
- Multiply or divide a row by any non-zero number.
- Add or subtract a multiple of any row to *another* row.

Example 1.1

Consider the system of equations:

$$\begin{cases} x - 2y + 3z = 1 \\ -x + y + 2z = -2 \\ x - 4y + 13z = -1 \end{cases}$$

Then we write the corresponding augmented matrix, and perform ele-

mentary row operations (where R_1, R_2, R_3 represent the rows):

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 1 \\ -1 & 1 & 2 & -2 \\ 1 & -4 & 13 & -1 \end{array}\right] \begin{array}{l} R_2 \to R_2 + R_1 \\ \longrightarrow \\ R_3 \to R_3 - R_1 \end{array}$$

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 1 \\ 0 & -1 & 5 & -1 \\ 0 & -2 & 10 & -2 \end{array}\right] \begin{array}{l} R_2 \to -R_2 \\ \longrightarrow \end{array}$$

$$\left[\begin{array}{ccc|c} 1 & -2 & 3 & 1 \\ 0 & 1 & -5 & 1 \\ 0 & 2 & 10 & -2 \end{array}\right] \begin{array}{l} R_1 \to R_1 + 2R_2 \\ \longrightarrow \\ R_3 \to R_3 - 2R_2 \end{array}$$

$$\left[\begin{array}{ccc|c} 1 & 0 & -7 & 3 \\ 0 & 1 & -5 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right] \iff$$

$$\begin{cases} x - 7z = 3 \\ y - 5z = 1 \\ 0 = 0 \end{cases}$$

So we have $x = 7z + 3$ and $y = 5z + 1$, so our solution is

$$\left\{ \left[\begin{array}{c} 7t+3 \\ 5t+1 \\ t \end{array}\right] \middle| \, t \in \mathbb{R} \right\} = t \begin{bmatrix} 7 \\ 5 \\ 1 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$$

and it is apparent that our solution is a line in $\mathbb{R}^3$.

The significance of the EROs is that, by performing an ERO, we do not change the solutions to the system of equations associated with a given matrix. This is how we were able to conclude in the example that the solution to the original system was precisely the solution to the final system.

Theorem 1.1

Suppose that an augmented matrix B is reached from matrix A by a finite sequence of EROs (they are **row-equivalent**). Then the solutions to the system associated with B are identical to the solutions to the system associated with A.

Proof. It is sufficient to show that any single ERO does not change the solutions to the equation. Suppose a certain ERO takes a matrix M to M'. Then each row of M' is a linear combination of the rows of M. Thus, any solution to M will also be a solution to M'. To show inclusion in the other direction, note that any ERO has an inverse which is also an ERO. Thus, we may perform the inverse ERO to take us from M' to M. By the same logic as above, every solution to

M' is also a solution to M. So the solutions are identical. By induction, this is true for any two row-equivalent matrices. □

Definition 1.4

A matrix is in **reduced row-echelon-form** (RREF) if:

- From left to right, the first nonzero entry of each row is 1 (leading 1).
- If a column contains a leading 1, then every other entry in the column is 0 (nonleading 1s are unaffected)
- From top to bottom, leading 1s are left to right.

Essentially, our matrix should have the form:

$$\begin{bmatrix} 0 & \dots & 1 & * & 0 & * & 0 & * \\ & & 0 & \dots & 1 & * & 0 & * \\ & & & & 0 & \dots & 1 & * \end{bmatrix}$$

It should also be noted that for augmented matrices, the augmented column is not considered when determining whether the matrix is in RREF.

For any matrix in RREF, we consider the column indices of the leading 1s: $i_1 < i_2 < \ldots < i_k$. Then we denote the corresponding variables $x_{i_1}, x_{i_2}, \ldots, x_{i_k}$. This motivates the following definition:

Definition 1.5

A **free variable** is a variable for which the column in the RREF does not have a leading 1. A **determined variable** is a variable which is not free.

Referring back to the example, we see that determined variables will have values in terms of the free variables, which result in parameters in our final solution set. In other words, free variables may be assigned arbitrary variables, which will uniquely determine values for the free variables that results in a solution to the system of equations.

Theorem 1.2

Any matrix can be put into RREF using EROs. Moreover, the RREF matrix of A is unique regardless of the EROs, which allows to refer to the RREF of a given matrix, which we denote RREF(A).

Gauss-Jordan elimination offers an algorithm for finding RREF(A):

1. Consider the first column. If column is all zeros, then move on. Otherwise, choose any row such that the first nonzero value is in the first column, and swap it into the first row.
2. Divide the row so that the first entry is 1.
3. Add multiples of the first row to the other rows so that every other entry in the first column is 0.
4. Repeat with each following column, except that in Step 1 we only use rows that have not already been used.

Example 1.2

Define the following RREF matrix:

$$A = \left[\begin{array}{ccccc|c} 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 4 & 0 \end{array}\right]$$

Denote the variables associated with the columns x, y, z, w, u, respectively. Then our leading variables are x, z, w, and our free variables are y, u. So we have

$$\begin{cases} x + 2y + u = 0 \\ z + u = 0 \\ w + 4u = 0 \end{cases}$$

We can then move the free variables to the right side:

$$\begin{cases} x = -2y - u \\ z = -u \\ w = -4u \end{cases}$$

Here, any choice of y, u will determine the values of x, z, w, giving a solution to the system:

$$\left\{ y \begin{bmatrix} -2 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + u \begin{bmatrix} -1 \\ 0 \\ -1 \\ -4 \\ 1 \end{bmatrix} \middle| \, y, u \in \mathbb{R} \right\}$$

1.3 Linear Systems

In the introduction, we found that the solution sets of a given linear system make take many different forms. The following theorem allows us to determine the form of the solutions of a linear system.

Theorem 1.3

Suppose an augmented matrix is in RREF:

$$\left[\begin{array}{cccccc|c} 1 & * & 0 & * & 0 & * & c_1' \\ & & 1 & * & 0 & * & c_2' \\ & & & & \vdots & & \vdots \\ & & & & 1 & * & c_k' \\ & & & & & & \vdots \\ & & & & & & c_n' \end{array}\right]$$

Then the associated linear system

1) has no solutions (is **inconsistent**) if and only if any $c_i' \neq 0$ for some $k+1 \leq i \leq n$.

2) has exactly 1 solution if and only if every variable is leading. In this case, we have

$$\text{RREF}(A) = \left[\begin{array}{ccc|c} 1 & & 0 & c_1' \\ & \ddots & & \vdots \\ 0 & & 1 & c_k' \\ & & & 0 \\ & & & \vdots \\ & & & 0 \end{array}\right]$$

3) has infinitely many solutions if and only if it is consistent and there is at least one free variable.

Corollary

If a linear system has m equations and n variables, with $m < n$, then the system is either inconsistent or has infinitely many solutions.

Proof. There is at most one leading variable per equation. So there are at most m leading variables, but $m < n$, so we have free variables. So the system either satisfies condition 1) or 3) above. □

For certain systems, we may make stronger statements about their solutions.

Definition 1.6

A system is **homogeneous** if $c_1' = \ldots = c_n' = 0$ (where c' denotes the constants of RREF(A)).

A homogeneous system is always consistent, since $\mathbf{0} = (0, \ldots, 0)$ is always a solution. Moreover, suppose a homogeneous system has m equations and n variables, with $m < n$. Since a homogeneous system is always consistent, the corollary implies there are infinitely many (nonzero) solutions.

We now consider some applications of our study of linear systems to geometry.

Definition 1.7

We call $X \subseteq \mathbb{R}^n$ **algebraic** if it is described by a system of polynomial equations in n variables.

Consider an algebraic set $X \subseteq \mathbb{R}^2$ defined by $\{(x, y) | g(x, y) = 0\}$ for some polynomial g in x, y.

- Suppose $\deg(g) = 1$. Then $g(x, y) = ax + by + c$, which describes a line in $\mathbb{R}^2$.
- Suppose $\deg(g) = 2$. Then $g(x, y) = ax^2 + bxy + cy^2 + dx + ey + f$. This describes a conic section of some kind. The nondegenerate conic sections are the ellipse, parabola, and hyperbola. The degenerate conic sections are the line (for instance $g(x, y) = x^2$), empty set ($g(x, y) = x^2 + 1$), or two lines ($g(x, y) = xy$).

We can use the facts we have learned about linear systems to prove the following statement:

Theorem

Through any 5 points in $\mathbb{R}^2$, there is always a conic $X \subseteq \mathbb{R}^2$ passing through the 5 points (where we demand that g is not identically 0).

Proof. Let our points be $(x_1, y_1), \ldots, (x_5, y_5)$. Then our conic must be described by a polynomial $g = ax^2 + bxy + cy^2 + dx + ey + f$ such that

$$\begin{cases} g(x_1, y_1) = 0 \\ \vdots \\ g(x_5, y_5) = 0 \end{cases} \iff \begin{cases} x_1^2 a + x_1 y_1 b + \ldots + f = 0 \\ \vdots \\ x_5^2 a + x_5 y_5 b + \ldots + f = 0 \end{cases}$$

$$\iff \left[\begin{array}{cccc|c} x_1^2 & x_1 y_1 & \ldots & 1 & 0 \\ & \vdots & & & \vdots \\ x_5^2 & x_5 y_5 & \ldots & 1 & 0 \end{array}\right]$$

Here we have a linear system, where the variables are a, b, c, d, e, f. Moreover, the system is homogeneous with 5 equations and 6 variables, so we know there is a nonzero solution, completing the proof. □

We can also consider systems of linear equations over other fields. Systems of linear equations over $\mathbb{Q}$ behave similarly to those over $\mathbb{R}$. In particular, we have the following result:

Proposition 1.4

For any homogeneous system of linear equations over $\mathbb{R}$ with rational coefficients, if there exists a real nonzero solution $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$, then there exists a rational nonzero solution $\mathbf{x'} = (x'_1, \ldots, x'_n) \in \mathbb{Q}^n$

Proof. Since the homogeneous system has a nonzero solution, it must have infinitely many solutions. This implies that there are one or more free variables. Setting the free variables to rational values ensures all variables are rational (if a determined variable doesn't depend on free variables, then its value will always be rational). □

1.4 Fields

We have so far assumed that all variables and coefficients of our linear systems are in $\mathbb{R}$, but this need not be so.

Example 1.3

Consider the system over $\mathbb{C}$

$$\begin{cases} x + iy = 1 \\ x - iy = i \end{cases}$$

Then we can use Gauss-Jordan elimination to find RREF(A):

$$\left[\begin{array}{cc|c} 1 & i & 1 \\ 1 & -i & 1 \end{array}\right] \xrightarrow{R_2 \to R_2 - R_1} \left[\begin{array}{cc|c} 1 & i & 1 \\ 0 & -2i & i-1 \end{array}\right] \xrightarrow{R_2 \to R_2/(-2i)}$$

$$\left[\begin{array}{cc|c} 1 & i & 1 \\ 0 & 1 & -\frac{i-1}{2i} \end{array}\right] \xrightarrow{R_1 \to R_1 - iR_2} \left[\begin{array}{cc|c} 1 & 0 & 1 + i\frac{i-1}{2i} \\ 0 & 1 & -\frac{i-1}{2i} \end{array}\right] \iff \begin{cases} x = \frac{1+i}{2} \\ y = -\frac{1+i}{2} \end{cases}$$

So we have seen that solving linear systems over $\mathbb{C}$ is very similar to solving linear systems over $\mathbb{R}$. This is also the case in $\mathbb{Q}$. We are led naturally to consider which algebraic structures we can apply similar methods to. This motivates the following definition, which generalizes the properties of $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$; namely, addition, subtraction, multiplication, and division, with distribution.

Definition 1.8

A **field** is a triple $(\mathbb{F}, +, *)$ such that $\mathbb{F}$ is a nonempty set together with two operations $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$, $* : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$ that satisfy the following axioms:

1. $\alpha + \mathcal{B} = \mathcal{B} + \alpha$ and $\alpha * \mathcal{B} = \mathcal{B} * \alpha$ (commutativity of $+, *$).
2. $(\alpha + \mathcal{B}) + \gamma = \alpha(\mathcal{B} + \gamma)$ and $(\alpha\mathcal{B})\gamma = \alpha(\mathcal{B}\gamma)$ (associativity of $+, *$).
3. $\exists 0 \neq 1 \in F$ such that $\forall \alpha \in \mathbb{F}, \alpha + 0 = \alpha, \alpha * 1 = \alpha$ (existence of additive and multiplicative identities).
4. $\forall \alpha \in \mathbb{F}$, $\exists \mathcal{B} \in \mathbb{F}$ s.t. $\alpha + \mathcal{B} = 0$ (existence of additive inverses).
5. $\forall \alpha \neq 0 \in \mathbb{F}$, $\exists \mathcal{B} \neq 0 \in \mathbb{F}$ s.t. $\alpha\mathcal{B} = 1$ (existence of multiplicative inverses for nonzero elements).
6. $\gamma(\alpha + \mathcal{B}) = \gamma\alpha + \gamma\mathcal{B}$ (distributivity of $*$ over $+$).

Example 1.4

$\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ are all examples of fields. $\mathbb{Z}$ is a nonexample, since multiplicative inverses do not exist for elements besides $1, -1$. $\mathbb{R}^2$ with addition, multiplication defined componentwise is not a field, since $(1, 0)$ has no additive inverse (this is in general true for any $\mathbb{F}^{n>1}$ with addition, multiplication defined componentwise).

Here are some immediate consequences of the definition of a field:

- The additive inverse of an element α is unique (denoted $-\alpha$)
- The multiplicative inverse of an element α is unique (denoted α^{-1})

Definition 1.9

Given a field $\mathbb{F}$, the **characteristic** of $\mathbb{F}$ is the smallest n such that $\underbrace{1 + \ldots + 1}_{n \text{ times}} = 0$. If no such n exists, then the field is of **characteristic 0**.

Remark

Gauss-Jordan elimination holds in any field. However, many results in this class will only hold for fields of characteristic 0, so we will assume that fields are characteristic 0 unless stated otherwise.

In search of a field that is less familiar than $\mathbb{R}, \mathbb{Q}, \mathbb{C}$, with nonzero characteristic, we make use of the following theorem.

Theorem 1.5

Let $a, n \in \mathbb{Z}$ be given, with $n \neq 0$. Then there exist unique q, r such that $a = qn + r$, with $0 \leq r < n$. In this case, q is the greatest integer such that $qn \leq a$, and r is $a \pmod{n}$.

The uniqueness from the above theorem allows us to define congruence modulo n.

Definition 1.10

We write $a \cong b \pmod{n} \iff n|a - b$.

We can use this definition to define the following class of fields:

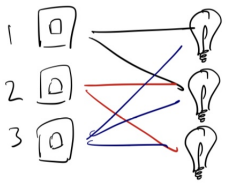
Definition 1.11

For p prime, we define the **finite field** $\mathbb{F}_p = \{\overline{0}, \ldots, \overline{p-1}\}$. We define the operations $+_p, *_p$ modulo p. That is, we have

- $\overline{a} +_p \overline{b} := \overline{(a+b) \pmod{p}}$
- $\overline{a} *_p \overline{b} := \overline{(a*b) \pmod{p}}$

It can be shown that for any $\overline{a} \in \mathbb{F}_p$ with $\overline{a} \neq \overline{0}$, there exists $\overline{b} \in \mathbb{F}_p$ such that $\overline{a} *_p \overline{b} = \overline{1}$. Moreover, associativity holds. So $(\mathbb{F}_p, +_p, *_p)$ is a field for every prime p. Conventionally, the bars are omitted, and the specific field is simply indicated.

In addition to being of interest due to having different characteristic than the fields we have worked in so far, finite fields are also practically useful.



Example 1.5

Suppose we have three switches and three lamps, all of which are initially off. The switches are connected to the lamps as shown above.

Suppose we want to make Lamp 1 turn on, but have Lamp 2 and 3 off.

Define the indicator variables $x_1, x_2, x_3 \in \mathbb{R}$, where each variable is 1 if the corresponding lamp is on, and 0 otherwise. Then the system can be represented as the following system of equations:

$$\begin{cases} x_1 + x_3 = 1 \\ x_1 + x_2 + x_3 = 0 \\ x_2 + x_3 = 0 \end{cases} \iff \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right]$$

$$\xrightarrow[\text{in } \mathbb{F}_2]{RREF(A)} \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right]$$

So we need to turn on switches 2 and 3, and leave 1 off.

Chapter 2

Matrices and Linear Transformations

2.1 Matrices

So far we have only used matrices to represent linear systems, and we have defined the EROs that we can use to act on matrices. We will now offer more definitions that will allow us to use matrices to represent linear maps.

Definition 2.1

Given a field $\mathbb{F}$, the set of $n \times m$ matrices with entries in $\mathbb{F}$ is denoted $M_{n\times m}(\mathbb{F})$.

Definition 2.2

Given two matrices $A, B \in M_{n\times m}(F)$ and a scalar $\lambda \in \mathbb{F}$, addition is defined entry-wise and scalar multiplication is done to all entries. That is, $(a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij})$ and $\lambda(a_{ij}) = (\lambda a_{ij})$.

Definition 2.3

The **dot product** of two column vectors of the same length is

$$\overrightarrow{A} \cdot \overrightarrow{B} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + \ldots + a_n b_n = \sum a_i b_i \in \mathbb{F}$$

Definition 2.4

Let $A \in M_{n\times m}(\mathbb{F})$, $x \in \mathbb{F}^m$. Suppose we denote the i-th row of A by w_i. Then we define

$$Ax \coloneqq \begin{bmatrix} w_1 \cdot x \\ \vdots \\ w_n \cdot x \end{bmatrix} \in \mathbb{F}^n$$

Remark

The observant reader may notice that the w_i are row vectors, not column vectors, and we have not defined the dot product of a row and column vector. This is easily remedied by instead considering $w_i' = w_i^T$.

Given these definitions, we can easily demonstrate basic matrix properties:

Theorem 2.1

Let $A \in M_{n\times m}(\mathbb{F})$, $x, y \in \mathbb{F}^m$, $\lambda \in \mathbb{F}$. Then

- $A(x + y) = Ax + Ay$
- $A(\lambda x) = \lambda(Ax)$

We define certain distinguished vectors:

Definition 2.5

Let $\mathbb{F}$ be a field. Consider $\mathbb{F}^m$. Then we define

$$e_i \coloneqq \begin{bmatrix} \delta_{ij} \\ \vdots \\ \delta_{ij} \end{bmatrix}$$

where the Kroncker delta δ_{ij} is 1 if $i = j$ and 0 otherwise.

Observe that for any matrix A, $Ae_i = V_i$ is simply the ith column of A. So we can rewrite the formula for matrix-vector multiplication as

$$Ax = \begin{bmatrix} V_1 & \dots & V_m \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} = \sum x_i V_i \in \mathbb{F}^m$$

Proof. Note that $x = \sum x_i e_i$. Applying the properties we have already discovered, we have $Ax = A\sum x_i e_i = \sum x_1(Ae_i) = \sum x_i V_i$. □

Remark

Let $A \in M_{m\times n}, b \in \mathbb{F}^m$. Then solving the augmented matrix system $[A|b]$ is equivalent to solving the equation $Ax = b$, where $x \in \mathbb{F}^n$.

To this point, we have not yet defined multiplication of two matrices. We will do so now.

Definition 2.6

Suppose $A \in M_{m\times k}(\mathbb{F})$, $B \in M_{k\times n}(\mathbb{F})$. Denote the columns of B $v_1, \ldots, v_n$. Then we define

$$AB \in M_{m\times n}(\mathbb{F}) = \begin{bmatrix} Av_1 & Av_2 & \ldots & Av_n \end{bmatrix}$$

Observe that the i, jth entry of AB is the dot product of the ith row of A and the jth column of B: $AB = (A_i \cdot B^j)$.

Remark

Note that the existence of AB does not imply that BA is even well defined. If it is, $AB \neq BA$ in general.

The key properties of matrices are as follows. Assume always that A, B, C are matrices over $\mathbb{F}$ of appropriate sizes so that products are defined, and that $\lambda \in \mathbb{F}$.

- $(AB)C = A(BC)$ (associativity)
- $A(B+C) = AB+AC, (A+B)C = AC+BC$ (left and right distributivity)
- $(\lambda A)B = \lambda(AB) = A(\lambda B)$ (commutativity with scalars)

In particular, if we consider vectors as $n \times 1$ matrices, we see that

$$A(Bv) = (AB)v,\ v \in \mathbb{F}^n$$

2.2 Linear Transformations

We will see that matrices have a very natural geometric interpretation. This is best seen by associating matrices with linear mappings between two spaces.

Definition 2.7

A **linear transformation** $T : \mathbb{F}^m \to \mathbb{F}^n$ is a function satisfying

- $T(x + y) = T(x) + T(y)$ for $x, y \in \mathbb{F}^m$
- $T(\lambda x) = \lambda T(x)$ for $x \in \mathbb{F}^m, \lambda \in \mathbb{F}$

Example 2.1

The function $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 2x$ is linear, while $g(x) = x^2$ is not. Rotation of $\mathbb{R}^2$ by 30 degrees counterclockwise is linear, which can be seen by putting elements of $\mathbb{R}^2$ in polar form and observing that rotation acts linearly on the angle (and not at all on the modulus).

The most important linear transformations for our study are those which can be associated with matrices.

Definition 2.8

Let $A \in M_{n\times m}(\mathbb{F})$. Define $T : \mathbb{F}^m \to \mathbb{F}^n$ by $T(x) \coloneqq Ax$ for $x \in \mathbb{F}^m$. We denote this by $T = L_A$.

Remark

Note that an $n \times m$ matrix is a transformation from $\mathbb{F}^m$ to $\mathbb{F}^n$.

The fact that the above transformation is indeed linear is immediate from the properties of matrix-vector multiplication from the previous section.

Remark

From the previous section, we know that $L_A(e_i) = Ae_i$ sends e_i to the ith column of A. Thus, the columns of A tell us precisely where the vectors e_i are sent after the mapping L_A. This seems to imply that the matrix of a linear transformation (and hence the transformation itself) is completely determined by where the e_i are sent. In fact we will see that this is true in general for linear transformations represented by matrices.

The corresponding between matrices and linear transformations will allow us to formulate corresponding results in both geometric and algebraic terms. We have already shown that matrices can be associated with linear transformations. It is of interest to determine when a linear transformation can be represented as a matrix.

Theorem 2.2

For any $T : \mathbb{F}^m \to \mathbb{F}^n$, there exists a unique $A \in M_{n\times m}(\mathbb{F})$ such that $T = L_A$. We denote this matrix $M(T)$. Specifically, the ith column of this matrix are given by $T(e_i)$.

Proof. Observe that if we consider the various e_i (standard basis vectors in F^m), then Me_i = the ith column of M for any matrix M. So we can construct a

matrix A which has as its ith column $T(e_i)$. Take some $x \in \mathbb{F}^m$. Then if

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

we can also say

$$x = x_1 e_1 + \ldots + x_m e_m$$

Since both $T(x)$ and Ax are linear, we only need to ensure that $T(e_i) = Ae_i$ for all $1 \leq i \leq m$. But by construction, Ae_i = the ith column of $A = T(e_i)$. So $T(x) = Ax$ for all $x \in \mathbb{F}^m$. Moreover, suppose that $T(x) = Bx$ as well for a matrix B. Then $T(e_i) = Ae_i$ = the ith column of A, and $T(e_i) = Be_i$ = the ith column of B. So the columns are identical and thus the matrices are identical. So we have a unique matrix. □

Moreover, a later theorem will show us that this is true of every linear transformation between finite vector spaces over a common field $\mathbb{F}$. Thus, we have shown a bijective correspondence between linear transformations from $\mathbb{F}^m \to \mathbb{F}^n$ and matrices in $M_{n\times m}(\mathbb{F})$.

Example 2.2

Suppose we define $T_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ by rotation of the plane counterclockwise by θ. Then to find $M(T_\theta)$, we simply find the values of $T_\theta(e_1)$ and $T_\theta(e_2)$. Consider first $T_\theta(e_1)$. Since e_1 has radius 1 and angle 0, we just need to find the coordinates of the point with radius 1 and angle θ. Geometrically, we can see that this is just $(\cos\theta, \sin\theta)$. Similarly, $T_\theta(e_2)$ has coordinates $(\cos(\theta + \pi/2), \sin(\theta + \pi/2) = (-\sin\theta, \cos\theta)$. So we can construct $M(T_\theta)$ by filling in the columns with $T_\theta(e_i)$, and thus

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

The correspondence between linear transformations and matrices gives us a convenient way to prove the properties of matrix multiplication.

Definition 2.9

Given two linear transformations $\phi : \mathbb{F}^m \to \mathbb{F}^n$ and $\psi : \mathbb{F}^n \to \mathbb{F}^k$, the **composition** $(\psi \circ \phi) : \mathbb{F}^m \to \mathbb{F}^k$ is defined by $(\psi \circ \phi)(x) = \psi(\phi(x))$ for all $x \in \mathbb{F}^m$.

Theorem 2.3

The composition of two linear transformations is a linear transformation.

Proof. Let $\phi : \mathbb{F}^m \to \mathbb{F}^n$, $\psi : \mathbb{F}^n \to \mathbb{F}^k$ be given. Let $x, y \in \mathbb{F}^m$ and $\lambda \in \mathbb{F}$ be arbitrary. Then $(\psi \circ \phi)(\lambda x + y) = \psi(\phi(\lambda x + y)) = \psi(\lambda\phi(x) + \phi(y)) = \psi(\lambda\phi(x)) + \psi(\phi(y) = \lambda\psi(\phi(x)) + \psi(\phi(y)) = \lambda(\psi \circ \phi)(x) + (\psi \circ \phi)(y)$. □

Using this definition, we can now prove the following:

Theorem 2.4

Given two linear transformations $\phi : \mathbb{F}^m \to \mathbb{F}^n$ and $\psi : \mathbb{F}^n \to \mathbb{F}^k$, we have $M(\psi \circ \phi) = M(\psi)M(\phi)$.

Proof. let $x \in \mathbb{F}^m$ be arbitrary. Then $(\psi \circ \phi)(x) = \psi(\phi(x)) = \psi(M(\phi)x) = M(\psi)(M(\phi)x) = (M(\psi)M(\phi))x$. But $M(\psi \circ \phi)x = (\psi \circ \phi)(x) = (M(\psi)M(\phi))x$. Since there is a unique matrix $M(\psi \circ \phi)$, we must have $M(\psi \circ \phi) = M(\psi)M(\phi)$.

Algebraically, we can also see that the ith column of $M(\psi \circ \phi)$ is given by $(\psi \circ \phi)(e_i)$. This is $\psi(\phi(e_i)) = \psi(M(\phi)e_i = \psi(i\text{th column of } M(\phi) = M(\psi)(i\text{th column of } M(\phi) = i\text{th column of } M(\psi)M(\phi)$ by definition. So we see that our definition of matrix multiplication makes sense here. □

Theorem 2.5

Matrix multiplication is associative, distributive, and commutes with scalar multiplication.

Proof. By translating these into linear transformations, we automatically get associativity (function composition is always associative). Showing commutativity with scalar multiplication is similarly easy. Left and distributivity is longer but not difficult in terms of linear transformations. □

Example 2.3

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation which first projects onto the line $x = y$ and then reflects across the y-axis. Then to find $M(T)$, we can first find the matrix representing each of the individual transformations, and then multiply them. To find the projection matrix, we see that both $(1, 0)$ and $(0, 1)$ should get projected to the point $(1/2, 1/2)$. So our projection matrix is

$$M(P) = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

The reflection matrix is given by

$$M(R) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$

So we can find $M(T)$ by multiplying:

$$M(T) = M(R \circ P) = M(R)M(P) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} \left(\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \right) = \frac{1}{2} \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$$

Summary

Matrix - Linear Transformation Correspondence

For any matrix $M \in M_{n \times m}(\mathbb{F})$, the associated linear transformation $L_M : F^m \to F^n$ is defined by $L_M(x) \coloneqq Mx$.

For any linear transformation $L : F^m \to F^n$, the associated matrix (assuming the standard basis for F^m) is $M(L)$, where the ith column of $M(L)$ is given by $L(e_i)$.

Given two linear transformations S, T, we have $M(S \circ T) = M(S)M(T)$, $M(S + T) = M(S) + M(T)$ (whenever these make sense).

Chapter 3

Vector Spaces

3.1 Vector Spaces and Subspaces

It is of interest to generalize the results that we have achieved without specifically invoking the specific choice of coordinates induced by $\mathbb{F}^n$ (or even any choice of coordinates at all). We do this by abstracting $\mathbb{F}^n$ into a vector space:

Definition 3.1

A **vector space** over a field $\mathbb{F}$ is a set V with an operation $+ : V \times V \to V$ and an operation $* : \mathbb{F} \times V \to V$ ($*$ is denoted with juxtapostion), such that the following conditions are satisfied:

- $x + y = y + x$ for any $x, y \in V$
- $(x + y) + z = x + (y + z)$ for any $x, y, z \in V$
- $(ab)v = a(bv)$ for any $a, b \in \mathbb{F}$, $v \in V$
- There exists $e \in V$ such that $e + v = v + e = v$ for any $v \in V$ (this element is unique and denoted $\overrightarrow{0}$).
- For any $x \in V$, there exists $y \in V$ such that $x + y = \overrightarrow{0}$ (this element is unique and denoted $-x$).
- $1v = v1 = v$ for any $v \in V$.
- $(a + b)v = av + bv$ for any $a, b \in \mathbb{F}$, $v \in V$
- $a(u + v) = au + av$ for any $a \in \mathbb{F}$, $u, v \in V$

A vector space over $\mathbb{R}$ is called a **real vector space** and a vector space over $\mathbb{C}$ is called a **complex vector space**. We call elements of the set V **vectors** and elements of the field $\mathbb{F}$ scalars.

Example 3.1

Some examples of vector spaces:

- $\mathbb{R}^n$ is a real vector space.
- Less trivially, a line in $\mathbb{R}^2$ through the origin is a real vector space (as well as any hyperplane through the origin in $\mathbb{R}^n$).
- If we denote by $\mathbb{R}[x]_{\leq n}$ the set of univariate polynomials in $\mathbb{R}$ with degree at most n, then this is another real vector space. This is also the case if we consider polynomials of any degree ($\mathbb{R}[x]$).
- If we consider the set of functions (or set of continuous functions, or smooth functions) from $[0, 1] \to \mathbb{R}$, then this is another real vector space over.
- Any field $\mathbb{F}$ is a one dimensional vector space over itself.
- Lastly, $\mathbb{C}$ can be considered a real vector space of dimension 2, or a complex vector space of dimension 1.

Some nonexamples:

- If we consider polynomials of degree *exactly* n, then it is not a real vector space, since there is no additive identity, and in any case the set is not closed under addition.
- The set of n-tuples in $\mathbb{R}^n$, with $a_1 \geq 0$ is not a vector space, since it's not closed under scalar multiplication and it is missing some additive identities.
- The x-axis unioned with the y-axis.

One way to construct vector spaces is to extract them from larger vector spaces; in other words, to identify a subspace:

Definition 3.2

If $(V, +_V, *_V)$ is a vector space over $\mathbb{F}$, then W is a **subspace** of V if $W \subseteq V$ and $(W, +_V, *_V)$ is a vector space.

Since we use the same operations, many of the axioms of vector spaces are automatically inherited for the subspace from the parent space. Thus, we only need to check a few conditions:

Theorem 3.1

Given a vector space V, a subset W is a subspace of V if and only if $\overrightarrow{0} \in W$ and W is closed under $+_V, *_V$.

Corollary

If W is a subspace of V and U is a subspace of W, then U is a subspace of V.

Proof. Since we use the same operations in W and in V, closure in U from W implies closure in U from V. Since $0_W = 0_V$, $0_W \in U$ implies $0_V \in U$. □

Example 3.2

$\mathbb{R}$ is a subspace of $\mathbb{R}^2$. Any line passing through the origin in $\mathbb{R}^2$ is a subspace of $\mathbb{R}^2$.

3.2 Span and Linear Independence

Definition 3.3

Let V be a vector space over $\mathbb{F}$. Then given some vectors $\{v_1, \ldots, v_m\}$, the **span** of these vectors is a subspace of V given by

$$\text{span}(v_1, \ldots, v_m) = \{a_1 v_1 + a_2 v_2 + \ldots + a_m v_m | a_i \in \mathbb{F}\}$$

Theorem 3.2

Given $v_1, \ldots, v_m \in V$, then $\text{span}(v_1, \ldots, v_m) \subseteq V$ is a subspace of V. Moreover, it is the smallest subspace of V containing each of the $v_1, \ldots, v_m$.

Example 3.3

The span of $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}$ is the xy-plane. In general, the span of n vectors in $\mathbb{R}^m$, $n \leq m$, is a k-dimensional hyperplane, where $k \leq n$ ($\leq$ because we may have collinear vectors).

Example 3.4

Let $V = C(\mathbb{R})$ (the set of continuous real functions). Then $\cos(2x) \in \text{span}(\cos^2(x), 1)$ since $\cos(2x) = 2\cos^2(x) - 1$.

Note that the span of a number of vectors is the set of linear combinations of the vectors. Then in this way, a set of vectors is distinguished if the entire space is a linear combination of the vectors.

Definition 3.4

A set of vectors $\{v_1, \ldots, v_m\} \subseteq V$ **spans** a vector space V if $V = \text{span}(v_1, \ldots, v_m)$.

In particular, though we have not defined dimension, we can use this definition to distinguish between finite and infinite dimensional vector spaces.

Definition 3.5

A vector space V is called **finite dimensional** if there is a finite list of vectors which spans it. It is called **infinite dimensional** if it is not finite dimensional.

Here, we are motivated by the key question of asking how we can use less data to encode a vector space. So given a subspace $W \subseteq V$, such that $W = \text{span}(v_1, \ldots, v_m)$, then we want to find fewer vector that still span W.

Definition 3.6

Given an ordered list of vectors $v_1, \ldots, v_m$, a vector, we say that v_i is **redundant** if $v_i \in \text{span}(v_1, \ldots, v_{i-1})$.

Example 3.5

Suppose we consider the ordered set

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

Then $\begin{bmatrix}1\\0\end{bmatrix}$ isn't redundant, since the first vector is never redundant (span of empty set is the trivial vector space). $\begin{bmatrix}1\\1\end{bmatrix}$ isn't redundant, since it

isn't in span($\begin{bmatrix}1\\0\end{bmatrix}$. However, $\begin{bmatrix}0\\1\end{bmatrix}$ is redundant, since

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

This leads to the observation that if we remove a redundant vector, then the span of the resulting set is the same as the original. Suppose we reduce our spanning vectors by recursively removing all redundant vectors. Is it possible that a vector can still be removed?

Theorem 3.3

If $\text{span}(v_1, \ldots, v_m) = \text{span}(\{v_1, \ldots, v_m\} \setminus \{v_i\})$, then $\{v_1, \ldots, v_m\}$ contains a redundant vector.

Proof. Since $v_i \in \text{span}(v_1, \ldots, v_m), v_i \in \text{span}(\{v_1, \ldots, v_m\} \setminus \{v_i\})$, so we can write

$$v_i = a_1 v_1 + \ldots + a_{i-1} v_{i-1} + a_{i+1} v_{i+1} + \ldots a_m v_m$$

Then we can rearrange this to write

$$a_1 v_1 + \ldots + a_m v_m = 0$$

where $a_i = -1$. If $a_m \neq 0$, then we could move v_m to the right and show that v_m is redundant. In the event that $a_m = 0$, we simply choose a_{m-1}, assuming that is nonzero. So we choose the largest k such that $a_k \neq 0$ (at least one such k exists since $a_i \neq 0$). Then

$$v_k = -\frac{a_1}{a_k} v_1 - \ldots - \frac{a_{k-1}}{a_1} v_{k-1} \quad (+0 v_{k+1} + \ldots 0 v_m)$$

So v_k is redundant. □

Definition 3.7

Given a vector space V over $\mathbb{F}$, we say that a set of vectors $\{v_1, \ldots, v_m\} \in V$ are **linearly independent** if $a_1 v_1 + \ldots + a_m v_m = 0$ implies that $a_1 = \ldots = a_m = 0$ (with $a_i \in F$). We call them **linearly dependent** if there is a linear relation $a_1 v_1 + \ldots + a_m v_m = 0$ with at least one $a_i \neq 0$.

Then we have the following:

Theorem 3.4

The following statements are equivalent:

- The list $\{v_1, \ldots, v_n\}$ is linearly independent.
- There is no redundant vector in the list.
- Removing any vector makes the span strictly smaller.

Lemma: Linear Independence Lemma

Let $\{v_1, \ldots, v_n\}$ be linearly independent. Let $v \in V$ be arbitrary. Then $v \in \text{span}(v_1, \ldots, v_n)$ or $\{v_1, \ldots, v_n, v\}$ are linearly independent. Moreover, only one of those cases can hold.

Proof. Suppose $v \notin \text{span}(v_1, \ldots, v_n)$. Note that there are no redundant vectors in $\{v_1, \ldots, v_n\}$ by assumption. But since v is not in the span, the list $\{v_1, \ldots, v_n, v\}$ also has no redundant vectors, and is thus linearly independent. Thus at least one of the two cases is true.

To show that only one can hold, suppose $v \in \text{span}(v_1, \ldots, v_n)$. Then by definition, $\{v_1, \ldots, v_n, v\}$ contains the redundant vector v, and is thus not linearly independent. Thus we cannot have both cases at once, so we have exactly one. □

3.3 Basis and Dimension

As we can see from the previous section, any vector in the span of a list of other vectors can be written as a linear combination of the vectors in the list. Thus, a spanning set allows us to write any vector in a vector space in terms of only the vectors in a certain list. Moreover, our discussion of linear independence shows that a linearly independent spanning set is something of an "optimal" set in terms of requiring a minimal number of vectors.

Definition 3.8

A **basis** of a vector space V is a (possibly finite) linearly independent list of vectors $\{v_1, \ldots\} \subseteq V$ that spans V.

Example 3.6

Here are some examples of bases of vector spaces.

1. $e_1, \ldots, e_n$ is a basis of $\mathbb{R}^n$.

2. $\begin{bmatrix}1\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}$ is a basis of $\mathbb{R}^2$.

3. $1, x, x^2, \ldots, x^n$ is a basis of $\mathbb{R}[x]_{\leq n}$.

4. $1, x, x^2, \ldots$ is a basis of $\mathbb{R}[x]$.

5. Any two nonzero, noncollinear vectors in $\mathbb{R}^2$ is a basis of $\mathbb{R}^2$.

6. Any three nonzero, nonplanar vectors in $\mathbb{R}^3$ is a basis of $\mathbb{R}^3$.

However, we will see that this definition actually implies many more significant results, especially when considering finite dimensional vector spaces. For instance, given a spanning set, every vector can be written as a linear combination of the spanning vectors. But by adding the condition of linear independence, we find that this representation is unique. In fact, this correspondence is even more powerful:

Theorem 3.5

$v_1, \ldots, v_n$ is a basis of V if and only if every vector $v \in V$ can be written uniquely as $v = a_1v_1 + \ldots + a_nv_n$ for $a_i \in \mathbb{F}$.

Proof. ($\implies$) Let $v \in V$ be arbitrary. Then since $v \in \text{span}(v_1, \ldots, v_n)$, $v = a_1v_1 + \ldots + a_nv_n$. To prove uniqueness, suppose that $v = a_1'v_1 + \ldots + a_n'v_n$. Then we have

$$(a_1 - a_1')v_1 + \ldots + (a_n - a_n')v_n = (v - v) = 0$$

But by linearly independence, $(a_i - a_i') = 0$ for every i, and thus $a_i = a_i'$. So this representation is unique.

($\impliedby$) If every $v \in V$ is a linear combination of $v_1, \ldots, v_n$, then $\{v_1, \ldots, v_n\}$ spans V. To show that $\{v_1, \ldots, v_n\}$ is linearly independent, suppose $a_1v_1 + \ldots + a_nv_n = 0$. Then since 0 has a unique representation, and $0v_1 + \ldots + 0v_n$, we must have $a_i = 0$ for each i. So $\{v_1, \ldots, v_n\}$ is linearly independent and must be a basis. □

Thus, if we have chosen a basis, then this justifies the use of a n-tuple of coordinates to represent each vector, at least in the finite dimensional case.

We now turn our attention to the study of how to construct bases. We first consider spanning lists. Suppose we have a list of vectors $v_1, \ldots, v_n \in V$ that spans V. If we remove all redundant vectors, then the new list is linearly independent. Moreover, since removing redundant vectors doesn't change the span, this new list still spans V, and is thus a basis of V. So we have proved the following theorem:

Theorem 3.6

Every finite spanning set S of a vector space V can be reduced to a basis $B \subseteq S$.

On the other hand, we can also take a linearly independent set and extend it to a basis. Suppose $v_1, \ldots, v_n$ is linearly independent. By the linear independence lemma, if $v \notin \text{span}(v_1, \ldots, v_n)$, then $\{v_1, \ldots, v_n, v\}$ is linearly independent. So we can choose some vector $v_{n+1} \notin \text{span}(v_1, \ldots, v_n)$ and add it to the list. The new list is still linearly independent, so we repeat until this process stops (which happens only in the finite dimensional case). We can also use results from the previous section to prove the following:

Theorem 3.7

Every finite dimensional vector space has a finite basis.

Proof. By definition, every finite dimensional vector space has a finite spanning set. Then by the previous theorem, it can be reduced to a basis. □

Corollary

Any linearly independent set of vectors in a finite dimensional vector space V can be extended to a basis.

Proof. Suppose we we have $\{v_1, \ldots, v_n\}$ linearly independent. Then suppose $V = \text{span}(w_1, \ldots, w_m)$. Then we simply take the set $v_1, \ldots, v_n, w_1, \ldots, w_m$ and remove redundant vectors, we have a linearly independent spanning set. Moreover, we know none of the original v_i were linearly independent, so they will still be in the new set. □

Example 3.7

Suppose we consider the plane $\{\langle x, y, z\rangle | x + y + z = 0\}$. Then if we choose any vector in the plane, and then any vector in the plane which is not in the span of the first, we have a basis of the plane.

We can summarize the results of this discussion in the following:

Theorem 3.8

The following statements are equivalent:

- $\{v_1, \ldots, v_n\}$ is a basis of V.
- Every $v \in V$ can be written uniquely as a linear combination of $\{v_1, \ldots, v_n\}$.
- $\{v_1, \ldots, v_n\}$ is a maximal linearly independent subset of V (that is, any set of which it is a strict subset is linearly dependent).
- $\{v_1, \ldots, v_n\}$ is minimal spanning set of V (that is, any strict subset does not span V).

Proof.

1. We have already proved 1 $\iff$ 2.
2. 1 $\implies$ 3 because any possible addition is already in the span, and thus by the linear independence lemma, the new list is linearly dependent.
3. 3 $\implies$ 1 because if it were not a spanning set, then we could add a new vector that is not in the span, and the new set would be linearly independent by the linear independence lemma.
4. 1 $\implies$ 4 because removing any vector from a linearly independent set makes the span strictly smaller.
5. 4 $\implies$ 1 If the set were not linearly independent, then there would be a redundant vector, that we could remove to create a smaller spanning set. Since this is assumed to not be the case, the set is linearly independent and thus is a basis. □

Moreover, we have shown that every finite dimensional vector space has a finite basis. There are stronger results we can show as well. First, we begin a discussion of linear independence of the columns of a matrix:

Consider a matrix A with column vectors $v_1, \ldots, v_n$. Then for any x, $Ax = x_1v_1 + \ldots + x_nv_n$. Thus, the column vectors are linearly independent if and only if the equation $Ax = 0$ has only the zero solution. In particular, if we row reduce the matrix, the independence of the column vectors does not change.

Example 3.8

Consider the matrix

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

If we calculate the RREF we get

$$\text{RREF}(A) = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Clearly, the third column vector is redundant; moreover, it corresponds to a free variable in the system. Thus we conclude that free variables correspond with redundant column vectors.

We can now show that every basis of a finite dimensional vector space has the same length (for infinite dimensional vector spaces, they are equicardinal).

Theorem 3.9

If $\{v_1, \ldots, v_n\}$ and $\{w_1, \ldots, w_m\}$ are bases of V, then $n = m$.

Proof. If $n \neq m$, suppose without loss of generality that $n > m$. Then we can set up a matrix A with column vectors $v_1, \ldots, v_n$. Then since we have more variables than equations, the homogeneous system $Ax = 0$ admits a nonzero solution and must have a free variable. This shows that $v_1, \ldots, v_n$ is linearly dependent, which contradicts the statement that it is a basis of V. So we must have that $n = m$. □

This allows us to define the dimension of a vector space:

Definition 3.9

Given a finite dimensional vector space V, the **dimension** of the vector space is the unique length of every basis of V.

It immediately follows that for any list of vectors $\mathcal{V}$, then if $\mathcal{V}$ is linearly independent, $|\mathcal{V}| \leq \dim V$, and if it is a spanning set, then $|\mathcal{V}| \geq \dim V$. As a result, we have

Theorem 3.10

If a vector space V has $\dim V = m$, then any list of vectors $\{v_1, \ldots, v_m\}$ the following statements are equivalent:

- $\{v_1, \ldots, v_m\}$ is a basis.
- $\{v_1, \ldots, v_m\}$ is linearly independent.
- $\{v_1, \ldots, v_m\}$ span V.

Proof. 1 implies 2, 3 by definition. If the list is linearly independent, then any nontrivial extension of the list will be longer than $\dim V$ and is thus not linearly independent. So the list is a maximal linearly independent set and therefore a basis. The proof for the other case is similar. □

Suppose we consider subspaces of the form $V \subseteq \mathbb{R}^n$ defined as the set of solutions to the equation $Av = 0$ for some matrix A (in other words, $V = \ker A$). We will now explore a general method for computing a basis of V.

Example 3.9

Let A be the matrix

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

If we compute $\mathrm{RREF}(A)$ we find

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

Since $\ker A = \ker \mathrm{RREF}(A)$, we simply observe that

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

In other words, we let each of the free variables equal 1 and the rest equal 0, and find the values of the determined variables.

Lemma

Suppose U is a subspace of V. If V is finite dimensional, then U is finite dimensional with $\dim(U) \leq \dim(V)$.

Proof. We seek to construct a basis of U. Choose a vector in U to begin our basis. Then choose a vector in U but not in the span of the previous vector. Continue this process until a basis is constructed. This must be finite because we cannot have a linearly independent list in V with length longer than $\dim(V)$. So we have a finite dimensional basis of U, and its length is $\dim(U) \leq \dim(V)$. □

Summary

Span, Linear Independence, Basis, Dimension

- The span of a set of vectors is the set of linear combinations of those vectors.
- A spanning set of a vector space V is a set of vectors v such that $V \subseteq \text{span}(v)$.
- A set of vectors is linearly independent if the only linear relation between them is the trivial one.
- A basis of a vector space V can be equivalently defined as:
  - A set of vectors which spans V and is linearly independent.
  - A minimal spanning set.
  - A maximal linearly independent set.
- Any spanning set can be reduced to a basis.
- Any linearly independent set can be extended to a basis.
- Every spanning list is longer than or equal length to every linearly independent list.
- Every basis of a finite dimensional vector space has equal length, denoted $\dim(V)$.

Chapter 4

Linear Maps and Vector Spaces

4.1 Linear Maps over Abstract Spaces

As refresher, we restate the definition of a linear mapping between vector spaces, this time in terms of abstract vector spaces:

Definition 4.1

A **linear map** between two vector spaces V, W over a common field $\mathbb{F}$ is a function $T : V \to W$ satisfying

- $T(v + w) = T(v) + T(w)$ for any $v, w \in V$
- $T(\lambda v) = \lambda T(v)$ for any $v \in V$, $\lambda \in \mathbb{F}$.

Example 4.1

Here are more examples of linear maps, this time between vector spaces other than $\mathbb{F}^n$:

- The transformation $F : \mathbb{R}[x] \to \mathbb{R}[x]$ defined by $F(p) \coloneqq x^2 p$.
- The transformation $D : \mathbb{R}[x] \to \mathbb{R}[x]$ defined by $D(p) = p'$.
- The transformation $T : \mathbb{R}[x] \to \mathbb{R}$ defined by $T(p) = \int_0^1 p$.

We can also generalize a previous theorem:

Lemma

Suppose $v_1, \ldots, v_n$ is a basis of V. Let $w_1, \ldots, w_n$ be an arbitrary list of n vectors in W. Then there is a unique linear map $T : V \to W$ such that $T(v_i) = w_i$ for each $1 \leq i \leq n$.

Proof. Suppose $T(v_i) = S(v_i) = w_i$ for $T, S : V \to W$. Then for any vector $v \in V$, there is a unique representation of v in terms of the basis vectors $v_1, \ldots, v_n$. Then by linearity, $T(v) = S(v)$ for each $v \in V$ and thus $T = S$. So such a map must be unique.

For any vector $v \in V$, there is a unique representation of v as $a_1v_1 + \ldots + a_nv_n$. Define $T(v) \coloneqq a_1w_1 + \ldots + a_nw_n$. Then this is a well-defined linear transformation with $T(v_i) = w_i$ by inspection. □

Now that we have considered more abstract spaces, we can construct new vector spaces from linear transformations over other spaces.

Definition 4.2

Let V, W be vector spaces over $\mathbb{F}$. Then the set of linear transformations from V to F is denoted $\mathcal{L}(V, W)$. Moreover, $\mathcal{L}(V, W)$ is a vector space over $\mathbb{F}$.

To show that $\mathcal{L}(V, W)$ is a vector space, we define addition and multiplication in the natural way:

$$(T + S)(v) = T(v) + S(v),\ (\lambda T)(v) = \lambda T(v)$$

Definition 4.3

Let $T : V \to W$ be a linear transformation. Then we define the **kernel** of T as the preimage of 0_W, in other words, $\ker T = \{v \in V : T(v) = 0\}$. We define the **image** of T as the set of points mapped to by T, in other words, $\operatorname{im} T = \{T(v) : v \in V\}$.

Clearly, $\ker T \subseteq V$, and $\operatorname{im} T \subseteq W$. Moreover, $\ker T$ is a subspace of V, and $\operatorname{im} T$ is a subspace of W.

If we consider linear transformations of the form $T = Ax$ for some A, then we can easily find the bases of $\operatorname{im} T$ and $\ker T$ by calculating RREF A. Then $\ker T$ is the solutions to the equation $Ax = 0$, and $\operatorname{im} T$ is the span of all the columns. So a basis of $\operatorname{im} T$ is the set of columns without the redundant vectors.

4.2 Injective and Surjective Maps

Definition 4.4

Let $f : X \to Y$ be a function between two sets (not necessarily a linear function, and not necessarily between vector spaces). Then we call f **injective** if $f(x) = f(y)$ implies $x = y$ for $x, y \in X$. We call f **surjective** if for all $y \in Y$, there is some $x \in X$ with $f(x) = y$.

In other words, f is injective if it distinguishes points in the domain, and f is surjective if it maps to the entire codomain.

Definition 4.5

We call $f : X \to Y$ **bijective** if it is both injective and surjective.

The importance of bijectivity is that the equation $f(x) = y$ has exactly one solution for every $y \in Y$. This allows us to construct an inverse function for f.

Proposition 4.1

A function $f : X \to Y$ is bijective if and only if there exists $g : Y \to X$ such that $g \circ f = \mathrm{id}_X$ and $f \circ g = \mathrm{id}_Y$.

In particular, we will explore bijectivity of linear transformations between vector spaces:

Proposition 4.2

Let $T : V \to W$ be a linear transformation between vector spaces. Then we have

- T is surjective if and only if $\operatorname{im} T = W$.
- T is injective if and only if $\ker T = \{0\}$.

Proof. 1 is clear.

Proof of 2: ($\implies$) Choose $v \in \ker T$. Then $T(v) = 0$, but $T(0) = 0$ for every linear transformation. By injectivity, we conclude $v = 0$. So $\ker T = \{0\}$.

($\impliedby$) Choose $v_1, v_2 \in V$ with $T(v_1) = T(v_2)$. Then by linearity, $T(v_1) - T(v_2) = T(v_1 - v_2) = 0$. Then $v_1 - v_2 \in \ker T \implies v_1 - v_2 = 0$. So $v_1 = v_2$ and thus T is injective. □

This allows us to prove one of the most fundamental results about linear mappings over finite dimensional vector spaces:

Theorem 4.3: Rank-Nullity Theorem

Let $T : V \to W$ be a linear transformation between finite dimensional vector spaces. Then $\dim \ker T + \dim \operatorname{im} T = \dim V$ (nullity + rank = dimension of source space).

Proof. Consider $\ker T$, a subspace of V. Then we know $\ker T$ is finite dimensional, so we can select a basis $\mathcal{B}_1 = \{v_1, \ldots, v_k\}$ for $\ker T$ (with $k = \dim \ker T$). Then $\mathcal{B}_1$ is linearly independent in V, so we can extend it to a basis of V: $\mathcal{B}_1 \to \mathcal{B}_1 \cup \mathcal{B}_2$. In particular, we have $\mathcal{B}_2 = \{w_1 \ldots, w_l\}$. Then we claim that $T(\mathcal{B}_2)$ is a basis of $\operatorname{im} T$.

To prove this, choose some $u \in \operatorname{im} T$. Then $u = T(x)$ for some $x \in V$. Then we can write $x = a_1 v_1 + \ldots + a_k v_k + b_1 w_1 + \ldots + b_l w_l$. Then by linearity, we have

$$u = T(x) = \underbrace{a_1 T(v_1) + \ldots + a_k T(v_k)}_{\text{equal to } 0} + b_1 T(w_1) + \ldots + b_l T(w_l)$$

So every $u \in \operatorname{im} T$ can be written as a linear combination of the $T(w_i)$, and thus $T(\mathcal{B}_2)$ spans $\operatorname{im} T$.

To prove linear independence, suppose $b_1 T(w_1) + \ldots + b_l T(w_l) = 0$. Then by linearity, $T(b_1 w_1 + \ldots + b_l w_l) = 0$, so $b_1 w_1 + \ldots + b_l w_l \in \ker T$. Since $\mathcal{B}_1$ is a basis of $\ker T$, we can set this equal to a combination of the v_i:

$$b_1 w_1 + \ldots + b_l w_l = a_1 v_1 + \ldots + a_k b_k$$

But if we move the $a_i v_i$ to the left side, we get that the combination is equal to 0, and since $\mathcal{B}_1 \cup \mathcal{B}_2$ is a basis, every coefficient (in particular, the b_i) must all be 0. So $T(\mathcal{B}_2)$ is linearly independent and thus a basis of $\operatorname{im} T$.

Then we have $k = |\mathcal{B}_1| = \dim \ker T$, and $l = |\mathcal{B}_2| = |T(\mathcal{B}_2)| = \dim \operatorname{im} T$, so $\dim \ker T + \dim \operatorname{im} T = |\mathcal{B}_1| + |\mathcal{B}_2| + |\mathcal{B}_1 \cup \mathcal{B}_2| = \dim V$. □

Just as bijectivity of general functions implies an inverse function, so too does bijectivity of linear transformations.

Definition 4.6

A linear transformation $T : V \to W$ is **invertible** if there exists a linear transformation $S : W \to V$ such that $S \circ T = \mathrm{id}_V$ and $T \circ S = \mathrm{id}_W$.

Definition 4.7

A matrix $A \in M_{n \times n}(\mathbb{F})$ is **invertible** if there exists a matrix $B \in M_{n \times n}(\mathbb{F})$ such that $AB = BA = I_n$.

Then $T : V \to W$ is invertible if and only if T is bijective and the inverse T^{-1} is also linear. However, we will see that linearity of a bijective T implies linearity of T^{-1}, so only the bijective condition is necessary to check.

Proposition 4.4

$T : V \to W$ is invertible if and only if T is bijective.

Proof. Suppose T is linear and bijective, with $T^{-1} : W \to V$. Then choose $w_1, w_2 \in W$ and $\lambda \in \mathbb{F}$. Then $w_1 = T(v_1), w_2 = T(v_2)$ for some $v_1, v_2 \in V$. Then we have $T^{-1}(\lambda w_1 + w_2) = T^{-1}(T(\lambda v_1 + v_2)) = \lambda v_1 + v_2$. □

Given this, we also have the following fact:

Theorem 4.5

Let $T : V \to W$ be a linear transformation, and let $\mathcal{B} = \{v_1, \ldots, v_m\}$ be a basis of V. Then T is invertible if and only if $T(\mathcal{B}) = \{T(v_1), \ldots, T(v_m)\}$ is a basis of W.

Proof. ($\implies$) By injectivity, $\dim \ker T = 0$, so the rank nullity theorem implies that $\dim V = \dim \operatorname{im} T$. By surjectivity, we then have $\dim V = \dim \operatorname{im} T = \dim W$. So we have a list of m linearly independent vectors (proof is the same as rank-nullity proof) in a m-dimensional vector space, so $T(\mathcal{B})$ is a basis.

($\impliedby$) By a lemma in the previous section, we know that there is a unique $S : W \to V$ such that $S(T(v_i)) = v_i$ for each i. So $(S \circ T)(v_i) = v_i = \mathrm{id}_V(v_i)$ for each i, so by uniqueness $(S \circ T) = \mathrm{id}_V$ and similarly $(T \circ S) = \mathrm{id}_W$. So T is invertible. □

4.3 Isomorphisms

The significance of linear transformations over vector spaces is that linear transformations respect linearity from the vector space; that is, they respect the underlying vector space structure. However, for noninvertible linear transformations, we may lose some of this structure, since there is no way to invert the transformation and return to the original structure. So we say that invertible linear transformations preserve vector space structure, which is precisely the notion captured by an isomorphism.

Definition 4.8

We say that a linear transformation $T : V \to W$ is an **isomorphism** if T is invertible. We say that V, W are **isomorphic** (denoted $V \cong W$) if there exists an isomorphism $\varphi : V \to W$.

It is immediate that isomorphic congruence is an equivalence relation, since φ^{-1} and $\varphi \circ \phi$ are both isomorphisms if φ, ϕ are isomorphisms.

Theorem 4.6

Given finite dimensional vector spaces V, W, $V \cong W$ if and only if $\dim V = \dim W$. In particular, any finite dimensional vector space V over $\mathbb{F}$ is isomorphic to $\mathbb{F}^{\dim V}$.

Proof. ($\implies$) Let $\mathcal{B}$ be a basis of V. Then there is an isomorphism $T : V \to W$ by assumption. Then by Theorem 4.5, $T(\mathcal{B})$ is a basis of W, and by bijectivity, $\dim V = |\mathcal{B}| = |T(\mathcal{B})| = \dim W$.

($\impliedby$) Let $\{v_1, \ldots, v_n\}$ be a basis of V and $\{w_1, \ldots, w_n\}$ a basis of W. Then there exists T such that $T(v_i) = w_i$, which shows that T is isomorphic by Theorem 4.5. □

This leaves us with a number of useful facts about isomorphisms:

Theorem 4.7

Given finite dimensional vector spaces V, W with $\dim V = \dim W$, and a linear transformation $T : V \to W$, the following are equivalent:

- T is an isomorphism (is invertible).
- T is injective.
- T is surjective.

Proof. Immediate by Rank-Nullity Theorem. □

Note that for $n \times n$ matrices A which are invertible, we must have $\mathrm{RREF}(A) = I_n$. Thus we can find A^{-1} through the following process:

Denote by $[A|I_n]$ the matrix which has I_n appended to A. Then if we perform row operations on the combined matrix until we reach $[I_n|B]$, then we will have $B = A^{-1}$. Note that the remark about $\mathrm{RREF}(A)$ guarantees that we will never have a result here if A is not invertible, since noninvertible matrices are not row equivalent to I_n.

One application of this is to the space of solutions to differential equations. In particular, given a homogeneous ordinary differential equation, the set of solutions forms a vector space (specifically, a vector subspace of $C^\infty(\mathbb{R})$).

Example 4.2

Consider the differential equation $f" + f = 0$. Then $\cos t$ and $\sin t$ are linearly independent solutions to the differential equation, and in

fact every solution takes the form $a\cos t + b\sin t$. So we can say that $\{\cos t, \sin t\}$ forms a basis of the solution space, and it has dimension 2.

Theorem 4.8

Suppose we have a differential equation

$$a_k f^{(k)} + \ldots + a_1 f' + a_0 f = 0$$

with $a_i \in c^\infty(\mathbb{R})$. Then if we denote the set of solutions V, that is,

$$V = \{f \in C^\infty(\mathbb{R}) | a_k f^{(k)} + \ldots + a_1 f' + a_0 f = 0\}$$

Then $\dim V \leq k$. Moreover, on a sufficiently small interval $[-\delta, \delta] \subseteq \mathbb{R}$, then

$$V^* = \{f \in C^\infty([-\delta, \delta]) | \ldots\}$$

has $\dim V^* = k$.

The following proof is incorrect in its current form (because the differentiable functions do not form a complete metric space with the metric given). However, it is retained out of interest.

Proof. Note that if we introduce initial conditions, then we can isolate unique solutions to the differential equations. In particular, given some $v \in \mathbb{R}^k$, there exists a unique solution $f \in C^\infty([-\delta, \delta])$ such that

$$\begin{bmatrix} f(0) \\ f'(0) \\ \vdots \\ f^{(k-1)}(0) \end{bmatrix} = v$$

Then define a mapping $\phi : \mathbb{R}^k \to V$ such that $\phi(v)$ is that unique solution to the differential equation. In particular, we see that this mapping is linear. Clearly, it is surjective, since for any $f \in V$ we can simply build v by evaluating $f(0), f'(0), \ldots, f^{(k-1)}(0)$. We will prove that ϕ is also injective. Thus ϕ is an isomorphism of vector spaces.

Now, by way of example, consider the differential equation $f'' + f = 0$. Then from the differential equation, we have $g'(t) = -f(t)$. Then we have

$$\begin{cases} g'(t) = -f(t) \\ f'(t) = g(t) \end{cases} \iff \begin{bmatrix} g(t) \\ f(t) \end{bmatrix}' = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} g(t) \\ f(t) \end{bmatrix}$$

If we define $\overrightarrow{x}(t) \coloneqq \begin{bmatrix} g(t) \\ f(t) \end{bmatrix}$, with $\overrightarrow{x} : [-\delta, \delta] \to \mathbb{R}$, then we have reformulated this to be $\overrightarrow{x}'(t) = A\overrightarrow{x}(t)$ (where in this case our A is constant but in general

could depend on t).

Now suppose we fix some $v_0 \in \mathbb{R}^2$ so that $\overrightarrow{x}(0) = v_0$. Then in the general case, we are looking for the solutions to

$$\begin{cases} \overrightarrow{x}'(t) = A(t)\overrightarrow{x}(t) \\ \overrightarrow{x}(0) = v_0 \end{cases}$$

Note that since $A(t)$ is continuous, there is some $\delta > 0$ such that

$$\int_{-\delta}^{\delta} ||A(t)||_{op} dt = \int_{-\delta}^{\delta} \max_{\overrightarrow{x} \in \mathbb{R}^n \setminus \{0\}} |\frac{A(t)(\overrightarrow{x})}{\overrightarrow{x}}| dt = c < 1$$

Let S_δ be the (complete) metric space of differentiable functions $\overrightarrow{x} : [-\delta, \delta] \to \mathbb{R}^n$, with the metric $d(\overrightarrow{x}, \overrightarrow{y}) = \max_{t \in [\delta, \delta]} |\overrightarrow{x}(t) = \overrightarrow{y}(t)|$. Then define an operator $\phi : S_\delta \to S_\delta$ such that

$$\phi(\overrightarrow{x})(t) = v_0 + \int_0^t A(u)\overrightarrow{x}(u)du$$

Then we can observe that the statement that $\overrightarrow{x}$ is a solution:

$$\phi(\overrightarrow{x}) = \overrightarrow{x} \iff \begin{cases} v_0 = \overrightarrow{x}(0) \\ A\overrightarrow{x}(t) = \overrightarrow{x}'(t) \end{cases}$$

is equivalent to the statement that $\overrightarrow{x}$ is a fixed point for ϕ.

Claim: ϕ is a contraction.

Let $\overrightarrow{x}, \overrightarrow{y} \in S_\delta$ be arbitrary. Then

$$\begin{aligned} |\phi(\overrightarrow{x})(t) - \phi(\overrightarrow{y})(t)| &= |\int_0^t A(u)[\overrightarrow{x}(u) - \overrightarrow{y}(u)]du| \\ &\leq \int_0^t |A(u)[\overrightarrow{x}(u) - \overrightarrow{y}(u)]|du \\ &\leq \int_0^t ||A(u)||_{op} |\overrightarrow{x}(u) - \overrightarrow{y}(u)|du \\ &\leq \int_0^t ||A(u)||_{op} d(\overrightarrow{x}, \overrightarrow{y})du \\ &= d(\overrightarrow{x}, \overrightarrow{y}) \int_0^t ||A(u)||_{op} du \\ &= cd(\overrightarrow{x}, \overrightarrow{y}) \end{aligned}$$

This is true for any $t \in [-\delta, \delta]$, so we have

$$d(\phi(\overrightarrow{x}), \phi(\overrightarrow{y})) = \max_{t \in [-\delta, \delta]} |\phi(\overrightarrow{x})(t) - \phi(\overrightarrow{y})(t)| \leq cd(\overrightarrow{x}, \overrightarrow{y})$$

So ϕ is contractive on S_δ. Then by the Banach fixed point theorem, since ϕ is contractive on a complete metric space S_δ, so there is some unique $u \in S_\delta$ such that $\phi(u) = u$, which is therefore the unique solution to the differential equation.

Thus we have demonstrated that given a differential equation, there is an isomorphism between $\mathbb{R}^k$, with vectors representing initial conditions, and V, containing solutions to the differential equation. Thus V has dimension k when δ is sufficiently small. □

4.4 Change of Basis

Recall that for any basis $\mathcal{B} = \{v_1, \ldots, v_n\} \subseteq V$, then for any $v \in V$, there exists unique scalar $c_i \in \mathbb{F}$ such that

$$v = c_1 v_1 + \ldots + c_n v_n$$

Definition 4.9

Given a basis $\mathcal{B} = \{v_1, \ldots, v_n\} \subseteq V$, for any $v \in V$, define

$$[v]_\mathcal{B} = M_\mathcal{B}(v) = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{F}^n$$

such that $c_1 v_1 + \ldots + c_n v_n$, where this is well defined because $\mathcal{B}$ is a basis. In other words, $[v]_\mathcal{B}$ satisfies $[v]_\mathcal{B} \cdot \langle v_1, \ldots, v_n \rangle = v$.

Example 4.3

Suppose we consider $V = \mathbb{R}[x]_{\leq 2}$. Define the standard basis $\mathcal{B} = \{1, x, x^2\}$ and another basis $\varepsilon = \{x^2 + 1, x, 2\}$. Then for the polynomial $p(x) = 5 + 3x + 7x^2$, we have

$$M_\mathcal{B}(p) = [p]_\mathcal{B} = \begin{bmatrix} 5 \\ 3 \\ 7 \end{bmatrix}, \; M_\varepsilon(p) = [p]_\varepsilon = \begin{bmatrix} 7 \\ 3 \\ -1 \end{bmatrix}$$

Definition 4.10

Given a vector space $\mathbb{F}^n$, then define the **standard basis** $\varepsilon = \{e_1, \ldots, e_n\}$.

Note that any $v \in \mathbb{F}^n$ then satisfies $M_\varepsilon(v) = v$.

Theorem 4.9

Given some vector space V over $\mathbb{F}$ with $\dim V = n$, and a basis $\mathcal{B}$ of V, the map $\phi : V \to \mathbb{F}^n$ defined by $\phi(v) \coloneqq M_{\mathcal{B}}(v)$ is an isomorphism.

Proof. Linearity is clear, surjectivity follows from closure, and injectivity follows from unique representation in terms of the basis. □

Definition 4.11

Let $T : V \to W$ be a linear transformation. Let $\mathcal{B} = \{v_1, \ldots, v_m\}$ be a basis of V and let $\mathcal{C} = \{w_1, \ldots, w_n\}$ be a basis of W. Then for any v_i, we can uniquely write $T(v_i) = a_{1,i}w_1 + \ldots + a_{n,i}w_n$. Then define the **transformation matrix** of T with respect to the bases $\mathcal{B}, \mathcal{C}$ as

$$M_{\mathcal{B},\mathcal{C}}(T) = [T]_{\mathcal{B}}^{\mathcal{C}} = \begin{bmatrix} a_{11} & \ldots & a_{1m} \\ a_{21} & \ldots & a_{2m} \\ \vdots & & \vdots \\ a_{n1} & \ldots & a_{nm} \end{bmatrix} = \begin{bmatrix} M_{\mathcal{C}}(T(v_1)) & M_{\mathcal{C}}(T(v_2)) & \ldots & M_{\mathcal{C}}(T(v_m)) \end{bmatrix}$$

Example 4.4

Define $\text{proj}_L : \mathbb{R}^2 \to \mathbb{R}^2$ to be projection onto the line $L = \text{span}(\langle 1, 1 \rangle)$. Let . Then under the standard basis, we have

$$M(\text{proj}_L) = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

But under the basis $\mathcal{B} = \{v_1, v_2\} = \{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix}\}$, we have $T(v_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $T(v_2) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Then $[T(v_1)]_{\mathcal{B}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $[T(v_2)]_{\mathcal{B}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. So we have

$$[\text{proj}_L]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} [T(v_1)]_{\mathcal{B}} & [T(v_2)]_{\mathcal{B}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

Now consider $[\text{proj}_L]_{\mathcal{B}}^{\varepsilon}$. Then the matrix is

$$[\text{proj}_L]_{\mathcal{B}}^{\varepsilon} = \begin{bmatrix} [T(v_1)]_{\varepsilon} & [T(v_2)]_{\varepsilon} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

Remark

Notice from the above example that when constructing the matrix $M_{\mathcal{B},\mathcal{C}}(T) = [T]^{\mathcal{C}}_{\mathcal{B}}$, we choose $v_1, \ldots, v_m$ based on the elements of $\mathcal{B}$, then apply T, then convert the resulting vectors into coordinates in $\mathcal{C}$. In other words, the source basis determines which vectors to use, and the target basis determines the coordinates we use.

A key proposition of these matrices in terms of basis is as follows:

Proposition 4.10

Let $T : V \to W$ be linear, with $\mathcal{B}$ a basis of V and $\mathcal{C}$ a basis of W. Then for any $v \in V$, we have

$$[T]^{\mathcal{C}}_{\mathcal{B}}[v]_{\mathcal{B}} = [T(v)]_{\mathcal{C}}$$

If we additionally take $S : W \to \mathcal{U}$, with $\mathcal{D}$ a basis of $\mathcal{D}$, then we have

$$[S \circ T]^{\mathcal{D}}_{\mathcal{B}} = [S]^{\mathcal{D}}_{\mathcal{C}}[T]^{\mathcal{C}}_{\mathcal{B}}$$

Example 4.5

Consider the differentiation operator $D; \mathbb{R}[x]_{\leq 2} \to \mathbb{R}[x]_{\leq 2}$. Let $p = 2 + x + 3x^2$. Then we can calculate $D(p)$ by writing

$$M(Dp) = M(D)M(p)$$

The matrix $M(D)$ is given by

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

And the matrix $M(p)$ is given by

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

So we have

$$M(Dp) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 0 \end{bmatrix} \iff D(p) = 1 + 6x$$

Suppose we have two vector spaces V, W, with ordered bases $\mathcal{B}, \mathcal{C}$, respec-

tively, and a linear transformation $T : V \to W$. Then there is an isomorphism from $V \to \mathbb{F}^n$ induced by $\mathcal{B}$, and an isomorphism from $W \to \mathbb{F}^m$ induced by $\mathcal{C}$. Then these can be related by the following commutative diagram:

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \downarrow{\scriptstyle \cong_{\mathcal{B}}} & & \downarrow{\scriptstyle \cong_{\mathcal{C}}} \\ \mathbb{F}^m & \xrightarrow{L([T]_{\mathcal{B}}^{\mathcal{C}})} & \mathbb{F}^n \end{array}$$

Now note that just as we can encode a transformation as a matrix with respect to a source and target basis, we can also encode the change of basis itself as a matrix.

Definition 4.12

Given a vector space V and two bases $\mathcal{B}, \mathcal{C}$ of V, define the **change of basis** matrix from $\mathcal{B}$ to $\mathcal{C}$ as

$$M_{\mathcal{B}\to\mathcal{C}} \coloneqq M_{\mathcal{B},\mathcal{C}}(\mathrm{Id}_V) = [\mathrm{Id}_V]_{\mathcal{B}}^{\mathcal{C}}$$

Then we have the inverse matrix $M_{\mathcal{C}\to\mathcal{B}}$ as being given by

$$[M_{\mathcal{B}\to\mathcal{C}}]^{-1} = M_{\mathcal{C}\to\mathcal{B}} = M_{\mathcal{C},\mathcal{B}}(\mathrm{Id}_V) = [\mathrm{Id}_V]_{\mathcal{C}}^{\mathcal{B}}$$

The we can observe that for any $T : V \to V$, we immediately have

$$M_{\mathcal{B}}(T) = M_{\mathcal{C}\to\mathcal{B}} M_{\mathcal{C}}(T) M_{\mathcal{B}\to\mathcal{C}}$$

which can be intuitively confirmed by "tracking" the basis through the RHS from right to left. More formally, we can use the previous results to confirm this:

$$M_{\mathcal{C}\to\mathcal{B}} M_{\mathcal{C}}(T) M_{\mathcal{B}\to\mathcal{C}} = [\mathrm{Id}]_{\mathcal{C}}^{\mathcal{B}} [T]_{\mathcal{C}}^{\mathcal{C}} [\mathrm{Id}]_{\mathcal{B}}^{\mathcal{C}} = [\mathrm{Id}T]_{\mathcal{C}}^{\mathcal{B}} [\mathrm{Id}]_{\mathcal{B}}^{\mathcal{C}} = [\mathrm{Id}T\mathrm{Id}]_{\mathcal{B}}^{\mathcal{B}} = [T]_{\mathcal{B}}^{\mathcal{B}} = M_{\mathcal{B}}(T)$$

Then since any choice of bases induces a different isomorphism to $\mathbb{F}^n$, we arrive at the following commutative diagram (where $\phi_{\mathcal{B}}$ is the isomorphism given by $\phi_{\mathcal{B}}(v) = M_{\mathcal{B}}(v)$, and similarly for $\phi_{\mathcal{C}}$).

$$\begin{array}{ccc} \mathbb{F}^n & \xrightarrow{M_{\mathcal{B}}(T)} & \mathbb{F}^n \\ {\scriptstyle \phi_{\mathcal{B}}}\uparrow & & \uparrow{\scriptstyle \phi_{\mathcal{B}}} \\ V & \xrightarrow{T} & V] \\ \downarrow{\scriptstyle \phi_{\mathcal{C}}} & & \downarrow{\scriptstyle \phi_{\mathcal{C}}} \\ \mathbb{F}^n & \xrightarrow{M_{\mathcal{C}}(T)} & \mathbb{F}^n \end{array}$$

4.5 Sums and Direct Sums of Subspaces

Given two subspaces of a vector space, we can consider the subspace formed by sums of the vectors in the original spaces.

Definition 4.13

Given U, W that are subspaces of V, we define the **sum** of U and W to be the smallest subspace of V containing both U and W. That is,

$$U + W \coloneqq \{u + w | u \in U, w \in W\}$$

In the general case, U and W may overlap, and we may then have multiple representations for some vectors in the intersection. Thus it is of particular interest to study spaces such that there is a unique representation of every vector, so that we can translate nice properties about subspaces to their sum.

Definition 4.14

Let U, W be subspaces of V. Then if every vector $x \in U + W$ can be written uniquely in the form $x = u + w$ for some $u \in U$, $w \in W$, then we say that $U + W$ is the **direct sum** of U and W, which we denote $U \oplus W$. We can extend this definition to a direct sum of any number of subspaces.

However, as we alluded to in the previous case, this is very closely linked with the intersection between $U + W$. That is, $U + W$ is a direct sum if and only if U and W are completely "opposed."

Proposition 4.11

Let U, W be subspaces of V. Then $W + U$ is a direct sum if and only if $W \cap U = \{0\}$.

Proof. If $W \cap U \neq \{0\}$, then choose some nonzero $v \in W \cap U$. Then we have $0 = v - v$ but also $0 = 0 + 0$, so 0 has two representations and thus we don't have a direct sum.

If $W \cap U = \{0\}$, then choose some $v \in U + W$ and write it as $v = u + w$ and $v = u' + w'$. Then we have $(u - u') + (w - w') = 0$. Then we have $u - u' = w' - w$, where the left side is in U and the right side in W. Since the only intersection is 0, we must have $u - u' = w' - w = 0$, so $u = u'$ and $w = w'$ and thus we have a unique representation. □

Proposition 4.12

Let V be a finite dimensional vector space and let U be a subspace of V. Then there exists some subspace W of V such that $U \oplus W = V$.

Proof. Choose some basis $u_1, \ldots, u_k$ of U. We can extend this to a basis $u_1, \ldots, u_k, w_1, \ldots, w_l$ of V. Let $W = \text{span}(w_1, \ldots, w_l)$, so that $w_1, \ldots, w_l$ is a basis of W. Then we clearly have $U + W = V$, so we only need to prove that

the intersection is trivial.

Choose some $v \in U \cap W$. Then since $v \in U$, we have $v = c_1u_1 + \ldots + c_ku_k$. We also have $v \in W$, so $v = d_1w_1 + \ldots + d_lw_l$. Then we have

$$c_1u_1 + \ldots + c_ku_k - d_1w_1 - \ldots - d_lw_l = v - v = 0$$

with $u_1, \ldots, u_k, w_1, \ldots, w_l$ linearly independent, so all the c_i, d_i are 0. Thus we have $v = 0$ and the intersection is trivial. So we have exhibited a W such that $U \oplus W = V$. □

Then we also have another way to combine vector spaces:

Definition 4.15

Let V_1, V_2 be vector spaces over $\mathbb{F}$. Then the product is defined as

$$V_1 \times V_2 = \{(v_1, v_2) | v_1 \in V_1, v_2 \in V_2\}$$

Addition and scalar multiplication are defined coordinatewise. For a set of vector spaces $V_1 \ldots, V_n$ over $\mathbb{F}$, then the product is

$$V_1 \times \ldots \times V_n = \{(v_1, \ldots, v_n | v_i \in V_i\}$$

It can be easily verified that $V_1 \times V_2$ as defined here satisfy the axioms for vector spaces, so we find that $V_1 \times V_2$ is a vector space over F. In particular, suppose we consider $\mathbb{R}^2 \times \mathbb{R}^3$. Then we have the set of 2-tuples where the first element is a 2-tuple and the second is a 3-tuple, so it is isomorphic to the set of 5-tuples $\mathbb{R}^5$. This leads to the general observation:

Proposition 4.13

Let $V_1, \ldots, V_m$ be finite dimensional vector spaces over $\mathbb{F}$. Then

$$\dim(V_1 \times \ldots \times V_m) = \dim(V_1) + \dim(V_2) + \ldots + \dim(V_m)$$

Theorem 4.14

Let $V_1, \ldots, V_m$ be subspaces of V, with $V_1 \oplus \ldots \oplus V_m = V$. Then the transformation $\phi : V_1 \times \ldots \times V_m \to V_1 + \ldots + V_m$ given by

$$\phi(v_1, \ldots, v_m) \mapsto v_1 + \ldots + v_m$$

is an isomorphism.

Proof. This transformation is clearly linear. It is surjective, since we can reconstruct any element in $V_1 + \ldots + V_m$ as a sum of elements of the individual spaces, and thus construct the $n$tuple. Since the sum is direct, only one tuple

will send to any given vector in the sum, so ϕ is also injective. So this is an isomorphism. □

Corollary

Let U, W be subspaces of V. Then we have

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

Proof. Consider the transformation $\phi : U \times W \to U + W$ given by $(u, w) \mapsto u + w$. Then this transformation is clearly linear and surjective. It is not injective, however, we only need to find the kernel, which is in fact equal to $\{(u, w) | u + w = 0, u \in U, w \in W\}$.

The kernel, by definition, is $\{(u, w) | u + w = 0\}$. This implies that $u = -w$, which implies that $u, w \in U \cap W$. Then we can define a map from $\ker \phi$ to $U \cap W$, which one can verify is also invertible. So $\ker \phi \cong U \cap W$, so by rank nullity we have

$$\dim(U \times W) = \dim U + \dim W - \dim(U \cap W)$$ □

Lemma

Suppose $V = V_1 \oplus \ldots \oplus V_r$. Suppose $\mathcal{B}_i$ is a basis of V_i for each i. Then the basis given by concatenating the $\mathcal{B}_i$ is a basis of V.

Proof. By associativity, we only need to prove this for the direct sum of two spaces. Suppose $V = U \oplus W$. Then suppose $\mathcal{B}_1 = \{u_1, \ldots, u_k\}$ and $\mathcal{B}_2 = \{w_1, \ldots, w_l\}$. We want to show that $\mathcal{B} = \{u_1, \ldots, u_k, w_1, \ldots, w_l\}$ is a basis of V.

To show that it spans V, pick some $v \in V$. Then $v = u + w$ for some $u \in U$, $w \in W$. Suppose $u = a_1 u_1 + \ldots + a_k u_k$ and $w = b_1 w_1 + \ldots + b_l w_l$. Then

$$v = a_1 u_1 + \ldots + a_k u_k + b_1 w_1 + \ldots + b_l w_l$$

so $\mathcal{B}$ spans V.

To show linear independence, suppose $c_1 u_1 + \ldots + c_k u_k + d_1 w_1 + \ldots + d_l w_l = 0$. 0 can also be represented as $0u_1 + \ldots + 0w_l$, and by unique representations we have that all the c_i, d_i are 0. So they are linearly independent. □

4.6 Quotient Spaces

Suppose we have a finite dimensional vector space V, and we have a subspace U. Then we can imagine translating U within V, by some offset vector $v \in V$. While this is not a subspace in general, it does contain similar (affine) structure to U.

Definition 4.16

Given a finite dimensional vector space V, a subspace U of V, and a vector $v \in V$, then define the **translate** of U by v to be

$$v + U \coloneqq \{v + u | u \in U\}$$

One important note is that translating a given subspace by multiple vectors may result in the same set. For instance, in $\mathbb{R}^3$, if we translate the xy-plane by $\langle 0, 0, 1\rangle$, or by $\langle 1, 0, 1\rangle$, we will have the same set. However, if we call two vectors v_1, v_2 equivalent when $v_1 + U = v_2 + U$, then we can see that this defines an equivalence relation on V, and thus U partitions V. Moreover, this partition has the structure of a vector space, which we call the quotient space:

Definition 4.17

Let V be finite dimensional and let U be a subspace. Define the **quotient space** of V by U to be

$$V/U \coloneqq \{v + U | v \in V\}$$

We define addition and scalar multiplication on V/U as follows:

$$(v_1 + U) +_q (v_2 + U) = (v_1 +_V v_2) + U, \quad v_1, v_2 \in V$$
$$\lambda *_q (v + U) = (\lambda *_V v) + U, \quad \lambda \in \mathbb{F}, v \in V$$

We can verify that the operations defined here are well defined; that is, if different representatives for the same translates are chosen, then the outputs are the same. Moreover, the operations as chosen also satisfy all the axioms for a vector space. Thus, the set we have constructed is indeed a vector space. It should be noted that V/U is not actually a subspace of V. While it is isomorphic to subspaces of V, that choice is not canonical.

Since each element of the quotient space corresponds to one of the partitions of V, we can naturally define a map which indicates which partition a given element is in.

Definition 4.18

Let V be a finite dimensional vector space and U a subspace. Then the **canonical map** or **quotient map** from V to V/U is the map $\pi : V \to V/U$ such that $\pi(v) = v + U$.

One can clearly see that this map is linear and surjective.

Proposition 4.15

$\dim V/U = \dim V - \dim U$.

Proof. Consider $\pi : V \to V/U$. By rank nullity, $\dim V = \dim \operatorname{im} \pi + \dim \ker \pi$. π is surjective so $\dim \operatorname{im} \pi = \dim V/U$, and $\ker \pi = U$, so $\dim V = \dim V/U + \dim U$ and thus $\dim V/U = \dim V - \dim U$. □

We can also prove this without using rank nullity, which will then allow us to prove rank nullity separately.

Alternate Proof. We know that there exists a subspace W such that $V = U \oplus W$. We know that $\dim W = \dim V - \dim U$ Then if $i_W : W \to V$ is the inclusion map, then $\pi \circ i_W : W \to V \to V/U$ is an isomorphism, so $\dim V/U = \dim W = \dim V - \dim U$. □

Thus we have found an identification from subspaces to linear maps. Similarly, we can also find an identification from linear maps to subspaces. This identification is a fundamental result which demonstrates the relationship between vector space structure and structure preserving mappings (i.e. linear maps).

Theorem 4.16: First Isomorphism Theorem

Let $T : V \to W$ be a linear map, and let $U = \ker T$. Define $\tilde{T} : V/U \to W$ with $\tilde{T}(v + U) = T(v)$. Then $\tilde{T}$ is a well-defined, linear injective map.

In other words, by quotienting out the kernel elements, we remove all the redundancy in T, which forces it to become injective.

The rank-nullity theorem is a consequence of the first isomorphism theorem. If we have some $T : V \to \operatorname{im} T$, then this is obviously surjective, and the first isomorphism theorem says that $\tilde{T} : V/\ker T \to \operatorname{im} T$ is also injective and thus $\dim \operatorname{im} T = \dim V/\ker T = \dim V - \dim \ker T$.

Chapter 5

Determinants

5.1 Invariants

Encoding a linear map as a matrix allows us to extract algebraic information from the matrix elements. However, since matrix representations are basis dependent, this is of little interest to us in general. Where this becomes important is when we are able to identify quantities which are invariant under a change of basis. That is, if $I(\cdot)$ is a function which retrieves some algebraic invariant of a matrix, then we should have $I(M_{\mathcal{B}}(T)) = I(M_{\mathcal{C}}(T))$ for any bases $\mathcal{B}, \mathcal{C}$ and any transformation T.

5.2 Laplace Expansion

The determinant is an invariant which is defined only for square matrices (that is, matrices which encode a endomorphic transformation $T : V \to V$). The determinant has a natural geometric interpretation as the scaling factor of the transformation. That is, if we consider the unit n-cube of $\mathbb{R}^n$, then the n-volume of that cube after applying T is $|\det T|$. Moreover, the sign of $\det T$ tells us whether the handedness of the space has changed; or whether space has been "flipped."

Here we will define a process called *Laplace expansion* which will allow us to calculate the determinant of an $n \times n$ matrix. However, this calculation should not be taken as the definition of the determinant, but simply as one method for calculating it that will also allow for convenient proofs of properties of the determinant.

For a 2×2 matrix, we simply take for granted the following formula:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

For a 3×3 and larger matrix, we will define this recursively. We can expand along any row or column of the matrix. In this discussion, we will assume that

we are expanding along a row for convenience; the notation is analogous for a column.

Let A be our original matrix. We denote by A_{ij} the matrix obtained by deleting the ith row and jth column of A. Then suppose we expand along the ith row. We define the Laplace expansion along the ith row to be

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

By recursively expanding along the first column, we can see that for any upper triangular matrix, the determinant will simply be the product of the diagonal entries:

$$\deg\begin{bmatrix} a_1 & * & * & * \\ 0 & a_2 & * & * \\ 0 & 0 & \ddots & * \\ 0 & 0 & 0 & a_n \end{bmatrix}$$

Similarly, we can see that any matrix with a row or columns of all zeroes has determinant zero.

Since the determinant is invariant under change of basis, we would expect many natural properties to follow. For instance, we have

Proposition 5.1

$\det(AB) = \det(A)\det(B)$ for any $n \times n$ matrices A, B.

Corollary

A is invertible if and only if $\det A \neq 0$, and in that case $\det(A^{-1}) = 1/\det A$.

So far we have used the fact that det is invariant under a change of basis. We will formalize this notion as follows:

Definition 5.1

Two matrices A and B are **similar** if there exists some invertible matrix S such that $A = S^{-1}BS$.

Then we have the following:

Proposition 5.2

If A, B are similar, then $\det A = \det B$.

Proof. Since $A = S^{-1}BS$, we have $\det A = \det S^{-1} \det B \det S = 1/\det S \det B \det S = \det B$. □

The entire point of our search for invariants was to be able to algebraically extract information about linear transformations, not just matrices. Specifically, we define the following:

Definition 5.2

Given $T : V \to V$, where V is finite dimensional, define the **determinant** of T to be $\det T \coloneqq \det M_{\mathcal{B}}(T)$.

By the change of basis formula, any choice of basis here will lead to the same answer.

5.3 Multilinearity

We now investigate the question of linearity for det. By a simple inspection, we can see that det is certainly not linear, since it fails under scalar multiplication for 2×2 matrices:

$$\det(2\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \det \begin{bmatrix} 2a & 2b \\ 2c & 2d \end{bmatrix} = 2a2d - 2b2c = 4 \det \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

We can perform a similar calculation with a 3×3 and get the following:

$$\det(2\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}) = 8 \det \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$$

This leads us to guess that $\det \lambda A = \lambda^n \det A$ for any $n \times n$ matrix and any scalar. In fact, this is true, and our search of a proof for this will lead us to a deeper understanding of the nature of det.

Suppose our initial matrix has columns as follows:

$$A = \begin{bmatrix} | & | & | \\ a_1 & \dots & a_n \\ | & | & | \end{bmatrix}$$

Then if we multiply any singular row by a scalar, leaving the others unchanged, we have:

$$A' = \begin{bmatrix} | & | & | \\ ca_1 & \dots & a_n \\ | & | & | \end{bmatrix}$$

By expanding along the row which has been modified, we see that

$$\det A' = c \det A$$

Similarly, if we add any other column vector $b \in \mathbb{F}^n$, so that

$$A'' = \begin{bmatrix} | & | & | \\ a_1 + b & \dots & a_n \\ | & | & | \end{bmatrix}$$

then we have

$$\det A'' = \det A + \det \begin{bmatrix} | & | & | \\ b & \dots & a_n \\ | & | & | \end{bmatrix}$$

So although det is not linear, it is linear in each column when modified individually. Since we could have expanded along either rows or columns, the same would apply for modifying any column. We call this property **multilinearity**.

Using multilinearity, the property that we guessed earlier is easy to prove:

Proposition 5.3

For any square matrix $A \in M_{n\times n}(\mathbb{F})$ and any scalar $\lambda \in \mathbb{F}$,

$$\det \lambda A = \lambda^n \det A$$

Proof. We can modify one column at a time by multiplying by λ. Each time, we will scale the determinant by λ, and since we need to do this once for each column, the total scaling factor is λ^n. □

Scalar multiplication and addition of columns or rows are of particular note because they are two of the elementary row operations that we use for row reduction and comparison of row-equivalent matrices. We should similarly investigate the last operation, switching rows.

Proposition 5.4

If two columns or rows in a matrix are the same, then the determinant is 0.

Proof. If we have a matrix of the form

$$A = \begin{bmatrix} | & | & | & | & | \\ \dots & a & \dots & a & \dots \\ | & | & | & | & | \end{bmatrix}$$

Then the columns are not linearly independent, so the transformation $L_A : \mathbb{F}^n \to \mathbb{F}^n$ is not injective or surjective. Thus $\det(A) = 0$. □

Corollary

If a matrix A can be obtained from B via switching two columns, then $\det A = -\det B$. That is,

$$\det \begin{bmatrix} | & | & | & | & | \\ \dots & a & \dots & b & \dots \\ | & | & | & | & | \end{bmatrix} = -\det \begin{bmatrix} | & | & | & | & | \\ \dots & b & \dots & a & \dots \\ | & | & | & | & | \end{bmatrix}$$

Proof. Using multilinearity of the determinant and the proposition above, we have

$$\begin{aligned} \det \begin{bmatrix} | & | \\ a & b \\ | & | \end{bmatrix} &= \det \begin{bmatrix} | & | \\ a & a \\ | & | \end{bmatrix} + \det \begin{bmatrix} | & | \\ a & b \\ | & | \end{bmatrix} \\ &= \det \begin{bmatrix} | & | \\ a & a+b \\ | & | \end{bmatrix} \\ &= \det \begin{bmatrix} | & | \\ -b & -b \\ | & | \end{bmatrix} + \det \begin{bmatrix} | & | \\ a & a+b \\ | & | \end{bmatrix} \\ &= \det \end{aligned}$$

□

As a result, we can now quantify the determinant of A after performing the row reduction process.

Theorem 5.5

Let A be a matrix and suppose that in the Gaussian reduction process, rows are switched s times and rows are divided by the quantities $k_1, \dots, k_r$. Then if A is invertible,

$$\det(A) = (-1)^s k_1 \dots k_r$$

and if it is not invertible

$$\det(A) = 0$$

5.4 Trace

We will now examine a second invariant of matrices which comes up in functional analysis, called the *trace*.

Definition 5.3

Let A be a square $n \times n$ matrix. Then the **trace** of A, $\text{tr}(A)$ is defined as the sum of the entries on the diagonal. That is,

$$\text{tr}(A) = \sum_{i=1}^{n} a_{ii}$$

Then by using variables for matrix entries, we can see that

Proposition 5.6

$\text{tr} : M_{n\times n}(\mathbb{F}) \to \mathbb{F}$ is linear.

We also have the following:

Proposition 5.7

For any square $n \times n$ matrices A, B, we have

$$\text{tr}(AB) = \text{tr}(BA)$$

Then we can show that the trace is indeed invariant under similarity:

Theorem 5.8

Suppose A and B are similar. Then $\text{tr}(A) = \text{tr}(B)$.

Proof. Since $B = S^{-1}AS$, we have

$$\text{tr}(B) = \text{tr}((S^{-1}A)S) = \text{tr}(SS^{-1}A) = \text{tr}(A) \qquad \square$$

This allows us to define the trace of a linear transformation just as we did for a determinant:

Definition 5.4

Let V be finite dimensional and let $T : V \to V$. Then the trace of V is T is defined to be

$$\text{tr}(T) \coloneqq \text{tr}(M_{\mathcal{B}}(T))$$

where $\mathcal{B}$ is any basis of V.

Then the preceding theorem shows that this is a well-defined value.

Chapter 6

Eigenvalues, Eigenvectors, and Diagonalization

6.1 Eigenvalues and Eigenvectors

We now turn our attention to more important information that we can extract about a transformation or a matrix. In particular, we will study eigenvalues and eigenvectors, both of which are strongly connected with *invariant subspaces*. Invariant subspaces are useful to us because, intuitively speaking, they don't interact under a transformation. Thus, we can analyze the action of a transformation on a space by decomposing the space into invariant subspaces and separately analyzing the transformation on each of the subspaces.

Definition 6.1

Let $T : V \to V$ be a transformation. Then we call a subspace $U \subseteq V$ *T-invariant* if $T(U) \subseteq U$. That is, we have $T(u) \in U$ for every $u \in U$.

Note that this definition does not necessarily mean that a subspace remains the same, but merely that it cannot be taken out of the subspace under T. That is, T may make the subspace smaller than itself. For instance, we have:

Proposition 6.1

Let $T : V \to V$. Then $\ker T$ and $\operatorname{im} T$ are both T-invariant subspaces of V.

Proof. We have $T(\ker T) = \{0\}$ by definition, and $\{0\} \subseteq \ker T$, so $\ker T$ is T-invariant.

$T(\operatorname{im} T) \subseteq T(V) = \operatorname{im} T$ so $\operatorname{im} T$ is T-invariant. □

We will be particularly interested in 1-dimensional invariant subspaces in this class, since those are related to the concepts of eigenvalues and eigenvectors.

Definition 6.2

We say that a scalar $\lambda \in \mathbb{F}$ is an **eigenvalue** of T if there exists some nonzero vector $v \in V$ such that $T(v) = \lambda v$.

Then $v \in V$ is an **eigenvector** of T corresponding to an eigenvector λ if $v \neq 0$ and $T(v) = \lambda v$.

Example 6.1

Suppose we have the matrix

$$\begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

Then we have eigenvalues of 5 and 3. The eigenvectors corresponding to 5 are the x-axis (excluding the origin) and the eigenvectors corresponding to 3 are the y-axis (excluding the origin).

Recall that since we are dealing with transformations of the form $T : V \to V$ where V is finite dimensional, we can make use of the useful fact that surjectivity, injectivity, and bijectivity are all equivalent.

Proposition 6.2

Let $T : V \to V$, where V is finite dimensional. Then $v \in V$ is an eigenvector of eigenvalue λ if and only if $v \in \ker(T - \lambda I)$, where I is the identity transformation. Moreover, $\lambda \in \mathbb{F}$ is an eigenvalue of T if and only if $\det(M(T) - \lambda I) = 0$.

Proof. For the first fact, note that

$$T(v) = \lambda v \iff (T - \lambda I)(v) = 0 \iff v \in \ker(T - \lambda I)$$

For the second, we know that λ is an eigenvalue of T if and only if there exists some nonzero $v \in V$ such that $T(v) = \lambda v$. Now this happens if and only if

$$v \in \ker(T - \lambda I)$$

which means that $T - \lambda I$ is not injective and thus $\det(T - \lambda I) = 0$. □

Definition 6.3

Let $T : V \to V$ with V finite dimensional and let λ be an eigenvalue of T. Then define the **eigenspace** of λ in $\mathbb{F}$ to be

$$E_{\lambda,T} \coloneqq \{0\} \cup \{v \in V : v \text{ is an eigenvector of eigenvalue } \lambda\} = \ker(T - \lambda I)$$

We should note that eigenvalues may be repeated for different 1-dimensional invariant subspaces. That is, there may be unrelated eigenvectors which have the same eigenvalue, and thus a single eigenspace may encapsulate multiple distinct invariant subspaces.

Definition 6.4

We define the **geometric multiplicity** of an eigenvalue λ to be $\operatorname{gemu} \lambda = \dim E_{\lambda,T}$.

Lastly, we will define the characteristic polynomial, which is another invariant that captures lots of important information about a transformation.

Definition 6.5

Given a transformation $T : V \to V$ finite dimensional, we define the **characteristic polynomial** of T to be $P_T = \det(T - \lambda I)$, with λ being taken as the variable.

Remark

An observant reader will note that the determinant of $T - \lambda I$ is not a value in the field, since the entries are of the form $a - \lambda$, where λ is taken to be a variable. However, these entries are elements of $K(\lambda)$, which is still a field, and thus the same results hold.

Then the discussion above easily shows that the roots of the characteristic polynomial are precisely the eigenvalues of the transformation.

We can similarly define these properties for a matrix. Specifically, given a matrix A, we can perform the following operations to find its eigenvalues and eigenvectors:

- Calculate the characteristic polynomial $\det(A - \lambda I)$. The roots are precisely the eigenvalues of A.
- Calculate the eigenspace $E_{\lambda,A} = \ker(A - \lambda I)$ for each eigenvalue λ. This allows us to find eigenvectors for each eigenvalue.

Theorem 6.3

Let $T : V \to V$, with V finite dimensional. Let $v_1, \ldots, v_m$ be eigenvectors for V, with corresponding eigenvalues $\lambda_1, \ldots, \lambda_m$. Suppose all the eigenvalues are distinct. Then $v_1, \ldots, v_m$ are linearly independent.

In other words, eigenvectors corresponding to distinct eigenvalues are linearly independent. This is a very intuitive result, since we can use the eigenvalues to decompose into distinct eigenspaces. These eigenspaces don't necessarily sum to the entire space, but they are certainly disjoint (besides 0). This makes sense, since the action of the transformation on each eigenspace is different, and thus they must be different subspaces. For a formal proof, we can use induction.

Proof. We induct on m. For the case $m = 1$, we trivially have linearly independent, since v_1 is nonzero and thus linearly independent. Now assume that the first k vectors are linearly independent. Then add the $k+1$th vector and create a linear relation:

$$c_1 v_1 + \ldots + c_k v_k + c_{k+1} v_{k+1} = 0$$

Then we can apply T:

$$T(c_1 v_1 + \ldots + c_k v_k + c_{k+1} v_{k+1}) = 0$$

and since they are eigenvectors:

$$c_1 \lambda_1 v_1 + \ldots + c_k \lambda_k v_k + c_{k+1} \lambda_{k+1} v_{k+1}$$

Then we can multiply the first equation by λ_{k+1} and subtract from the second to get

$$0 = (c_1 \lambda_1 - c_1 \lambda_{k+1}) v_1 + \ldots + (c_k \lambda_k - c_k \lambda_{k+1}) v_k$$

Since the first k vectors are linearly independent, we must have $c_i(\lambda_i - \lambda_{k+1}) = 0$ for each i. Since the coefficients are distinct, we must have $c_i = 0$, and thus the original coefficients were all 0. So $v_1, \ldots, v_{k+1}$ are linearly independent. □

As a corollary, we are able to formalize the intuitive argument made in the paragraph before this proof:

Corollary

Let $T : V \to V$ with V finite dimensional. Let $\lambda_1, \ldots, \lambda_m$ be distinct eigenvalues. Then the sum $E_{\lambda_1} + \ldots + E_{\lambda_m}$ is a direct sum; that is, $E_{\lambda_1} + \ldots + E_{\lambda_m} = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_m}$.

Proof. Choose some $v \in E_{\lambda_1} + \ldots + E_{\lambda_m}$. Suppose we have two representations

$$v_1 + \ldots + v_m = v = v_1' + \ldots + v_m'$$

such that $v_i, v_i' \in E_{\lambda_i}$. Then

$$0 = (v_1 - v_1') + \ldots + (v_m - v_m')$$

Since these are linearly independent by the theorem, the $(v_i - v_i')$ must all be zero, so the representations are the same. □

6.2 Diagonalization

We have observed that eigenvectors are of particular importance because they define 1-dimensional subspaces of V which are invariant under a transformation. We now turn our attention to the problem of *diagonalization*. This problem asks us, given some matrix, to find a similar matrix which is diagonal. This is a powerful tool because it allows us to represent the transformation very simply, as scaling in certain directions and nothing more. Then we need to decide when such a matrix can be found.

Definition 6.6

Given a transformation $T : V \to V$ with V finite dimensional, n vectors $v_1, \ldots, v_n$ is an **eigenbasis** of V for T if it is a basis, and each v_i is eigenvector of T.

Definition 6.7

A transformation $T : V \to V$ is **diagonalizable** if there exists some basis $\mathcal{B} \subseteq V$ such that $M_{\mathcal{B}}(T)$ is diagonal.

In particular, we have:

Proposition 6.4

T is diagonalizable if and only if there exists an eigenbasis of V for T.

Proof. ($\Longrightarrow$) Suppose we have some basis $\mathcal{B} \subseteq V$ such that $M_{\mathcal{B}}(T)$ is diagonal. Then

$$M_{\mathcal{B}}(T(v_i)) = M_{\mathcal{B}}(T)M_{\mathcal{B}}(v_i) = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{bmatrix} \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ a_i \\ \vdots \\ 0 \end{bmatrix} = a_i v_i$$

so each v_i is an eigenvalue and thus we have an eigenbasis.

($\Longleftarrow$) Suppose we have an eigenbasis $\mathcal{B}$. The central idea here is that when we

have an eigenbasis, each coordinate is acted on separately. Thus, if $v_i \in \mathcal{B}$ has eigenvalue λ_i, then $T(v_i) = \lambda_i v_i$. If we have an eigenbasis, this holds for any i and thus

$$M_{\mathcal{B}}(T) = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

so T is diagonalizable. □

From the proof we see that not only is a transformation diagonalizable if and only if it has an eigenbasis, but also the diagonal matrix is precisely one with eigenvalues along the diagonal.

Similarly, we can ask about the diagonalization of a matrix.

Definition 6.8

A **diagonalization** of an $n \times n$ matrix A is a presentation

$$D = S^{-1}AS$$

where D is diagonal and S is invertible.

This question is equivalent to the diagonalization problem for transformations:

Proposition 6.5

A matrix A is diagonalizable if and only if there is some basis $\mathcal{B} \subseteq \mathbb{F}^n$ such that $M_{\mathcal{B}}(L_A)$ is diagonal.

Proof. ($\implies$) If A is diagonalizable, then we have $D = S^{-1}AS$ for some S invertible and D diagonal. Then $A = M_e(L_A)$. Choose $\mathcal{B}$ such that $v_i \in \mathcal{B}$ is the ith column of S. Since S is invertible, it follows that the v_i are a basis. Then $S = M_{\mathcal{B}\to e}$ and similarly $S^{-1} = M_{e\to\mathcal{B}}$. By change of basis, then we have $D = M_{e\to\mathcal{B}}M_e(L_A)M_{\mathcal{B}\to e} = M_{\mathcal{B}}(L_A)$.

($\impliedby$) If there exists a basis $\mathcal{B}$ such that $M_{\mathcal{B}}(L_A)$ is diagonal. We have $A = M_e(L_A)$, and we pick S to be $M_{\mathcal{B}\to e}$. Then by change of basis this choice works, since $M_{\mathcal{B}}(L_A) = M_{e\to\mathcal{B}}M_e(L_A)M_{\mathcal{B}\to e}$. □

In particular, the procedure for calculating the diagonalization of a matrix is to calculate an eigenbasis $\mathcal{B}$ of $\mathbb{F}^n$ for T_A. Then we have $S = M_{\mathcal{B}\to e}$, which is the matrix with column i equal to $v_i \in \mathcal{B}$. By change of basis, we will then have D as the diagonal matrix with eigenvalues on the diagonal.

One particular advantage is that because a diagonalized matrix acts on each vector in the eigenbasis independently, it is very easy to calculate repeated

applications of the transformation. This manifests itself as cancellation of the change of basis matrices S and S^{-1}.

Proposition 6.6

If A is diagonalizable with $A = SDS^{-1}$, then $A^n = SD^nS^{-1}$.

Now that we have seen some of the power of diagonalization, we turn our attention to the question of which transformations and matrices can be diagonalized.

Proposition 6.7

If T has n distinct eigenvalues, then T is diagonalizable.

Proof. We can pick n eigenvectors, each corresponding to one of the eigenvalues. These are linearly independent since they correspond to distinct eigenvalues. Since we have n linearly independent vectors, it is a basis, so this is an eigenbasis and thus T is diagonalizable. □

A similar proposition holds for matrices. Not that the converse is certainly not true, since we may have repeated eigenvalues (consider the identity).

Proposition 6.8

Suppose there exists an eigenbasis of V for T, and T has distinct eigenvalues $\lambda_1, \ldots, \lambda_m$. Then $V = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_m}$.

Proof. ($\implies$) Suppose $v_1, \ldots, v_n$ are an eigenbasis of V for T. Let i_j be the number of eigenvectors in the eigenbasis of eigenvalues λ_j. Then we clearly have $\sum_j i_j = n$. We must have $i_j \leq \dim E_{\lambda_j}$ for each j, because the v_i are linearly independent. This gives us

$$n = \sum_j i_j \leq \sum_j \dim E_{\lambda_j}$$

Since we know that the sum of the eigenspace is direct (though not necessarily equal to n), we have

$$\dim \bigoplus_j E_{\lambda_j} = \sum_j \dim E_{\lambda_j}$$

Moreover $\bigoplus E_{\lambda_j} \subseteq V$ so

$$n = \dim V \geq \dim \bigoplus_j E_{\lambda_j} = \sum_j \dim E_{\lambda_j} \geq \sum_j i_j = n$$

So the only way this holds is if the inequality is in fact an equality, proving the direct sum.

($\Longleftarrow$) Suppose $V = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_m}$. Since this is a direct sum, we can pick a basis of each and concatenate them to find a basis of V. But each basis vector is an eigenspace and thus an eigenvector, so we have an eigenbasis. $\square$

To summarize the results on diagonalizability so far, we have the following:

Theorem 6.9

Let V be a finite dimensional vector space over $\mathbb{F}$. Let $T : V \to V$ be a transformation. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of T with repetition. Then the following are equivalent:

1. There exists an eigenbasis of V for T.
2. $M_{\mathcal{B}}(T)$ is diagonal for some basis $\mathcal{B}$.
3. $V = \bigoplus E_{\lambda_i}$.
4. $\dim V = \sum \dim E_{\lambda_i} = \sum \text{gemu}(\lambda_i)$.

6.3 Permutations ($*$)

We will briefly take a detour here to discuss applications of permutations to linear algebra. Doing so will allow us to derive *Leibniz's formula* for the determinant, which is phrased in terms of permutations.

Definition 6.9

A **permutation** of a set X is a bijection $\pi : X \to X$. In particular, we consider permutations of the set $\{1, \ldots, n\}$, and we will represent such a permutation as $i_1 i_2 \ldots i_n$, where $i_j = \pi(j)$.

For instance, the identity permutation on 3 elements is 123. Other permutations would be 213, 231, and so on. In particular, we consider *swaps*, which are permutations that switch exactly two elements.

Definition 6.10

A permutation is **even** if there exists a decomposition into an even number of swaps $\tau_1, \ldots, \tau_{2k}$ such that $\pi = \tau_{2k} \circ \ldots \circ \tau_1$. A permutation is **odd** if there exists an odd number of swaps. The **sign** of a permutation is $\text{sgn}(\pi)$, which is defined to be 1 if π is even and -1 if π is odd.

Theorem 6.10: Leibniz's Formula

Let A be a matrix. Then

$$\det(A) = \sum \operatorname{sgn}(i_1 \dots i_n) a_{1i_1} \dots a_{ni_n}$$

where the sum is over all possible permutations of n elements.

Proof. Let us first consider the 2×2 case. The only permutations are 12 and 21, so

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc = \operatorname{sgn}(12) a_{11} a_{22} + \operatorname{sgn}(21) a_{12} a_{21}$$

Though we will not provide a full proof here, the strategy is to apply induction using Laplace expansion. □

In other words, each term is constructed by picking one element from each row, such that each column is only represented once.

6.4 Further Study of Eigenvalues and Eigenvectors

We will now continue to study diagonalization and related problems.

Definition 6.11

Given an eigenvalue λ of T, the **algebraic multiplicity** of λ is $\operatorname{almu} \lambda$, which is the multiplicity of λ as a root of the characteristic polynomial $P_T(x)$.

We should note that this is inequivalent to the geometric multiplicity in general.

Example 6.2

Consider the matrix $\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$. Since this is a skew matrix, the only eigenspace is the x axis and thus $\operatorname{gemu} 1 = 1$. But the characteristic polynomial is $P_A(\lambda) = (1 - \lambda)^2$, and thus $\operatorname{almu} 1 = 2$.

As we just saw, the algebraic multiplicity may contain more multiplicities than the dimension of the eigenspace. To prove that this only works in one direction, we use the following lemma:

Lemma

Suppose $A \in M_{m\times m}$, $B \in M_{m\times n}$, and $C \in M_{n\times n}$. Then the determinant of the block matrix is

$$\det \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} = \det A \det C$$

We have the following:

Proposition 6.11

For any eigenvalue λ, $\operatorname{gemu} \lambda \leq \operatorname{almu} \lambda$.

Proof. Suppose $m = \operatorname{gemu} \lambda = \dim E_\lambda$. Choose some basis $v_1, \ldots, v_m$ of E_λ. Then extend this to a basis of V, given by $\mathcal{B} = \{v_1, \ldots, v_m, \ldots, v_n\}$. So we have

$$M_{\mathcal{B}}(T) = \begin{bmatrix} \lambda & & & & * & \ldots \\ & \lambda & & & * & \ldots \\ & & \ddots & & * & \ldots \\ & & & \lambda & * & \ldots \\ & & & & * & \ldots \end{bmatrix}$$

As a block matrix, this is

$$M_{\mathcal{B}}(T) = \begin{bmatrix} \lambda I_m & B \\ 0 & C \end{bmatrix}$$

for some B, C. Then

$$P_T(x) = \det(M_\beta(T) - xI_{n+m}) = \det \begin{bmatrix} \lambda I_m - xI_m & B \\ 0 & C - xI_n \end{bmatrix} = (\lambda - x)^m P_C(x)$$

Thus $\operatorname{almu} \lambda \geq m$. □

We note here that in $\mathbb{C}$, every characteristic polynomial will decompose into linear factors, and thus every matrix (over either $\mathbb{R}$ or $\mathbb{C}$) has exactly n complex eigenvalues, counting algebraic multiplicity. So $n = \sum \operatorname{almu} \lambda \geq \sum \operatorname{gemu} \lambda$. For diagonalizable real matrices, $\sum \operatorname{gemu} \lambda = n$, so we have $\operatorname{gemu} \lambda = \operatorname{almu} \lambda$ for each λ.

This means that over the complex numbers, every matrix has at least one (possibly complex) eigenvalue.

Proposition 6.12

For a finite dimensional vector space V over $\mathbb{R}$, if $\dim V = n$ is odd, then every linear transformation $T : V \to V$ has a real eigenvalue.

Proof. If n is odd, then $\deg(p_T(x)) = n$ is also odd. Thus, the limits at infinity are different, with $\lim_{x\to\infty} p = \pm\infty$ and $\lim_{x\to-\infty} p = \mp\infty$. By the intermediate value theorem, there exists some real λ such that $p(\lambda) = 0$. □

We can also prove this using facts about polynomials. Since V is a real vector space, $p_T(x)$ has real coefficients. Thus for any nonreal root λ of p_T, $\overline{\lambda}$ is also a root of p_T. thus, the nonreal complex roots (and thus eigenvalues) come in conjugate pairs. Thus we can decompose every real polynomial into a product of degree 1 and degree 2 real polynomials. Since we have n odd, at least one must be of degree 1, which gives a real polynomial.

Proposition 6.13

Let $T : V \to V$ with V finite dimensional. Then the coefficient of the first, second, and last terms of the characteristic polynomial are $p_T(x) = (-1)^n x^n + (-1)^{n-1}\operatorname{tr}(T)x^{n-1} + \ldots + \det(T)$.

Proof. Let A be the matrix of T in some basis. To construct the characteristic polynomial, we have

$$p_T(x) = \det\begin{bmatrix} a_{11}-x & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22}-x & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn}-x \end{bmatrix}$$

We will evaluate this using Leibniz's rule. In particular, the term for the product along the diagonal is

$$(a_{11}-x)\ldots(a_{nn}-x)$$

where sgn of this permutation is 1. Certainly, this term has degree n. For any other term, there is at least one off-diagonal entry. Moreover, since each row or column can only be represented by one entry. Thus, the other terms must have degree at most $n-2$. Thus, the coefficients of x^n and x^{n-1} in $p_T(x)$ are *precisely* the coefficients in $(a_{11}-x)\ldots(a_{nn}-x)$.

Calculating, we can see that the x^n coefficient is simply $(-1)^n$. Moreover, if we expand, the terms with coefficients x^{n-1} are of the form $a_{ii}(-1)^{n-1}x^{n-1}$. Summing over all possible choices, the x^{n-1} coefficient is $(-1)^{n-1}(a_{11}+\ldots+a_{nn})x^{n-1} = (-1)^{n-1}\operatorname{tr}(A)x^{n-1} = (-1)^{n-1}\operatorname{tr}(T)x^{n-1}$.

Lastly, note that the last term is given by $p(0) = \det(A-0I) = \det(A) = \det(T)$. □

Corollary

For any 2×2 matrix A, the characteristic polynomial is given by

$$p_a(x) = x^2 - \operatorname{tr}(A)x + \det(A)$$

The upshot of this is that we get some interesting facts about the trace and determinant:

Corollary

For any real or complex matrix A with eigenvalues $\lambda_1, \ldots, \lambda_n$, repeated by algebraic multiplicity, we have:

- $\operatorname{tr}(A) = \sum_i \lambda_i$
- $\det(A) = \prod_i \lambda_i$

Proof. We have $p_A(x) = (-1)^n(x - \lambda_1)\ldots(x - \lambda_n)$. So the second term is $(-1)^n(-1)(\lambda_1 + \ldots + \lambda_n)x^{n-1} = (-1)^{n-1}(\lambda_1 + \ldots + \lambda_n)x^{n-1}$. But we also know the characteristic polynomial has second term given by $(-1)^{n-1}\operatorname{tr}(A)x^{n-1}$.

Similarly, the last term is $(-1)^n(-1)^n\lambda_1 \ldots \lambda_n = \lambda_1 \ldots \lambda_n = \det(A)$. □

For diagonalizable matrices, this can be shown by diagonalizing A into a diagonal matrix with the λ_i along the diagonal. However, not every matrix can be diagonalized, even with n complex eigenvalues, since here we are counting by algebraic multiplicity and not geometric multiplicity.

Example 6.3

Find all eigenvalues of the $n \times n$ matrix A, which has all entries as 1.

We have $E_0 = \ker A$. Since we only have one linearly independent column, we find that $\operatorname{rank} A = n$ and thus $\ker A = n - 1$. Thus $\operatorname{almu} 0 \geq \operatorname{gemu} 0 = n - 1$, so the trace is given by $\underbrace{0 + \ldots + 0}_{n-1} + \lambda$, where λ is unknown. But $\operatorname{tr} A = n$, so $\lambda = n$. Thus the only eigenvalues are 0, with $\operatorname{almu} 0 = \operatorname{gemu} 0 = n - 1$ and n, with $\operatorname{almu} n = \operatorname{gemu} n = 1$.

6.5 Minimal Polynomials

We will now consider other ways of working with polynomials to represent matrices. First, we need to discuss how to evaluate polynomials using operators.

Definition 6.12

Let $p(x) \in \mathbb{F}[x]$ be a polynomial given by

$$p(x) = c_0 + c_1 x + \ldots + c_m x^m, c_i \in \mathbb{F}$$

Let A be an $\mathbb{F}$-matrix or operator over an $\mathbb{F}$-vector space. Then we define

$$p(A) = c_0 I + c_1 A + \ldots + c_m A^m$$

Example 6.4

Let $p(x) = 2 + x^2$ and $A = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$. Then

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

Because matrices do not commute, we cannot say that in general $p(A)q(B) = q(B)p(A)$. However, we can say the following:

Lemma

Given $p, q \in \mathbb{F}[x]$ and $A \in M_{n\times n}(\mathbb{F})$, we have $p(A)q(A) = (pq)(A)$.

Proof. Suppose $p(x) = \sum c_i x^i$ and $q(x) = \sum d_i x^i$. Then

$$(pq)(x) = \sum_k \sum_{i+j=k} c_i d_j x^k$$

so

$$(pq)(A) = \sum_k \sum_{i+j=k} c_i d_i A^k$$

On the other hand, we have

$$p(A)q(A) = \left(\sum c_i A^i\right)\left(\sum d_j A^j\right) = \sum_k \sum_{i+j=k} c_i A^i d_j A^j = \sum_k \sum_{i+j=k} c_i d_j A^k$$

□

Proposition 6.14

Let $A \in M_{n\times n}(\mathbb{F})$. Then there exists some polynomial $p(x) \in \mathbb{F}[x]$ such that $p(A) = 0$.

Proof. Consider the matrices $I_n, A, \ldots, A^{n^2}$. These are all elements of the n^2-dimensional vector space $M_{n\times n}(\mathbb{F})$, so these are linearly dependent. Thus, we

can create a linear relation

$$c_0 I_n + c_1 A + \ldots + c_{n^2} A^{n^2} = 0$$

So we pick the polynomial

$$p(x) = c_0 + c_1 x + \ldots + c_{n^2} x^{n^2} = 0$$

and thus $P(A) = 0$. □

Now let us make some remarks on the significance of this result. First, suppose that the polynomial has degree 2, with $A^2 + 2A + I_n = 0$. Then we have

$$\begin{gathered} A^2 = -2A - I_n \\ A^3 = A^2 A = -2A^2 - A = -2(-2A - I_n) - A = 4A + 2I_n - A = 3_A + 2I_n \\ A^4 = A^3 A = 3A^2 + 2A = -4A - 3I_n \\ \vdots \\ A^n = cA + dI_n \end{gathered}$$

and thus we see that for any n, $A^n \in \operatorname{span}(I_n, A)$. We also have that

$$-A(A + 2I_n) = I_n$$

Thus $A^{-1} = -A - 2I_n$.

Definition 6.13

We call a polynomial $p(x)$ **monic** if the highest nonzero coefficient of p is 1.

We now introduce the minimal polynomial, an important polynomial in our study of linear algebra.

Definition 6.14

Let A be a matrix. We call a monic polynomial $\mu(x) \in \mathbb{F}[x]$ a **minimal polynomial** of A if $\mu(A) = 0$ and, for any other monic polynomial h with $h(A) = 0$, we have $\deg M \leq \deg h$.

Moreover, this polynomial is unique. We can prove this as follows:

Proposition 6.15

Suppose $h(A) = 0$ and μ is a minimal polynomial for A. Then $\mu | h$.

Proof. Applying the division algorithm, $h(x) = q(x)\mu(x) + r(x)$ for some q, r. Then we must have $r = h - q\mu$, and in particular $r(A) = h(A) - q(A)\mu(A) = 0$. So r also satisfies $r(A) = 0$. But if $r \neq 0$, $\deg r < \deg \mu$, contradicting minimality. So we must have $r = 0$ and thus $h = q\mu$, so $\mu | h$. □

Corollary

The minimal polynomial of a matrix is unique.

Proof. Suppose μ, μ' are both minimal polynomials. Then $\mu | \mu'$ and $\mu' | \mu$, but they are both monic, so $\mu = \mu'$. □

Note that we have already constructed a polynomial p such that $p(A)$, and we know that $\deg p \leq n^2$. So the minimal polynomial has degree at most n^2. However, a much stronger bound exists due to Cayley and Hamilton:

Theorem 6.16: Cayley-Hamilton Theorem

Let $A \in M_{n\times n}(\mathbb{F})$ and let $p_A(x)$ be the characteristic polynomial of A. Then $p_A(A) = 0$.

Of course, this means that the minimal polynomial always divides the characteristic polynomial (up to sign), so $\deg \mu \leq n$.

Example 6.5

Let A be an invertible 2×2 matrix. From a previous disucssion, we know that $p_A(x) = x^2 - \operatorname{tr}(A)x + \det(A)$. Then by Cayley-Hamilton, we have $A^2 - \operatorname{tr}(A)A + \det(A)I = 0$. This tells us that

$$A(A - \operatorname{tr}(A)I) = \det(A)I \implies A^{-1} = \frac{1}{\det(A)}(A - \operatorname{tr}(A)I)$$

Here, we easily recover the formula

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

6.6 Matrix Exponentiation (∗)

So far, we have been able to define basic arithmetic operations on matrices, including addition, multiplication, subtraction, and raising to natural powers. We will now define the function given by an exponent power.

Recall that one definition of the function e^x is given by its Taylor expansion,

$$e^x := \sum_{i=1}^{\infty} \frac{x^n}{n!}$$

We will define matrix exponentiation analogously:

Definition 6.15

Let A be an $n \times n$ matrix over $\mathbb{R}$ or $\mathbb{C}$. Then define the **matrix exponent** as

$$e^A = \exp(A) := \sum_{i=0}^{n} \frac{A^n}{n!}$$

Note that $A^0 := I_n$ for all A, and that this sum always converges.

Example 6.6

Suppose $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$. Then the exponent is given by

$$e^A = I + \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} + \frac{1}{2}\begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix} + \ldots = \begin{bmatrix} e^a & 0 \\ 0 & e^b \end{bmatrix}$$

and similar results hold for other diagonal matrices.

This observation gives us the following fact:

Proposition 6.17

Suppose λ is an eigenvalue of A. Then e^λ is an eigenvalue of e^A.

Similarly, the following allows us to calculate exponents more easily:

Proposition 6.18

Suppose A is diagonalizable. Then if the diagonalization is given by $A = SDS^{-1}$, then we have

$$e^A = Se^D S^{-1}$$

Proof. We have

$$e^A = I + (SDS^{-1}) + \frac{1}{2}(SD^2S^{-1} + \ldots = S(I + D + \frac{1}{2}D^2 + \ldots)S^{-1} = Se^D S^{-1}$$

□

Now consider a function of the form e^{At} for some matrix A. If we differentiate this with respect to t, we have the following:

$$\begin{aligned} e^{At} &= I_n + At + \frac{1}{2}A^2t^2 + \dots \\ \frac{d}{dt}e^{At} &= 0 + A + \frac{1}{2}(2A^2t) + \frac{1}{3!}(3A^3t^2) + \dots \\ &= A(I_n + \frac{1}{2}At + \frac{1}{3!}A^2t^2 + \dots) \\ &= Ae^{At} \end{aligned}$$

analogously to the normal exponential function.

Let us now consider the following application. Suppose we have a differential equation with $x(t) : \mathbb{R} \to \mathbb{R}^n$ and $x'(t) = Ax(t)$ for some $A \in M_{n\times n}(\mathbb{R})$. Consider a solution of the form $x(t) = e^{At}x_0$, where x_0 is some vector in $\mathbb{R}^n$ representing the inital state. Then the by the previous calculation, we see that $x'(t) = Ae^{At}x_0$, so functions of this form are indeed solutions. Moreover, this form is in fact the only form, though we do not prove this here.

Example 6.7

Consider the oscillation differential equation, given by

$$\begin{cases} g'(t) = 0g(t) - f(t) \\ f'(t) = g(t) + 0f(t) \end{cases}$$

Then this is equivalent to

$$x'(t) = Ax(t)$$

where we have

$$x(t) = \frac{g(t)}{f(t)}, A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

We can diagonalize A to get

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

and thus

$$At = \begin{bmatrix} 0 & -t \\ t & 0 \end{bmatrix} = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix} \begin{bmatrix} it & 0 \\ 0 & -it \end{bmatrix} \frac{1}{2} \begin{bmatrix} -i & 1 \\ i & 1 \end{bmatrix}$$

Then by the previous discussion, we know that solutions of the form $x(t) = e^{At}x_0$ will work. We can calculate

$$e^{At} = Se^{Dt}S^{-1} = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{bmatrix} \frac{1}{2} \begin{bmatrix} -i & 1 \\ i & 1 \end{bmatrix} = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix}$$

So if we have the original condition $x_0 = \begin{bmatrix} a \\ b \end{bmatrix}$, then our solution is

$$x(t) = e^{At}x_0 = \begin{bmatrix} \cos(t) & -\sin(t) \\ \sin(t) & \cos(t) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a\cos(t) - b\sin(t) \\ a\sin(t) + b\cos(t) \end{bmatrix}$$

Recall from the previous section that for a 2×2 matrix, we have $A^2 - \text{tr}(A)A + \det(A)I_2 = 0$. In other words, we have $A^2 \in \text{span}(I_2, A)$, and by repeatedly multiplying this equation by A, we see that in general, $A^n \in \text{span}(I_2, A)$. Since e^A is a linear combination of these, we have $e^A \in \text{span}(I_2, A)$.

6.7 Complex and Real Vector Spaces (∗)

As we showed before, real transformations do not in general have to have eigenvalues, but complex transformations do. Since $\mathbb{R} \subseteq \mathbb{C}$, this means that every real transformation, considered as a complex transformation, has a (possibly complex) eigenvalue. So far we have glossed over the formalization of this extension, since we were working with matrices. Indeed, it is fairly easy to extend a transformation from $\mathbb{R}^n$ to $\mathbb{C}^n$. However, this becomes more difficult when we are working with abstract vector spaces and would like to maintain this relationship.

If we suppose that $T : V \to V$ for some V over $\mathbb{R}$, then we would like to extend this to a complex transformation $T_{\mathbb{C}} : V_{\mathbb{C}} \to V_{\mathbb{C}}$, where $V_{\mathbb{C}}$ is a complex vector space which "contains" V. One way to do this is is as follows:

Definition 6.16

Given a vector space V, define the **complexification** of V to be

$$V_{\mathbb{C}} \coloneqq \{(v, w) | v, w \in V\} = \{v + iw | v, w \in V\}$$

where the addition operator is defined as in $V \times V$, and scalar multiplication is defined as follows:

$$\lambda = a+bi \in \mathbb{C} \implies \lambda(v+iw) = (a+bi)(v+iw) = (av-bw)+i(bv+aw)$$

Then if we consider the vector space $\mathbb{R}^n$, its complexification is

$$\mathbb{R}^n_{\mathbb{C}} = \left\{ \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} + i \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} | v_i, w_i \in \mathbb{R} \right\} = \mathbb{C}^n$$

as we would expect.

Chapter 7

Inner Product Spaces

7.1 Inner Products

Remember that one of the original motivations for our definition of a vector space was to capture the notion of a space in which we have magnitude and direction. These both exist informally in our abstract definition, with basis vectors representing different directions, and scalar multiplication representing changing the magnitude of a vector. However, none of these can be explicitly quantified in general. We will do so using the *inner product* operator.

Definition 7.1

Let $v, w \in \mathbb{R}^n$ be two vectors. Then the **dot product** of v, w is given by

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \coloneqq v_1 w_1 + \ldots + v_n w_n = \begin{bmatrix} v_1 & \ldots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$$

One of the uses of the dot product is that it allows us to have a (relative) notion of direction. Explicitly, if we have two vectors $v, w \in \mathbb{R}^2$, then we have

$$v \cdot w = \cos(\theta)||v|| \cdot ||w||$$

where θ represents the angle between v, w. For $\mathbb{R}^n$ with larger n, then we simply define the angle to be

$$\theta(v, w) \coloneqq \arccos\left(\frac{v \cdot w}{||v||||w||}\right)$$

Now consider the line $L = \text{span}(u)$ for some u with $||u|| = 1$. Then the transformation which gives the projection onto L is given by

$$\text{proj}_L(v) = (v \cdot u)u$$

For arbitrary vectors w which may not have $||w|| = 1$, we simply consider the direction vector, so that

$$\text{proj}_L(v) = \left(v \cdot \frac{w}{||w||}\right) \frac{w}{||w||} = \frac{v \cdot w}{||w||^2} w = \frac{v \cdot w}{w \cdot w} w$$

Here are some further properties of the dot product:

- $(u + v) \cdot w = u \cdot w + v \cdot w$
- $(\lambda u) \cdot w = \lambda(u \cdot w)$
- $u \cdot w = w \cdot u$
- $v \cdot v \geq 0$, and $v \cdot v = 0$ if and only if $v = 0$.

We should note here that these properties only hold in $\mathbb{R}$. In $\mathbb{C}$, we can have some properties which will become clear when we discuss general inner products. In fields such as $\mathbb{F}_2$, these certainly do not work. For instance, we have

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2 = 0$$

so there are nonzero vectors with zero magnitude (under the dot product).

In order to define a similar product over an abstract vector space, we note that the definition of a dot product does not necessarily work. Even for finite dimensional spaces, which we know we can cast into coordinates with an appropriate isomorphism, there is no canonical choice of isomorphism.

Instead, we will simply use our study of the dot product to capture the key properties that a general product should have, and then work with any such product.

Definition 7.2

Let V be a finite dimensional vector space over $\mathbb{R}$. Then a function $\langle -, - \rangle : V \times V \to \mathbb{R}$ is an **inner product** if it satisfies the following axioms:

- $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$
- $\langle \lambda u, w \rangle = \lambda \langle u, w \rangle$
- $\langle u, v \rangle = \langle v, u \rangle$
- $\langle v, v \rangle > 0$ for any $v \in V$ with $v \neq 0$.

As we have already discussed, the dot product over $\mathbb{R}^n$ satisfies these definitions, as does the dot product composed with any isomorphism from an n dimensional real vector space to $\mathbb{R}^n$. For some more interesting examples, consider the following:

Example 7.1

Let $V = C([-1, 1], \mathbb{R})$ be the set of continuous, real valued functions on the compact interval $[-1, 1]$. Then we define

$$\langle f, g \rangle \coloneqq \int_{-1}^{1} f(t)g(t)dt$$

To check positive definiteness, we have the following:

$$\langle f, f \rangle = \int_{-1}^{1} f^2(t)dt \geq 0$$

Moreover, $\langle f, f \rangle = 0$ if and only if f is nonzero on a set of measure zero. But f is continuous, so this only happens if f is identically zero. So we have positive definiteness. The other axioms follow from properties of the integral.

Example 7.2

Consider the vector space $V = \mathbb{R}[x]_{\leq n}$ of real valued polynomials of degree at most n. We can first define an inner product using the isomorphism to $\mathbb{R}^n$ given by

$$a_0 + a_1 x + \ldots + a_n x^n \mapsto \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}$$

such that

$$\langle a_0 + a_1 x + \ldots + a_n x^n, b_0 + b_1 x + \ldots + b_n x^n \rangle = a_0 b_0 + a_1 b_1 + \ldots + a_n b_n$$

We can also use the previous example, since polynomials are continuous:

$$\langle p, q \rangle = \int_{-1}^{1} p(t)q(t)dt$$

Lastly, we can define a separate inner product, given by

$$\langle p, q \rangle = \int_{0}^{\infty} p(t)q(t)e^{-t}dt$$

where the e^{-t} factor forces this integral to converge for polynomials.

Then we see that the choice of inner product is not unique for a given vector

space. Thus, we need to specify the inner product we use:

Definition 7.3

An **inner product space** over $\mathbb{R}$ is a vector space V over $\mathbb{R}$ together with an inner product $\langle -, - \rangle : V \times V \to \mathbb{R}$, denoted $(V, \langle -, - \rangle)$.

Then for an inner product space, we can finally define what is meant by magnitude and direction:

Definition 7.4

Given an inner product space, the **magnitude** of a vector v is $||v|| := \sqrt{\langle v, v \rangle}$.

Definition 7.5

Given an inner product space, we say that two vectors u, v are **orthogonal** if $\langle u, v \rangle = 0$, written $u \perp v$.

Example 7.3

If we consider the inner product given by

$$\langle f, g \rangle = \int_{-1}^{1} fg dt$$

then the norm is $||f|| = \sqrt{\int_{-1}^{1} f^2 dt}$ which is the L^2 norm on $[-1, 1]$.

Proposition 7.1

If $u \perp v$, then $||u + v||^2 = ||u||^2 + ||v||^2$.

Proof.

$$\begin{aligned} ||u + v||^2 &= \langle u + v, u + v \rangle \\ &= \langle u, u \rangle + \underbrace{\langle v, u \rangle + \langle u, v \rangle}_{0} + \langle v, v \rangle \\ &= ||u||^2 + ||v||^2 \end{aligned}$$

□

Theorem 7.2: Cauchy-Schwarz Inequality

For two vectors $u, v \in V$ in some real inner product space, we have

$$|\langle u, v\rangle| \leq ||u|| \cdot ||v||$$

with equality if and only if u, v are linearly dependent.

Proof. If we consider some scalar α, we have

$$0 \leq \langle u - \alpha v, u - \alpha v\rangle = \langle u, u\rangle - 2\alpha\langle v, u\rangle + \alpha^2\langle v, v\rangle$$

Specifically, we will make the choice

$$\alpha = \frac{\langle u, v\rangle}{\langle v, v\rangle}$$

Geometrically, the intuition for this choice is that we want to find the minimum value of $||u - \alpha v||$, which occurs when $u - \alpha v \perp v$. Then we have

$$\begin{aligned} 0 &\leq \langle u, u\rangle - 2\frac{\langle u, v\rangle}{\langle v, v\rangle}\langle u, v\rangle + \frac{\langle u, v\rangle^2}{\langle v, v\rangle^2}\langle v, v\rangle \\ \frac{\langle u, v\rangle^2}{\langle v, v\rangle} &\leq \langle u, u\rangle \\ \langle u, v\rangle &\leq \langle u, u\rangle\langle v, v\rangle \end{aligned}$$

proving the inequality. Moreover, we see that equality holds if and only if u, v are linearly dependent. □

Example 7.4

Considering again the L^2 norm on continuous functions from $[-1, 1] \to \mathbb{R}$, the Cauchy-Schwarz inequality shows that

$$\left(\int_{-1}^{1} fg\right)^2 \leq \left(\int_{-1}^{1} f\right)\left(\int_{-1}^{1} g\right)$$

Lastly, we have the triangle inequality, we have

$$||u + v|| \leq ||u|| + ||v||$$

Many of the results we have shown in this section are important results in various analysis fields. In particular, we see that it is quite common to quantify abstract spaces in such a way that lines up with our definition of an inner product space here, which makes the study of inner products extremely useful.

7.2 Orthogonality

We will now focus our study on the concept of orthogonality, which we defined in the previous section. Recall that vectors are orthogonal, written $u \perp v$, when $\langle u, v \rangle = 0$.

Theorem 7.3

Suppose $v_1, \ldots, v_m$ are nonzero mutually orthogonal vectors, such that $\langle v_i, v_j \rangle = 0$ for any $i \neq j$. Then $v_1, \ldots, v_m$ are linearly independent.

Proof. Consider an arbitrary linear relation

$$c_1 v_1 + \ldots + c_m v_m = 0$$

Then choose some v_i. If we take the inner product of the linear relation with v_i, we must have 0 by the properties of the inner product:

$$\langle c_1 v_1 + \ldots + c_m v_m, v_i \rangle = c_1 \langle v_1, v_i \rangle + \ldots + c_i \langle v_i, v_i \rangle + \ldots + c_m \langle v_m, v_i \rangle = 0$$

But the all the terms in the middle 0 by the orthogonality assumption, except the $\langle v_i, v_i \rangle$ term:

$$c_i \langle v_i, v_i \rangle = 0$$

Since $v_i \neq 0$, $\langle v_i, v_i \rangle \neq 0$ and thus $c_i = 0$. This holds for all the c_i so $v_1, \ldots, v_m$ are linearly independent. □

Definition 7.6

A list of vectors $v_1, \ldots, v_m \in V$ are **orthonormal** if they are mutually orthogonal and $||v_i|| = 0$ for all i. In other words, $\langle v_i, v_j \rangle = \delta_{ij}$.

Proposition 7.4

Given a set of orthonormal vectors $u_1, \ldots, u_m$, and any scalars, we have

$$||a_1 u_1 + \ldots + a_m u_m|| = a_1^2 + \ldots + a_m^2$$

Proof. Using orthogonality, we can apply the Pythagorean identity m times to get

$$||a_1 u_1 + \ldots + a_m u_m|| = ||a_1 u_1||^2 + \ldots + ||a_m u_m||^2$$

which by the normality assumptions becomes

$$a_1^2 + \ldots + a_m^2$$

□

Definition 7.7

A basis $\mathcal{B}$ of an inner product space V is an **orthonormal basis** if it is orthonormal.

For instance, the standard basis of $\mathbb{R}^n$ is an orthonormal basis, and we can even consider an orthonormal basis to be an abstraction of the standard basis. Observe that if we have some orthonormal basis $\mathcal{B} = \{u_1, \ldots, u_n\}$ and any vector v, then we have

$$v = \langle v, u_1 \rangle u_1 + \ldots + \langle v, u_n \rangle u_n$$

Moreover, we can consider the $\langle v, u_i \rangle$ to be projections of v onto the corresponding basis vectors. In particular, this means that if we put v into $\mathcal{B}-$ coordinates, we have

$$M_{\mathcal{B}}(v) = \begin{bmatrix} \langle v, u_1 \rangle \\ \vdots \\ \langle v, u_n \rangle \end{bmatrix}$$

Definition 7.8

Let W be a subspace of an inner product space $(V, \langle -, - \rangle)$. Suppose $\mathcal{B} = \{u_1, \ldots, u_m\}$ is some basis of W. Then define the **projection map** from V onto W by

$$\text{proj}_W : V \to W$$

such that

$$\text{proj}_W(v) \coloneqq \text{proj}_{u_1}(v) + \ldots + \text{proj}_{u_m}(v)$$

where $\text{proj}_{u_i}(v) = \langle v, u_i \rangle u_i$.

Example 7.5

We will informally discuss an application of inner product projections to Fourier analysis. However, this discussion is mainly for the purpose of intuition and interest, and should not be taken to be rigorous. Consider the set of continuous functions from $[-\pi, \pi] \to \mathbb{R}$. Then the set $1, \cos x, \cos 2x, \ldots, \sin x, \sin 2x, \ldots$ is orthogonal. By way of example,

$$\langle \sin x, \cos x \rangle = \int_{-\pi}^{\pi} \sin x \cos x dx = \frac{1}{2} \int_{-\pi}^{\pi} \sin 2x dx = 0$$

Moreover, if we consider their norms,

$$\langle \sin x, \sin x \rangle = \pi$$

except for the constant function, which has norm

$$\sqrt{\int_{-\pi}^{\pi} 1} = \sqrt{2\pi}$$

So to get normality, we divide the functions:

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \dots$$

Moreover, for periodic differentiable functions, Fourier analysis tells us that we can decompose into trigonometric functions. Thus, we will hand-wave away some of the details and declare that the above set is an orthonormal basis.

In this case, then, we can use this to find the Fourier series of a periodic function. Remember that Fourier analysis tells us that every periodic function can be uniquely decomoposed as

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \sqrt{n} a_n \cos nx/\sqrt{n} + \sqrt{n} b_n \sin nx/\sqrt{n}$$

Then to calculate the a_i, b_i, we project f onto $\cos nx, \sin nx$:

$$\sqrt{n} a_n = \langle f, \frac{\cos nx}{\sqrt{n}} \rangle \implies a_n = \frac{1}{n} \langle f, \cos nx \rangle, b_n = \frac{1}{n} \langle f, \cos nx \rangle$$

Thus, we have used the inner product to (informally) recover the formula for the coefficients of a Fourier transform.

7.3 Gram-Schmidt

As we have seen, orthonormal bases of vector spaces are particularly powerful. Thus, it is of interest to us to find a way to build an orthogonal basis out of any basis of the space. This process is known as Gram-Schmidt orthonormalization.

The basic idea is that given some basis $v_1, \dots, v_m \in V$, we wish to recursively build an orthogonal basis. We do this as follows:

- First, we take the normalized version of v_1 to be our first vector: $u_1 = \frac{v_1}{||v_1||}$.
- Then, we obtain the perpendicular component of v_2 with respect to u_1: $v_{2\perp} = v_2 - \langle v_2, u_1 \rangle u_1$.
- This new vector is orthogonal to u_1, but it is not normalized, so we normalize it: $u_2 = \frac{v_{2\perp}}{||v_{2\perp}||}$.

- We continue with v_3, obtain the perpendicular component with respect to the plane spanned by u_1, u_2: $v_{3\perp} = v_3 - \langle v_3, u_1\rangle u_1 - \langle v_3, u_2\rangle u_2$.
- We then normalize.
- Repeat until we have m basis vectors, with the recursive definition

$$u_n = \frac{v_n - \sum_{i=1}^{n-1}\langle v_n, u_i\rangle u_i}{||v_n - \sum_{i=1}^{n-1}\langle v_n, u_i\rangle u_i||}$$

or more concisely, $v_{n\perp} = v_n - \sum_{i=1}^{n-1}\langle v_n, u_i\rangle u_i$, $u_n = \frac{v_{n\perp}}{||v_{n\perp}||}$.

Example 7.6

Consider the basis of $\mathbb{R}[x]_{\leq 2}$ of $1, x, x^2$. Suppose we have the inner product $\langle p, q\rangle = \int_0^1 pq$. Then we normalize the first basis vector:

$$u_1 = \frac{1}{||1||} = 1$$

We take the perpendicular component of x with respect to 1:

$$v_{2\perp} = x - \langle x, 1\rangle 1 = x - \int_0^1 x dx = x - \frac{1}{2}$$

To normalize, we calculate:

$$||x - \frac{1}{2}|| = \langle x - \frac{1}{2}, x - \frac{1}{2}\rangle = \langle x, x\rangle - \langle 1, x\rangle + \frac{1}{4}\langle 1, 1\rangle = \frac{1}{3} - \frac{1}{2} + \frac{1}{4} = \frac{1}{12}$$

So we have

$$u_2 = \sqrt{12}(x - \frac{1}{2})$$

Then, we take the perpendicular component of x^2:

$$\begin{aligned} v_{3\perp} &= x^2 - \langle x^2, \sqrt{12}(x - \frac{1}{2})\rangle\sqrt{12}(x - \frac{1}{2}) - \langle x^2, 1\rangle 1 \\ &= x^2 - 12\langle x^2, x - \frac{1}{2}\rangle(x - \frac{1}{2} - \frac{1}{3} \\ &= x^2 - (x - \frac{1}{2}) - \frac{1}{3} = x^2 - x + \frac{1}{6} \end{aligned}$$

So our orthonormal basis is

$$\{1, \sqrt{12}(x - \frac{1}{2}, \frac{x^2 - x + \frac{1}{6}}{||x^2 - x + \frac{1}{6}||}\}$$

Theorem 7.5: QR Factorization

For any $M \in M_{n\times m}$ with linearly independent columns (nonsingular), then $M = QR$ where Q is an $n \times m$ matrix with orthonormal columns, and R is an $m \times m$ upper triangular matrix.

We will not discuss this process in depth. However, note that when M has columns

$$M = \begin{bmatrix} | & | & | \\ v_1 & \dots & v_m \\ | & | & | \end{bmatrix}$$

then we will obtain Q by performing Gram-Schmidt on $v_1, \dots, v_m$, then using the resulting vectors $u_1, \dots, u_m$ to form the columns of Q:

$$Q = \begin{bmatrix} | & | & | \\ u_1 & \dots & u_m \\ | & | & | \end{bmatrix}$$

In this case, then our R should be

$$R = \begin{bmatrix} ||v_1|| & \langle v_2, u_1\rangle & \dots & \langle v_m, u_1\rangle \\ & ||v_{2\perp}|| & \dots & \langle v_m, u_2\rangle \\ & & \ddots & \vdots \\ & & & ||v_{m\perp}|| \end{bmatrix}$$

As a few consequences of Gram-Schmidt, we have the following:

Corollary

For a finite dimensional inner product space $(V, \langle -, -\rangle)$, there is an orthonormal basis.

Proof. Pick any basis of V and apply Gram-Schmidt. □

Corollary

If $u_1, \dots, u_k$ is an orthonormal list of vectors in V, then it can be extended to an orthonormal basis.

Proof. Extend to a basis, then apply Gram-Schmidt. Because the first k vectors are already orthonormal to the previous vectors, Gram-Schmidt will not change them. □

7.4 Transpositions and Projections in $\mathbb{R}^n$

We will now discuss the transpose of a matrix. From a computational viewpoint, the reader may have already learned that given an $m \times n$ matrix A, the transpose is given by an $n \times m$ matrix A^T which is A "flipped" along the primary diagonal.

Proposition 7.6

We have the following properties:

- $(A^T)^T = A$
- $(AB)^T = B^T A^T$
- For (square) invertible matrices, $(A^{-1})^T = (A^T)^{-1}$
- For column vectors $v, w \in \mathbb{R}^n$, $v \cdot w = v^T w$

Proposition 7.7

If $V \subseteq \mathbb{R}^n$ has an orthonormal basis $u_1, \ldots, u_m$, then the matrix of the projection onto V is $A = QQ^T$, where

$$Q = \begin{bmatrix} | & | & | \\ u_1 & \ldots & u_m \\ | & | & | \end{bmatrix}$$

Proof. We have $\text{proj}_V(x) = (x \cdot u_1)u_1 + \ldots + (x \cdot u_m)u_m$, so in matrix form we have

$$\begin{bmatrix} | & | & | \\ u_1 & \ldots & u_m \\ | & | & | \end{bmatrix} \begin{bmatrix} x \cdot u_1 \\ \vdots \\ x \cdot u_m \end{bmatrix}$$

which is also

$$\begin{bmatrix} | & | & | \\ u_1 & \ldots & u_m \\ | & | & | \end{bmatrix} \begin{bmatrix} - & u_1 & - \\ & \vdots & \\ - & u_m & - \end{bmatrix} x = QQ^T x$$

So $A = QQ^T$. □

It is often useful to denote the orthogonal "parts" of a vector space with respect to some subspace. We call this the *orthogonal complement*:

Definition 7.9

Given a subspace U of V, we define the **orthogonal complement** of U to be

$$U^\perp \coloneqq \{v \in V | \langle u, v \rangle = 0 \forall u \in U\}$$

Proposition 7.8

$U \cap U^\perp = \{0\}$.

Proof. Pick some $v \in U \cap U^\perp$. By definition, we have $\langle v, v\rangle = 0$ which implies $v = 0$. □

Proposition 7.9

If U is a subspace of V, then $V = U \oplus U^\perp$.

Proof. Pick some orthonormal basis $u_1, \ldots, u_k$ of U. Then extend this to an orthonormal basis $u_1, \ldots, u_k, w_1 \ldots, u_l$ of V. Then we have $V = \text{span}(u_1, \ldots, u_k) \oplus \text{span}(w_1, \ldots, w_l)$. The first term is clearly U. We want to show that the second term is $U^\perp$.

Let $W = \text{span}(w_1, \ldots, w_l)$. Choose some $w \in W$. Then $w = d_1 w_1 + \ldots + d_l w_l$. Pick $u \in U$ and write $u = c_1 u_1 + \ldots + c_k u_k$. By orthonormality, $\langle u, w\rangle$ so $W \subseteq U^\perp$.

In the other direction, choose $v \in U^\perp$. Then $v = \langle v, u_1\rangle u_1 + \ldots + \langle v, u_k\rangle u_k + \langle v, w_1\rangle w_1 + \ldots + \langle v, w_l\rangle w_l$. The $\langle v, u_i\rangle$ terms are all 0, so $v = \langle v, w_1\rangle w_1 + \ldots + \langle v, w_l\rangle w_l \in W$. So $U^\perp \subseteq W$.

Thus $V = U \oplus W = U \oplus U^\perp$. □

Thus, we see that we can calculate the orthogonal complement by finding an orthonormal basis of U, extending to an orthonormal basis of the full space, and simply taking $U^\perp$ to be the span of the added vectors.

Similarly, we see that for any vector V, we can write it as $v = \text{proj}_U(v) + \text{proj}_{U^\perp}(v)$. This is generally true for subspaces forming a direct some of the original space.

Proposition 7.10

The following are consequences of the above discussion:

- $\dim U^\perp = \dim V - \dim U$
- $(U^\perp)^\perp$
- $U^\perp = \{0\} \iff U = V$

Since we have $\dim U^\perp = \dim V - \dim U = \dim V/U$, there is an isomorphism between $U^\perp$ and V/U. But more significantly, there is a canonical such isomorphism. For instance, if we quotient $\mathbb{R}^3$ by the xy plane, then the orthogonal complement is the z axis, and the quotient is the set of parallel planes. Then we can identify a plane with its z coordinate, constructing the isomorphism.

More generally, for each coset, we can pick a representative u such that $u+U$ is the coset, and u is in the orthogonal complement. This constructs the canonical isomorphism.

To investigate the theoretical basis of transposition, consider the dot product between two vectors $u, w \in \mathbb{R}^n$. We can write this as

$$u \cdot w = u^T w$$

under the basic definition of transposition. If we consider this under some transformation, we have

$$Av \cdot w = (Av) \cdot w = (Av)^T w = v^T A^T w = v \cdot (A^T w)$$

Proposition 7.11

Let $A \in M_{n\times n}(\mathbb{R})$. Then $(\text{im}(A))^\perp = \ker(A^T)$.

Proof. Let $x \in (\text{im}(A))^\perp$. Suppose the columns of A are $u_1, \ldots, u_n \in \text{im}\, A$. Then for any u_i, we have $x \cdot u_i = 0$. Since this is true for all i, we get

$$\begin{bmatrix} - & u_1 & - \\ \vdots & \vdots & \vdots \\ - & u_n & - \end{bmatrix} x = \mathbf{0}$$

But this matrix is precisely A^T, and thus $x \in \ker(A^T)$. This holds in the other direction as well, so $\ker(A^T) = (\text{im}(A))^\perp$. □

However, we can prove the above proposition with a different method, using the below observation:

Proposition 7.12

If V is an inner product space and there is some $v \in V$ such that $\langle v, u\rangle = 0$ for all $u \in V$. Then $v = 0$.

Proof. Take $u = v$. Then $\langle v, v\rangle = 0$ but by the axioms of an inner product, we must have $v = 0$. □

This gives an alternate proof for the previous proposition:

Alternate Proof. Let $x \in (\text{im}(A))^\perp$. Then for any $v \in \mathbb{R}^n$, we have $x \cdot Av = 0$ (since $Av \in \text{im}\, A$). Using the transposition trick, we have $(A^T x) \cdot v = 0$ for all v, and by the above fact we then must have $A^T x = 0$. So $x \in \ker A^T$. □

We will now discuss the *covariance matrix*, which is useful in many statistical and probabilistic applications of linear algebra.

Definition 7.10

Let $A \in M_{n\times n}(\mathbb{R}^n)$. Then the **covariance matrix** of A is defined as A^TA.

Suppose our matrix A is given by

$$\begin{bmatrix} | & | & | \\ v_1 & \dots & v_n \\ | & | & | \end{bmatrix}$$

Then consider the covariance matrix. By the process of matrix multiplication, the (i,j)th entry of A^TA is given by $v_i \cdot v_j$. This becomes important because if each of the v_i is a single data point in $\mathbb{R}^n$, wrapped into a matrix, then the covariance matrix tells us how correlated two data points are.

Example 7.7

Suppose users are asked to rate two movies on a scale of -5 to 5, and their responses are recorded. Suppose three users have the following scores:

$$u_1 = \begin{bmatrix} 5 \\ 5 \end{bmatrix}, u_2 = \begin{bmatrix} 5 \\ -5 \end{bmatrix}, u_3 = \begin{bmatrix} -5 \\ 5 \end{bmatrix}$$

Note that here, u_2 and u_3 are linearly dependent, and thus maximally correlated. Meanwhile, if we calculate the dot products between u_1 and u_2, we see that they are uncorrelated. If a company wants to tailor movie recommendations to different user, then it is more computationally efficient to classify the users by taste – that is, to consider users with correlated tastes to be the same.

Proposition 7.13

Suppose the columns of A, $v_1, \dots, v_m$ are linearly independent. Then A^TA is invertible.

Proof. The kernel of A^TA is given by $\{v \in \mathbb{R}^n : A^TAv = 0\} = \{v : Av \in \ker(A^T)\}$. Then we have $Av \in \operatorname{im} A$. We also have $Av \in \ker(A^T) = (\operatorname{im} A)^\perp$. But $(\operatorname{im} A)^\perp \cap \operatorname{im} A = \{0\}$, so we have $Av = 0$. But A has linearly independent columns, and thus $v = 0$. So $\ker(A^TA) = \{0\}$ and we have invertibility. □

We now consider the application of inner products and transposes to least squares regression. Suppose some equation $Ax = b$ has no solutions. However, we would like to find the "closest solution." That is, we want to find the vector $x^* \in \mathbb{R}^n$ such that $||Ax^* - b||$ is minimized.

Note that the image of A is, in general, a hyperplanar subspace of $\mathbb{R}^n$. Moreover, $Ax = b$ has solutions precisely when $b \in \operatorname{im} A$, or when b lies in this hyperplane. On the other hand, if b is not in the hyperplane, then the vector $v \in \operatorname{im} A$ which minimizes the value $||b - v||$ is the projection of b into the hyperplane. Then to find the "closest solution," we simply need to find some x^* such that $Ax^* = \operatorname{proj}_V b$, where $V = \operatorname{im} A$.

In order to actually solve this equation, we could solve this normally, but we can also use transposes to find this as well. Note that we have $b - \operatorname{proj}_V b \perp V$, so $Ax^* - b \in V^\perp = \ker A^T$. Thus we have $A^T(Ax^* - b) = 0$. So our least square solution is a solution to

$$A^T A x^* = A^T b$$

When A has linearly independent columns, we have that $A^T A$ is invertible, so we can find a solution by calculating

$$x^* = (A^T A)^{-1} A^T b$$

Now in the case of least squares regression, we are given some dataset of points, and we want to draw some curve. If this is linear, then the curve is $y = ax + b$; if it is quadratic, the curve is $y = ax^2 + bx + c$, and so on. This also generalizes to higher dimensions. The key is that if b is the prediction, then we need to find the parameters x such that $Ax = b$ where A is our data matrix.

Example 7.8

Suppose we have some dataset with linear relations betwen features and price, and we have the system

$$\begin{cases} 50x + 300y = 600 \\ 200x \quad\quad = 450 \\ 100x + 100y = 250 \end{cases}$$

Then we want to find the parameters x, y such that the linear model

$$\begin{bmatrix} 150 & 300 \\ 200 & 0 \\ 100 & 100 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

is closest to

$$\begin{bmatrix} 600 \\ 450 \\ 250 \end{bmatrix}$$

This can be done by calculating

$$(A^T A)^{-1} A^T b$$

which gives parameters

$$\begin{bmatrix} 2.18 \\ 0.85 \end{bmatrix}$$

In the quadratic case, we would record the values of $1, x, x^2$; in the bivariate quadratic case we would record $1, x, y, x^2, y^2, xy$, and so on.

Then in a general inner product space, we can similarly say that the best approximation for a vector $v \in V$ by a vector in the subspace $w \in W$ is given by $w = \text{proj}_W v$. To calculate this, we just pick some orthonormal basis $u_1, \ldots, u_m$ of W, and we calculate the projections onto each basis vector:

$$\text{proj}_W v = \langle v, u_1 \rangle u_1 + \ldots + \langle v, u_m \rangle u_m$$

7.5 Isometries and Orthogonal Matrices

Definition 7.11

Let Q be an nonsquare $n \times m$ real matrix, with $n > m$. We call Q **semiorthogonal** when the columns of Q are orthonormal. When n, m, we call Q semiorthogonal when the rows are orthonormal instead. However, in the following we will assume $n > m$ unless stated otherwise.

Note that since each columnn vector is in $\mathbb{R}^n$, we must have $n > m$ in order to have orthonormality.

Proposition 7.14

Q is semiorthogonal if and only if $Q^T Q = I_m$.

Proof. If the columns of Q are $u_1, \ldots, u_m$, then the covariance matrix $Q^T Q$ is made of the dot products between u_i. But they are orthonormal, so the covariance matrix is the identity, and thus $Q^T Q = I_m$. This argument works in both ways. □

Definition 7.12

Let Q be a real square matrix. We call Q **orthogonal** if its columns are orthonormal (equivalently, if its columns form an orthonormal basis of $\mathbb{R}^n$).

Proposition 7.15

If A, B are $n \times n$ orthogonal matrices, then we have the following:

1. $\det A = \pm 1$.
2. AB is orthogonal.

Proof. 1. We have $\det A = \det A^T$ (by Laplace expansion), so $1 = \det I = \det A^T A = \det A^T \det A = (\det A)^2$. So $\det A = \pm 1$.

2. We have $(AB)^T AB = B^T A^T AB = B^T I_n B = I_n$.

□

Proposition 7.16

A is orthogonal if and only if $A^T A = AA^T = I_n$, and thus if and only if $A^{-1} = A^T$.

Orthogonal matrices are of particular importance to computing because they allow for robust methods of solving matrix equations of the form

$$Ax = b$$

In particular, we would like a stable method, such that small changes in the data do not affect our solution much. One method of doing this is to factor A using QR factorization, giving

$$QRx = b$$

for some Q semiorthogonal and R upper triangular. Then we use the properties of semiorthogonality to get

$$Q^T QRx = Rx = Q^T b$$

(Note that this is not an equivalent statement but an implication, so we are now looking for the least square solution rather than the true solution). If we suppose R takes the form

$$\begin{bmatrix} c_{11} & \cdots & c_{1n} \\ & \ddots & \vdots \\ & & c_{nn} \end{bmatrix}$$

then this gives the system

$$\begin{cases} \vdots \\ c_{n-1,n-1}x_{n-1} + c_{n-1,n}x_n = d_{n-1} \\ c_{n,n}x_n = d_n \end{cases}$$

where $d_1, \ldots, d_n$ are the entries of $Q^T b$. Then we can solve this using only n divisions, which allow for a much more stable procedure.

Definition 7.13

Let $(V, \langle -, - \rangle_V)$ and $(W, \langle -, - \rangle_W)$ be two inner product spaces. Let $T : V \to W$ be a linear transformation. Then we call T an **isometry** or inner product preserving if, for any $u, v \in V$ we have $\langle u, v \rangle = \langle Tu, Tv \rangle$.

In particular, an isometry preserves the norm of any vector. In the opposite direction, we want an isometry to preserve lengths and angles between vectors. But the polarization identity says that

$$\langle v, w\rangle = \frac{1}{2}(||v+w||^2 - ||v||^2 - ||w||^2)$$

which shows that it is acutally sufficient to only check norm. Thus, T is an isometry if and only if it preserves norm.

Proposition 7.17

If T is an isometry, then T is injective.

Proof. If $v \in \ker T$ then $Tv = 0$ and thus $||0|| = ||Tv|| = ||v||$ so $v = 0$. □

Proposition 7.18

Let $v_1, \ldots, v_m \in V$ be linearly independent and let $T : V \to W$ be an isometry. Then $T(v_1), \ldots, T(v_m)$ are linearly independent.

Moreover, they preserve orthogonality:

Proposition 7.19

Let $v_1, \ldots, v_m \in V$ be orthonormal. Let $T : V \to W$ be an isometry. Then $T(v_1), \ldots, T(v_i)$ are orthonormal.

Lemma

If A, B are square real $n \times n$ matrices and

$$x^T Ay = x^T By$$

for all $x, y \in \mathbb{R}^n$, then $A = B$.

Proof. Let $x = e_i, y = e_j$ for some i, j. Then $e_i^T Ae_j$ is the i, jth entry of A. But by assumption it is also the i, jth entry of B, so $A = B$. □

Proposition 7.20

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a transformation. Then T is an isometry if and only if $M(T)$ is semiorthogonal.

Proof. Let T be a isometry and let $A = M(T)$. Let $x, y \in \mathbb{R}^n$. Then we have

$$x^T y = x \cdot y = Ax \cdot Ay = (Ax)^T Ay = x^T A^T Ay$$

We also have

$$x \cdot y = x^T I y$$

Then we have $x \cdot y = Ax \cdot Ay \implies I = A^T A$ by the lemma, and the reverse direction follows as well. □

Corollary

$T : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible isometry if and only if $M(T)$ is orthogonal.

7.6 The Spectral Theorem

In this section, we will introduce the spectral theorem for linear algebra. In fact, there are many forms of the spectral theorem in various fields, but we will work with the theorem in this form:

Theorem: Spectral Theorem

Let A be an $n \times n$ real matrix. If A is symmetric, then it is diagonalizable. Moreover, A is symmetric if and only if it is orthogonally diagonalizable.

Definition 7.14

Let $(V, \langle -, - \rangle)$ be an inner product space and $T : V \to V$ be linear. We say $v_1, \ldots, v_n$ is an **orthonormal eigenbasis** of V for T if it is an eigenbasis and an orthonormal basis. If such a basis exists, we say that T is **orthogonally diagonalizable**

Definition 7.15

A matrix $A \in M_{n \times n}(\mathbb{R})$ is **orthogonally diagonalizable** if there exists an orthogonal $n \times n$ matrix S such that $A = SDS^T$ where D is diagonal.

Note that for orthogonal matrices, $S^T = S^{-1}$, so what this definition essentially says is that A is diagonalizable, and the change of basis matrix is orthogonal.

Proposition 7.21

$T : \mathbb{R}^n \to \mathbb{R}^n$ is orthogonally diagonalizable if and only if $M(T)$ is orthogonally diagonalizable.

Proof. ($\Longleftarrow$) Let the orthonormal eigenbasis be the columns of S.

($\Longrightarrow$) Let $A = M(T)$. Let $v_1, \ldots, v_n$ be an orthonormal eigenbasis of $\mathbb{R}^n$ for T

with associated eigenvalues $\lambda_1, \ldots, \lambda_n$. Then we set our diagonalization to be

$$S = \begin{bmatrix} | & | & | \\ v_1 & \ldots & v_n \\ | & | & | \end{bmatrix}, D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

But by the assumption that the v_i are orthonormal, we see that S is orthogonal and thus $S^{-1} = S^T$, so we have

$$A = SDS^{-1} = SDS^T$$

with S orthogonal. So $A = M(T)$ is orthogonally diagonalizable. □

Example 7.9

Consider the matrix $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This is reflection over the line $y = x$, so the eigenspaces are the lines $y = x$ ($\lambda = 1$) and the line $y = -x$ ($\lambda = -1$). Then we pick a unit vector in each line and we have an orthonormal eigenbasis.

We now consider an algorithm which will allow us to orthogonally diagonalize a matrix. Recall that in the general diagonalization case, we use the following process:

1. Find the eigenvalues using the characteristic polynomial.
2. Find a basis for each eigenspace.
3. Concatenate the bases to get an eigenbasis $v_1, \ldots, v_n$.
4. Let S be the matrix with columns $v_1, \ldots, v_n$. Let D be the corresponding eigenbases.

In the case of orthogonal diagonalization, we need to make sure that our basis is orthogonal. To do so, we modify step two by applying Gram-Schmidt to find an *orthonormal* basis of of each eigenspace. However, we need to guarantee that the concatenation process preserves orthogonality.

Recall that for diagonalizability, an equivalent condition was that the eigenbases direct summed to the entire space. For orthogonalizability, we have the following instead:

Proposition 7.22

T is orthogonally diagonalizable if and only if $V = E_{\lambda_1} \oplus \ldots \oplus E_{\lambda_k}$ and $E_{\lambda_i} \perp E_{\lambda_j}$ for any $\lambda_i \neq \lambda_j$.

Thus, the additional knowledge that the eigenspaces are mutually orthogonal guarantees that the concatenated orthonormal eigenbases forms an overall orthonormal eigenbasis.

We will now progress to proving the spectral theorem.

Theorem: Spectral Theorem

Let A be an $n \times n$ real matrix. If A is symmetric, then it is diagonalizable. Moreover, A is symmetric if and only if it is orthogonally diagonalizable.

Proof. ($\Longleftarrow$) Assume A is orthogonally diagonalizable with $A = SDS^T$. Applying the transpose gives us $A^T = (SDS^T)^T = (S^T)^T D^T S^T = SDS^T$ (since D is diagonalizable). So $A^T = A$ and A is symmetric. □

In the forward direction, we make use of the following lemma.

Lemma

Let A be symmetric and real. Let $\lambda_1 \neq \lambda_2$ be distinct real eigenvalues. Then $E_{\lambda_1} \perp E_{\lambda_2}$.

Proof. Let v_1, v_2 be eigenvectors with respective eigenvalues λ_1, λ_2. We want to show that $v_1 \perp v_2$. Consider the product

$$\lambda_1 v_1 \cdot v_2 = (Av_1) \cdot v_2 = v_1 \cdot A^T v_2 = v_1 \cdot Av_2 = \lambda_2 v_1 \cdot v_2$$

So $\lambda_1(v_1 \cdot v_2) = \lambda_2(v_1 \cdot v_2)$, but $\lambda_1 \neq \lambda_2$, so $v_1 \cdot v_2 = 0$. □

Using this lemma, the proof reduces to the problem of demonstrating that A is real diagonalizable. To do so, we will need to investigate the complex inner product.

Definition 7.16

Given a complex number $z = a + bi$, the **complex magnitude** of z is

$$|z| \coloneqq \sqrt{a^2 + b^2} = \sqrt{z\overline{z}} \in \mathbb{R}$$

Definition 7.17

Given two vectors $z, w \in \mathbb{C}^n$, then the complex dot product is given by

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} \coloneqq z_1\overline{w}_1 + \ldots + z_n\overline{w_n}$$

In this case, the norm of z is given by $||z|| = \sqrt{z \cdot z} = \sqrt{z_1\overline{z}_1 + \ldots + z_n\overline{z}_n} = \sqrt{|z_1|^2 + \ldots + |z_n|^2}$. Then just as we generalized the real dot product to real inner products, we can generalize the complex dot product to complex inner products.

Definition 7.18

Let V be a complex vector space. A **complex inner product** on V is a function $\langle -, - \rangle : V \times V \to \mathbb{C}$ such that

1. $\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle$ (additivity in first argument)
2. $\langle \lambda u, w \rangle = \lambda \langle u, w \rangle$ (homogeneity in the first argument)
3. $\langle u, w \rangle = \overline{\langle w, u \rangle}$ (conjugate symmetry)
4. $\langle v, v \rangle > 0$ for all $v \neq 0 \in V$.

Note that the comparison in axiom 4 makes sense because $\langle v, v \rangle = \overline{\langle v, v \rangle}$, so $\langle v, v \rangle \in \mathbb{R}$. Note also that we still have additivity in the second argument. However, we do not have homogeneity in the second argument, but instead conjugate homogeneity:

Proposition 7.23

If $u, w \in V$ for some complex inner product space V, then for any $\lambda \in \mathbb{C}$,

$$\langle u, \lambda w \rangle = \overline{\lambda} \langle u, w \rangle$$

Proof.

$$\begin{aligned} \langle u, \lambda w \rangle &= \overline{\langle \lambda w, u \rangle} \\ &= \overline{\lambda \langle w, u \rangle} \\ &= \overline{\lambda} \overline{\langle w, u \rangle} \\ &= \overline{\lambda} \langle u, w \rangle \end{aligned}$$

□

Under this definition, the dot product is then given by $z \cdot w = z^T \overline{w}$.

Definition 7.19

Let A be a complex square matrix. Then the **Hermitian conjugate** of A is A^*, where $A^* = \overline{A^T}$ is the conjugate transpose of A.

Proposition 7.24

If $v, w \in \mathbb{C}^n$ and A is a complex square matrix, then $(Av) \cdot w = v \cdot (A^* w)$.

Proof.

$$\begin{aligned}(Av) \cdot w &= (Av)^T \overline{w} \\ &= v^T A^T \overline{w} \\ &= v^T \overline{A^* w} \\ &= v \cdot (A^* w)\end{aligned}$$

□

Proposition 7.25

Let A be a real symmetric matrix. Then every complex eigenvalue of A is real.

Proof. Let λ be a complex eigenvalue. Let $v \in \mathbb{C}^n$ be an eigenvector with eigenvalue λ. Then we have

$$(Av) \cdot v = \lambda v \cdot v$$

But since A is symmetric, $A^* = \overline{A}$. Since it is also real, $\overline{A} = A$, so $A^* = A$.

$$\begin{aligned}(Av) \cdot v &= v \cdot (A^* v) \\ &= v \cdot (Av) \\ &= v \cdot (\lambda v) \\ &= \overline{\lambda} v \cdot v\end{aligned}$$

So we have

$$\overline{\lambda} v \cdot v = \lambda v \cdot v$$

and since $v \cdot v > 0$, we conclude that $\lambda = \overline{\lambda}$, so $\lambda \in \mathbb{R}$. □

Theorem 7.26

Let V be a finite dimensional real vector space. Let $T : V \to V$ be linear with no nonreal complex eigenvalues. Then there exists a basis $\mathcal{B} = v_1, \ldots, v_n$ such that $M_{\mathcal{B}}(T)$ is upper triangular.

Proof. We sketch this proof only since the complex version was proved for homework.

Proceed by induction. Let λ be an eigenvalue of T (which we know exists because there is at least one complex eigenvalue, and all complex eigenvalues are real. Then define

$$W = \operatorname{im}(T - \lambda \operatorname{Id})$$

Then we have $\dim W \leq \dim V - 1$ by rank nullity. By construction, W is T-invariant. Then define $T_W : W \to W$ to be the restriction of T to W. T_W cannot have eigenvalues other than the eigenvalues of T, so the hypothesis applies. By the inductive hypothesis, there exists a basis $\mathcal{B}'$ of W such that

$M_{\mathcal{B}'}(T)$ is upper triangular, meaning that $T(w_i) \in \text{span}(w_1, \ldots, w_i)$ for any i. Then we extend this to a basis $w_1, \ldots, w_k, v_1, \ldots, v_l$ of V. For any of the v_i, we have $T(v_i) = (T - \lambda \,\text{Id})(v_i) + \lambda v_i \in W + \text{span}(v_i) \subseteq \text{span}(w_1, \ldots, w_k, v_1, \ldots, v_l)$, So T is upper triangular with respect to this basis. □

Remark

In fact, we can use orthonormal bases instead of bases in the above proof. So instead we have that every real matrix with only real eigenvalues is upper triangular with respect to some orthonormal basis.

Given this, let $\mathcal{B}$ be this orthonormal basis and suppose we let $U = M_{\mathcal{B}}(T)$. Let $A = M(T)$. Then A, U are similar so $A = SUS^{-1}$ for some S. But the columns of S are the elements of $\mathcal{B}$, and the elements of $\mathcal{B}$ are orthonormal, so S is orthogonal and thus A is orthogonally diagonalizable. This allows us to finish the proof of the spectral theorem. We repeat the statement and first half of the proof here for convenience.

Theorem 7.27: Spectral Theorem

Let A be an $n \times n$ real matrix. If A is symmetric, then it is diagonalizable. Moreover, A is symmetric if and only if it is orthogonally diagonalizable.

Proof. ($\implies$) Let A be a real symmetric matrix. We showed that all complex eigenvalues of A are real. Then A is similar to some upper triangular matrix U by some orthogonal matrix S. Then we have

$$\begin{aligned} A &= SUS^T \\ U &= S^T AS \\ U^T &= (S^T AS)^T \\ &= S^T A^T S \\ &= S^T A^T S \\ &= S^T AS = U \end{aligned}$$

So $U^T = U$, and thus U is both upper triangular and symmetric, and thus must be diagonal. So A is orthogonally diagonalizable.

($\impliedby$) Assume A is orthogonally diagonalizable with $A = SDS^T$. Applying the transpose gives us $A^T = (SDS^T)^T = (S^T)^T D^T S^T = SDS^T$ (since D is diagonalizable). So $A^T = A$ and A is symmetric. □

7.7 Inner Products on $\mathbb{R}^n$

We will now discuss arbitrary inner products on $\mathbb{R}^n$. Of course, the dot product satisfies the inner product axioms, but there are many other ways of constructing a dot product. Suppose we have some inner product of $\mathbb{R}^n$, given by $(\mathbb{R}^n, \langle -, - \rangle)$.

Example 7.10

We can define the following inner products on $\mathbb{R}^2$:

- The normal dot product.
- $\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rangle \coloneqq 2x_1y_1 + 3x_2y_2$
- $\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rangle \coloneqq x_1y_1 + x_1y_2 + x_2y_1 + 3x_2y_2 = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$.
  To check positive definiteness, we can verify that

  $$\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rangle = x_1^2 + 2x_1x_2 + 3x_2^2 = (x_1 + x_2)^2 + 2x_2^2 \geq 0$$

  with equality if and only if $x_1 = x_2 = 0$.

To see a non example, consider

$$\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rangle \coloneqq x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2$$

This violates positive definiteness because we have $\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rangle = x_1^2 + 2x_1x_2 + x_2^2 = (x_1 + x_2)^2 \geq 0$, but this can have equality for nonzero vectors.

The general idea here is that we can associate inner product with matrices that encode the transformation. Consider some inner product. Using linearity, we have

$$\langle x, y \rangle = \langle x_1e_1 + \ldots + x_ne_n, y_1e_1 + \ldots + y_ne_n \rangle = \sum_{i,j} \langle x_ie_i, y_je_j \rangle$$

Thus, every inner product is identified by the coefficients of x_iy_j. We can encode this by writing a matrix A that has entries $a_{ij} = \langle e_i, e_j \rangle$, then this becomes

$$\sum_{i,j} \langle x_ie_i, y_je_j \rangle = \sum_{i,j} x_iy_j \langle e_i, e_j \rangle = \sum_{i,j} x_iy_ja_{ij}$$

On the other hand, if we consider the multiplication x^TAy, then we would get

$$\begin{bmatrix} \ldots & x_i & \ldots \end{bmatrix} \begin{bmatrix} & & \\ & a_{ij} & \\ & & \end{bmatrix} \begin{bmatrix} \vdots \\ y_j \\ \vdots \end{bmatrix} = \begin{bmatrix} \ldots & x_1 & \ldots \end{bmatrix} \begin{bmatrix} \vdots \\ \ldots a_{ij}y_j \ldots \\ \vdots \end{bmatrix} = \ldots + a_{ij}x_iy_j + \ldots$$

So the total multiplication is

$$x^TAy = \sum_{i,j} a_{ij}x_iy_j = \langle x, y \rangle$$

Thus, we can associate the inner product with a matrix encoding the coefficients of the various $x_i y_j$ terms. This allows us to easily translate between matrices and explicit expressions of inner products.

Example 7.11

$\langle \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rangle := 2x_1y_1 + 3x_2y_2$ has matrix $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$.

Of course, not every matrix can be associated with a valid inner product, as the nonexample in the first example in this section showed. This leads us to investigate the question of which matrices are able to represent inner products.

Definition 7.20

Let A be a square real matrix. Then define $\langle -, - \rangle_A$ such that $\langle x, y \rangle_A = x^T A y$.

By the properties of matrix multiplication, we automatically get additivity and homogeneity for any matrix. To check symmetricity, we observe that linearity means it is enough to check the standard basis. By our definition, we have

$$\langle e_i, e_j \rangle_A = a_{ij}, \langle e_j, e_i \rangle_A = a_{ji}$$

So we have symmetricity if and only if A itself is symmetric. Then the last condition we need is positive definiteness: that for any $x \neq 0$, we have $x^T A x > 0$. Since there is not an easy condition that we already know that is equivalent to this, we simply define:

Definition 7.21

A real square matrix A is called **positive definite** when it is symmetric and $x^T A x = \langle x, x \rangle_A > 0$ for any $x \neq 0$.

Note that by the spectral theorem, we automatically know that A is (orthogonally) diagonalizable. Thus, we are able to make claims in terms of its eigenvalues:

Proposition 7.28

A symmetric $n \times n$ real matrix A is positive definite if and only if all its eigenvalues are positive.

Remark

Suppose A, B are "orthogonally similar," such that there exists S orthogonal with $A = S^T BS$. Then we have $\langle x, y\rangle_A = x^T Ay = x^T S^T BSy = (Sx)^T B(Sy) = \langle Sx, Sy\rangle_B$. This means that A and B encode the same inner product, but acting on different orthonormal eigenbases (in the sense that the corresponding coefficients are the same). Specifically, if we take A to act on the standard basis, then B acts on the orthonormal eigenbasis which is given by the columns of S. Then if the columns of S are $\mathcal{B} = \{v_1, \ldots, v_n\}$, we have $\langle v_i, v_j\rangle_B = \langle e_i, e_j\rangle_A$. Alternatively, using change of basis, we can summarize this as

$$\langle x, y\rangle_A = \langle M_{\mathcal{B}}(x), M_{\mathcal{B}}(y)\rangle_B$$

In other words, if $A = M_e(T)$ and $M_{\mathcal{B}}(T)$, with $\mathcal{B}$ orthonormal, then the above equality holds.

Proof. By the spectral theorem, we can write $A = S^T DS$ for S orthogonal and D diagonal. Then $\langle x, x\rangle_A = \langle Sx, Sx\rangle_D$. Letting $\lambda_1, \ldots, \lambda_n$ be the diagonal entries of D (or the eigenvalues of A), we have $\langle Sx, Sx\rangle_D = \lambda_1(Sx)_1^2 + \ldots + \lambda_n(Sx)_n^2$. If all the λ_i are positive, then this is positive for all nonzero Sx, and if any are zero or negative, then we lose the condition. Since S is invertible, this is true of x as well. Thus A is positive definite if and only if all the eigenvalues are positive. □

The conclusion of the above discussion is that there is a direct correspondence between inner products on $\mathbb{R}^n$ and postiive definite $n \times n$ real matrices.

We now look to extend this concept to inner products on abstract finite dimensional vector spaces.

Proposition 7.29

Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis of V. Then there exists a unique inner product $\langle -, -\rangle_{\mathcal{B}}$ on V such that $\mathcal{B}$ is orthonormal.

Proof. We begin with uniqueness. Suppose we have an inner product $\langle -, -\rangle$ such that $\mathcal{B}$ is orthonormal. Then let $x = c_1v_1 + \ldots + c_nv_n$ and $y = d_1v_1 + \ldots + d_nv_n$ be arbitrary. Then by orthonormality,

$$\langle x, y\rangle = c_1d_1 + \ldots + c_nd_n$$

Then since this is the definition, we automatically get uniqueness. For existence, we simply define this to be the inner product $\langle x, y\rangle_{\mathcal{B}} \coloneqq c_1d_1 + \ldots + c_nd_n$. □

Note that this definition is given by changing the identity matrix into the basis $\mathcal{B}$, so that $\langle x, y\rangle_{\mathcal{B}} = M_{\mathcal{B}}(x) \cdot M_{\mathcal{B}}(y)$.

This discussion has suggested that inner products are related to each other up to a change of basis.

Definition 7.22

Two inner product vector spaces $(V, \langle -, - \rangle_V), (W, \langle -, - \rangle_W)$ are **isomorphic** in the category of inner product spaces if there exists an invertible isometry between them.

Proposition 7.30

Any real finite dimensional inner product space is isomorphic to $(\mathbb{R}^n, \cdot)$

Proof. Follows from the above proposition. □

Thus we have seen that inner products are related to matrices, and also to bases. Thus, we are led to ask the question of how matrices and bases are related through inner products.

Using the observations we have made so far, we make the identification

$$S = M_{e \to \mathcal{B}} = \begin{bmatrix} | & | & | \\ v_1 & \dots & v_n \\ | & | & | \end{bmatrix}^{-1}$$

such that $\langle x, y \rangle_{\mathcal{B}} = M_{\mathcal{B}}(x) \cdot M_{\mathcal{B}}(y) = Sx \cdot Sy = (Sx)^T Sy = x^T S^T S y = \langle x, y \rangle_{S^T S}$. In the case that S is orthogonal (or that $\mathcal{B}$ is orthonormal), this reduces to the dot product. This results in the important observation that although there is a one to one correspondence between positive definite matrices and inner products, the correspondence between inner products and bases is not one-to-one.

This allows us to easily calculate inner products that make bases orthogonal.

Example 7.12

Given the basis $\{\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}\}$, the change of basis matrix is

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

So the matrix for which they are orthonormal is

$$S^T S = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$$

So the inner product is

$$\langle x, y \rangle_{S^T S} = x_1 y_1 - x_1 y_2 - x_2 y_1 + 2 x_2 y_2$$

Definition 7.23

Let V, W be inner product spaces and $T : V \to W$ be a linear transformation. We say that a transformation $T^* : W \to V$ is an **adjoint** of T if $\langle Tv, w\rangle_W = \langle v, T^*w\rangle_V$ for any $v \in V, w \in W$.

Proposition 7.31

Let V, W be finite dimensional. Every $T : V \to W$ has a unique adjoint T^*.

Proof. Let $v_1, \ldots, v_n$ be an orthonormal basis of V. For any v, we must have

$$T^*w = \langle Tv_1, w\rangle v_1 + \ldots + \langle Tv_n, w\rangle v_n$$

which we also take to be the definition. □

Proposition 7.32

Let $S : \mathbb{R}^m \to \mathbb{R}^n$. Then $M(S^*) = M(S)^T$.

Proof. Note that we have $\langle Sv, w\rangle = \langle v, S^*w\rangle$ for any v, w (where the inner product is the standard dot products). Letting $A = M(S), B = M(S^*)$, we have $Av \cdot w = v \cdot B$. Then setting $v = e_i, w = e_j$, we get $Ae_i \cdot e_j = a_{ji}$ and $e_i \cdot Be_j = b_{ij}$, so we find that $B = A^T$. □

This gives us the notion of transpose for abstract vector spaces.

Definition 7.24

$T : V \to V$ is **self-adjoint** when $T^* = T$.

In the case of $\mathbb{R}^n$, the self-adjoint operators are exactly the symmetric matrices. So being self-adjoint is analogous to symmetricity for abstract spaces.

7.8 Singular Value Decomposition

Example 7.13

Let $A = \begin{bmatrix} -3 & 3 \\ 2 & 2 \end{bmatrix}$. Then

$$A\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 4\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$A\begin{bmatrix} -1 \\ 1 \end{bmatrix} = 6\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

We can graphically track where the vectors are sent to to see what this transformation does geometrically. Specifically, it rotates the vectors, and then scales in different directions (so that circles map to ellipses).

Moreover, a perhaps surprising result is that this behavior occurs for any matrix.

Theorem 7.33

For any $T : V \to U$ between inner product spaces, there exists orthonormal bases $\mathcal{B} = \{v_1, \ldots, v_n\}$ of V and a basis $\mathcal{C} = \{u_1, \ldots, u_m\}$ of U such that $T(v_i) = \sigma_i u_i$ with all $\sigma_i \geq 0$. The σ_i are known as singular values of the matrix.

Note that by convention, we reorder the singular values so that they are in decreasing order. We also note that for any σ_i which are outside the range, we set σ_i to be 0.

This theorem, then, says that every transformation can be viewed as the composition of some isometry and some diagonalizable operator with an orthonormal eigenbasis.

If we then consider what $M_{\mathcal{B}\to\mathcal{C}}(T)$ looks like, we will have $a_{ii} = \sigma_i$, such that the matrix is essentially diagonal (it may not be square, but the main diagonal will have the singular values.

To investigate how we may compute this basis, let us first relax the orthonormality requirement, and also assume invertibility. Suppose we have some invertible transformation $T : V \to V$. Then we if we pick $\mathcal{B}$ to be any basis of V and set $\mathcal{C}$ to be the basis vectors after applying T, then we get

$$M_{\mathcal{B}\to\mathcal{C}}(T) = I_n$$

If we relax the invertibility assumption, we have the following:

1. Let $W = \ker T$. Pick some basis $w_1, \ldots, w_k$ of W.
2. Extend to a basis $\mathcal{B} = v_1, \ldots, v_l, w_1, \ldots, w_k$ of V.
3. Extend the set $T(v_1), \ldots, T(v_l)$ to a basis $\mathcal{C} = T(v_1), \ldots, T(v_l), u_1, \ldots, u_r$ of W.

Then we have a matrix with l 1's on the diagonal and zero everywhere else.

Now in order to ensure the bases $\mathcal{B}, \mathcal{C}$ are orthonormal, we will need to adjust the process.

Definition 7.25

A symmetric real matrix A is **positive semidefinite** if $x^T Ax \geq 0$ for any x, or equivalently if all the eigenvalues are nonnegative.

Proposition 7.34

Let B be an $m \times n$ matrix. Then $B^T B$ is symmetric and positive semidefinite.

Proof. $(B^T B)^T = B^T B$, so $B^T B$ is symmetric. To calculate positive semidefiniteness, we calculate $x^T B^T B x = (Bx)^T Bx = Bx \cdot Bx \geq 0$. □

We can now prove that every matrix has a singular value decomposition.

Proof. Let $T : V \to U$. Then $T^*T : V \to V$ is symmetric, so the spectral theorem says we can pick an orthonormal eigenbasis $\mathcal{B} = \{v_1, \ldots, v_m\}$. Let us also order them such that $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_m \geq 0$ (where ≥ 0 follows from positive semidefiniteness). Suppose the first r eigenvalues are nonzero. Define $w_i = T(v_i)$ for each $1 \leq i \leq r$. Then we claim that $w_1, \ldots, w_r$ are orthogonal with $||w_i|| = \sqrt{\lambda_i}$.

To show this, pick w_i, w_j. Then $\langle w_i, w_j \rangle_U = \langle T(v_i), T(v_j) \rangle_U = \langle v_i, T^* \circ T(v_j) \rangle_V = \lambda_j \langle v_i, v_j \rangle$ which is λ_j when $i = j$ and 0 otherwise. Thus we have orthogonality and we know the lengths of the vectors, which are all nonzero. Then we finish the proof by normalizing each of the w_i and extending to an orthonormal basis of U. In this case, our singular values will be $\sigma_i = \sqrt{\lambda_i}$, $1 \leq i \leq r$, and 0 otherwise. □

Corollary

Any matrix $A \in M_{n \times m}(\mathbb{R})$ can be written as $A = U \Sigma V^T$, where $U \in M_{n \times n}(\mathbb{R})$ is orthogonal, $\Sigma \in M_{n \times m}(\mathbb{R})$ has the singular values along the diagonal and zero elsewhere, and $V \in M_{m \times m}(\mathbb{R})$ is orthogonal.

Proof. Let $T : \mathbb{R}^m \to \mathbb{R}^n$ be L_A, so $A = M_{e \to e}(T)$. Apply singular values decomposition to get bases $\mathcal{B}, \mathcal{C}$. Then let $U = M_{e \to \mathcal{C}}$ and $V_{\mathcal{B} \to e}$, and set $\Sigma = M_{\mathcal{B}, \mathcal{C}}(T)$. □

Remark

If there are r nonzero singular values, then we can set Σ to be $r \times r$ diagonal, U to be $n \times r$ semiorthogonal, and V to be $m \times r$ semiorthogonal, since the zero singular values can be removed without affecting the decomposition.

7.9 Quadratic Forms

In this section, we will discuss the nature of the zero loci of quadratic multivariate polynomials.

Example 7.14

Consider the equation $x^2 + 2xy + 3y^2 = 7$. If we pick the change of basis $a = x - \sqrt{3}y$, and $b = x + \sqrt{3}y$. Then letting $\lambda_i = \frac{1}{2} \pm \frac{\sqrt{3}}{6}$, we get

$$\lambda_1 a^2 + \lambda_2 b^2 = 7$$

which is an ellipse in the (possibly rotated) basis of a, b.

Definition 7.26

A **quadratic form** on $\mathbb{R}$ is a function $q : \mathbb{R}^n \to \mathbb{R}$ such that $q \in \text{span}\{x_i x_j | i, j \leq n\}$.

In other words, quadratic forms are linear combinations of x_i^2 and $x_i x_j$.

Example 7.15

Let $q = x_1^2 + x_2^2$. Then we can write this as $q(x) = x \cdot x = x^T I x$.

Let $q = 4x_1^2 + 9x_2^2$. Then

$$q(x) = x^T \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix} x$$

Because the terms of quadratic forms correspond precisely to terms in the expansion of $\langle x, x \rangle$, we find that we can precisely record any quadratic form as $q(x) = x^T A x = \langle x, x \rangle_A$ for some A, where a_{ii} is the coefficient of x_i^2 in q and a_{ij} is half the coefficient of $x_i x_j$ in q for $i \neq j$ (half because we split $c x_i x_j = c/2(x_i x_j + x_j x_i)$). Note that this implies A is symmetric and thus can be associated with a (semi) inner product.

Theorem 7.35

Let $q = \langle x, x \rangle_A$. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be an orthonormal eigenbasis of A with associated eigenvalues $\lambda_1, \ldots, \lambda_n$. Let $x \in \mathbb{R}^n$ be such that

$$M_{\mathcal{B}}(x) = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Then $q(x) = \lambda_1 c_1^2 + \ldots + \lambda_n c_n^2$.

Proof. We orthogonally diagonalize A as SDS^T, where

$$S = \begin{bmatrix} | & | & | \\ v_1 & \dots & v_n \\ | & | & | \end{bmatrix} = M_{e\to\mathcal{B}}$$

Then $M_{\mathcal{B}}(x) = M_{e\to\mathcal{B}}M_e(x) = S^{-1}x = S^Tx$. Then we have

$$q(x) = \langle x, x\rangle_A = x^TAx = x^TS^TDSx = (S^Tx)^TDS^Tx = \langle S^Tx, S^Tx\rangle_D = \lambda_1c_1^2+\ldots+\lambda_nc_n^2$$

□

Then if we have an equation $q = a$, then we have some conic section. Specifically,

- $\lambda_1, \lambda_2 > 0$ gives an ellipse.
- $\lambda_1 > 0, \lambda_2 < 0$ gives a hyperbola.
- $\lambda_1 = 0$ gives two lines.

7.10 Jordan Canonical Form

So far, we have obtained important results about which matrices may be diagonalized. However, one important observation we made was that matrices without enough eigenvalues may not be diagonalized. Using the Jordan canonical form, we can derive an alternate form which has many properties similar to diagonalization, which can be applied to many more types of matrices.

First recall that the minimal polynomial for a square matrix A over $\mathbb{F}$ is the monic polynomial $\mu(x) \in \mathbb{F}[x]$ of lowest degree such that $\mu(A) = 0$. We proved uniqueness by showing that any two minimal polynomials must divide each other and be monic, and we proved existence by consider the linearly dependent list $I, A, A^2 \ldots, A^{n^2}$. Moreover, if the degree of minimal polynomial is d, then d is the smallest integer such that the list $I_n, A, \ldots, A^d$ is linearly dependent.

Example 7.16

The minimal polynomial of $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ is $x - 2$, since $A - 2I = 0$.

The minimal polynomial of $\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$ is $(x - 2)(x - 3)$.

In general, if we consider a block matrices, we get

$$\begin{bmatrix} A & O \\ O & B \end{bmatrix}\begin{bmatrix} C & O \\ O & D \end{bmatrix} = \begin{bmatrix} AC & O \\ O & BD \end{bmatrix}$$

Thus, the minimal polynomial of

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

is $(x-2)(x-3)$.

More generally, a diagonal matrix of the form

$$\begin{bmatrix} \lambda & & \\ & \ddots & \\ & & \lambda \end{bmatrix}$$

is $(x-\lambda)$, and if the values are all distinct, then the minimal polynomial of

$$\begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

is $(x-\lambda_1)\dots(x-\lambda_n)$.

For nondiagonalizable matrices, such as

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

we can calculate $A^2 = 0$, so x^2 is the minimal polynomial.

Similarly, the minimal polynomial of

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

is x^3, since $A^3 = 0$.

If we modify this by adding diagonal terms, such that

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

then $B - \lambda I = A$, so $(B-\lambda I)^3 = A^3 = 0$ and thus $(x-\lambda)^3$ is the minimal polynomial.

Let us now consider an arbitrary matrix of the form $J_n(\lambda)$, which has λ along the diagonal, 1 above the diagonal, and 0 elsewhere:

$$\begin{bmatrix} \lambda & 1 & \dots & 0 \\ & \lambda & \ddots & 0 \\ & & \ddots & 1 \\ & & & \lambda \end{bmatrix}$$

We refer to this as a Jordan block. Then in this case, the minimal polynomial of $J_n(\lambda) = (x-\lambda)^n$. Moreover, this is also its characteristic polynomial.

Now using what we have learned about minimal polynomials of block matrices, we can see that by combining multiple block matrices, we get

$$\mu\left(\begin{bmatrix} J_m(\lambda_1) & O \\ O & J_k(\lambda_2) \end{bmatrix}\right) = (x-\lambda_1)^m(x-\lambda_2)^k$$

so long as $\lambda_1 \neq \lambda_2$, and we can extend this to further eigenvalues.

The major result of Jordan canonical forms is that *every* complex matrix is similar to a matrix which is composed of Jordan blocks, and moreover that this matrix is unique (up to a conventional ordering).

Theorem 7.36

Let A be an $n \times n$ complex square matrix. Then A is similar to a matrix J of the form

$$J = \begin{bmatrix} J_{n_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{n_k}(\lambda_k) \end{bmatrix}$$

where $\lambda_1, \ldots, \lambda_k$ are all the complex eigenvalues of A, possibly repeated, and $n_1 + \ldots + n_k = n$. Moreover, this matrix is canonical in the sense that the pairs (n_i, λ_i) are uniquely determined by A up to ordering.

Diagonal matrices, then, are a special case of the Jordan canonical form where each Jordan block has size 1. In this way, we see that the Jordan canonical form gives us a generalization of diagonalization which holds for every complex matrix.

Corollary

The characteristic polynomial of J is given by

$$P_J = (x-\lambda_1)^{n_1} \ldots (x-\lambda_k)^{n_k}$$

Incidentally, this allows us to prove the Cayley-Hamilton theorem:

Theorem

If A is square over $\mathbb{F}$ and $p_A(x)$ is the characteristic polynomial of A, then $p_A(A) = 0$.

Proof. We write $A = SJS^{-1}$, so $p_A = p_J$. Plugging A into p_J, we get $p_J(A) = p_J(SJS^{-1}) = Sp_J(J)S^{-1}$. We also have $p_J(J) = 0$ since this works for all Jordan blocks and an induction argument shows it works for block matrices of Jordan blocks. □

Note that for any Jordan block $J_n(\lambda)$, its characteristic polynomial is $p_{J_n(\lambda)}(x) = (x-\lambda)^n$. Thus, almu$(\lambda) = n$. Moreover, note that $J_n(\lambda) - \lambda I$ has $n-1$ linearly independent columns, so this dimension of its kernel is 1 and gemu$(\lambda) = 1$.

Thus we see that for a matrix in Jordan form, we can directly read off the algebraic and geometric multiplicities. Suppose that λ is an eigenvalue of A. Then gemu(λ) is the number of Jordan blocks with eigenvalue λ. Moreover, almu(λ) is the sum of the sizes of those blocks.

This also helps us calculate the minimal polynomial of a Jordan canonical form. Supposing that an eigenvalue has two blocks, $J_m(\lambda)$ and $J_k(\lambda)$, then $J_m(\lambda)$ will go to 0 after m exponents, and $J_k(\lambda)$ goes to 0 after k exponents. So the exponent of $(x-\lambda)$ in the minimal polynomial is the maximum size of a Jordan block corresponding to λ, given by $\max\{n_k : \lambda_k = \lambda\}$.

Thus we have seen that the Jordan canonical form is especially powerful because it allows us to directly read off many invariants of a matrix, including its eigenvalues with algebraic and geometric multiplicity, its characteristic polynomial, and its minimal polynomial.

We will now discuss how to calculate the Jordan canonical form.

1. Separate the different eigenvalues of A.
2. Consider the special case for each where $\mu(x) = (x-\lambda)^d$. Then the only eigenvalue of $A - \lambda I$ is λ, and it is nilpotent. Thus we have reduced the problem to classifying nilpotent matrices.

To accomplish this, we will consider these matrices from a geometric perspective. Suppose that $T : V \to V$ has an eigenvalue λ. Then we have previously defined the eigenspace of λ to be

$$E_\lambda \coloneqq \{v \in V : (T - \lambda I)v = 0\}$$

We expand the eigenspace as follows:

Definition 7.27

Let $T : V \to V$ have an eigenvalue λ. Then the **generalized eigenspace** of λ in T is

$$G_\lambda \coloneqq \{v \in V : \exists k \in \mathbb{N} s.t. (T - \lambda I)^k v = 0\}$$

where the corresponding vectors v are called **generalized eigenvectors**.

Lemma

G_λ is a T-invariant subspace.

Proof. L let $v_1, v_2 \in G_\lambda$, $t \in \mathbb{F}$, so that $(T-\lambda I)^{k_1} v_1 = (T-\lambda I)^{k_2} v_2 = 0$. Then $(T-\lambda)^{k_1+k_2}(v_1 + tv_2) = 0 + t^{k_1+k_2} 0 = 0$, so it is closed under addition and homogeneity. Moreover, for any $v \in G_\lambda$, if $(T-\lambda I)^k v = 0$, then $(T-\lambda)^k Tv = T(T-\lambda)^k v = Tv = 0$ so $Tv \in G_\lambda$. Note that we can do this switch because T commutes with T^k and with scalars. Thus G_λ is a T-invariant subspace. □

Proposition 7.37

If $\lambda_1 \neq \lambda_2$, then $G_{\lambda_1} \cap G_{\lambda_2} = \{0\}$.

Proof. Let $v \in G_{\lambda_1} \cap G_{\lambda_2}$. Suppose $v \neq 0$. Let $k_1 > 0$ be the smallest integer such that $(T-\lambda_1 I)^{k_1} v = 0$. Then $(T-\lambda_1 I)^{k_1-1} v \neq 0$. Let $w = (T-\lambda_1)^{k_1-1}$. Then $(T-\lambda_1)w = 0$. So $Tw = \lambda_1 w$ and thus $w \in G_{\lambda_1}$. We can expand w into a linear combination of v and repeated applications of T to v. Since G_{λ_2} is T-invariant, $w \in G_{\lambda_2}$. Then we can find a value k_2 such that $(T-\lambda_2)^{k_2} w = 0$, then this implies that $(\lambda_1 - \lambda_2)^{k_2} w = 0$, which is not true unless $w = 0$. But we chose k_1 so $w \neq 0$, so this is impossible. Thus we must have $v = 0$ and thus $G_{\lambda_1} \cap G_{\lambda_2} = \{0\}$. □

Proposition 7.38

Suppose that $\lambda_1, \ldots, \lambda_r$ are distinct eigenvalues. Then for any choice of nonzero generalized eigenvectors $v_i \in G_{\lambda_i}$, the v_i are linearly independent.

Proof. Induction on r using the previous proposition. □

Theorem 7.39

Let $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of $T : V \to V$ with V finite dimensional over $\mathbb{C}$. Then $V = G_{\lambda_1} \oplus \ldots \oplus G_{\lambda_k}$. Equivalently, for any $v \in V$, v has a unique decomposition into a sum of generalized eigenvectors given by $v = v_1 + \ldots + v_r$, $v_i \in G_{\lambda_i}$.

Of course, in the diagonalizable case, we already knew this was possible using eigenvectors and eigenspaces. In order to extend to the general case, we need to use generalized eigenvectors and generalized eigenspaces.

Thus we see that we have reduced our space into a direct sum of generalized eigenspaces, each with only eigenvalue, and thus we can consider them individually.

Proposition 7.40: Bezout's Identity

Let $f(x), g(x) \in \mathbb{C}[x]$ have no common roots. Then there exist $p(x), q(x) \in \mathbb{C}[x]$ such that $pf + qg = 1$ identically. This result holds over arbitrary fields by changing the "no common roots" condition to "no common irreducible factors."

Example 7.17

If $f = x^2, g = x - 1$, then $p = 1$ and $q = (-x - 1)$ gives $pf + qg = 1$.

Recall that two integers $m, n \in \mathbb{Z}$ are coprime if there exist $a, b \in \mathbb{Z}$ such that $ma + nb = 1$. Thus, we can consider the "no common roots" or "no common irreducible factors" to be the equivalent notion of being coprime for polynomials.

An analogous fact holds for more polynomials: Suppose that $f_1(x), \ldots, f_r(x) \in \mathbb{C}[x]$ such that there is no value which is a root of every polynomial (there may be pairwise roots or higher, but no roots of all polynomials). Then there exists $p_1, \ldots, p_r \in \mathbb{C}[x]$ such that $f_1 p_1 + \ldots + f_r p_r = 1$.

This allows us to prove Theorem 7.39.

Proof of Theorem 7.39. Let $\mu(x)$ be the minimal polynomial of T. Then $\mu(x) = (x - \lambda_1)^{n_1} \ldots (x - \lambda_r)^{n_r}$ where the λ_i are distinct.

Define μ_i to be $\frac{\mu(x)}{(x-\lambda_i)^{n_i}}$, that is, the minimal polynomial without the $(x-\lambda_i)$ factors. Then $\mu_1, \ldots, \mu_r$ have no common roots. Then by Bezout's identity, we have $\mu_1 p_1 + \ldots + \mu_r p_r = 1$ for appropriate polynomials. I claim that $\mu_i(T) p_i(T) v \in G_{\lambda_i}$ for any $v \in V$. To prove this, note that $(T - \lambda_i)^{n_i} \mu_i(T) = \mu(T) = 0$. Moreover, these terms sum to v since the sum of the polynomials is 1. So we can decompose v into generalized eigenvectors. □

Thus, we can now assume that $T : V \to V$ has only one eigenvalue λ. In this case, the minimal polynomial must be $\mu(x) = (x - \lambda)^k$, so T is nilpotent.

Definition 7.28

A transformation $T : V \to V$ is **nilpotent** if $T^k = 0$ for some k. A square matrix A is nilpotent if $A^k = 0$ for some k. Without loss of generality, we may take $k = n$.

Proposition 7.41

For $T : V \to V$, if $v \in V$ has k such that $T^k v = 0$ and $T^{k-1} v \neq 0$, then $v, Tv, \ldots, T^{k-1} v$ are linearly independent.

Proof. Take a linear combination and recursively apply T to conclude the combination is trivial. □

Proposition 7.42

If $T : V \to V$ is nilpotent, then there exists $v_1, \dots, v_k$ such that $T^{k_1-1}v_1, \dots, Tv_1, v_1, \dots, T^{k_k-1}v_k, \dots, v_k$ form a basis $\mathcal{B}$ of V for appropriate integers $k_1, \dots, k_k > 0$.

Proof. We induct on $\dim V$. Let $W = \operatorname{im}(T)$. W is T-invariant, and because T is nilpotent, $W \neq V$. Then we use induction ot pick a basis of w given by $w_1, Tw_1, \dots, T^{k_1-1}w_1, \dots, w_r, Tw_r, T^{k_r-1}w_r$. We extend each chain by writing $w_1 = Tv_1$, and thus we claim that $v_1, Tv_1, \dots, T^{k_1}v_1, \dots, v_r, Tv_r, \dots, T^{k_r}v_r$ is linearly independent. The proof is similar to the proof for the previous proposition.

We now extend to a basis of V by adding $u_1, \dots, u_r$. In order for the inductive hypothesis to work, we want $u_1, \dots, u_r$ to be in $\ker T$, but we also need them to not overlap with the vectors we have already chosen. Let W' be the span of the vectors we have chosen so far. Then $T(u_i) \in \operatorname{im} T = W = T(w')$ so there is some $w_i' \in W'$ such that $T(w_i') = T(u_i)$. Then replace u_i with $u_i - w_i'$ such that $T(u_i - w_i') = 0$. Lastly, we verify that this final list is still linearly independent. Since it is of length $\dim V$, it is a basis. We can also check that it is in the form requested. □

Corollary

The matrix of T under the above basis $\mathcal{B}$ is the Jordan canonical form of $M_{\mathcal{B}}(T)$:

$$M_{\mathcal{B}}(T) = \begin{bmatrix} J_{k_1}(0) & & & \\ & J_{k_2}(0) & & \\ & & \ddots & \\ & & & J_{k_k}(0) \end{bmatrix}$$

Then in the general case, if T has only one eigenvalue, then $T - \lambda I$ is nilpotent, so

$$T - \lambda I \sim \begin{bmatrix} J_{n_1}(0) & & \\ & \ddots & \\ & & J_{n_k}(0) \end{bmatrix} \implies T \sim \begin{bmatrix} J_{n_1}(\lambda) & & \\ & \ddots & \\ & & J_{n_k}(\lambda) \end{bmatrix}$$

If T has multiple eigenvalues, then V is a direct sum of the generalized eigenspaces, so we consider T on each generalized eigenspace and use the direct sum of the derived matrices.

This proves the existence of the Jordan canonical form. In arbitrary fields,

this form only exists for matrices whose characteristic polynomials may be split into linear factors over the field (which follows automatically for algebraically complete fields such as $\mathbb{C}$).