“Advanced Linear Algebra”

Math 462 - Fall 2009 - Section 15224

California State Univeristy Northridge

This version was LaTeXed on December 17, 2009

This is a Rough Draft and Contains Errors!

This is an incomplete working document that will be revised many times over the semester. Save a tree and just print what you need - there will be lots of corrections, additions, updates, and changes, and probably pretty frequently!

Note to Students

Don't try to use this as a replacement for your textbook. These are notes and just provide an outline of the subject material, not a complete presentation. I have provided a copy to you to use only as a study aid. Their real purpose is to remind me what to talk about about during my class lectures. They are loosely based on the textbook *Linear Algebra Done Right* by Sheldon Axler, and contain some material from other sources as well, but the presentation in the textbook is more thorough. You should read the textbook in preparation for class, and just use these notes to aide your own note-taking during class.

Some Legal Stuff

Please report any errors to bruce.e.shapiro@csun.edu. All feedback, comments, suggestions for improvement, etc., is appreciated, especially if you've used these notes for a class, either at CSUN or elsewhere, from both instructors and students.

Contents

Please remember that this is a draft document, and so it is incomplete and buggy. More sections will be added as the course progresses, the ordering of topics is subject to change, and errors will be corrected as I become aware of them. This version was last LaTeXed on December 17, 2009.

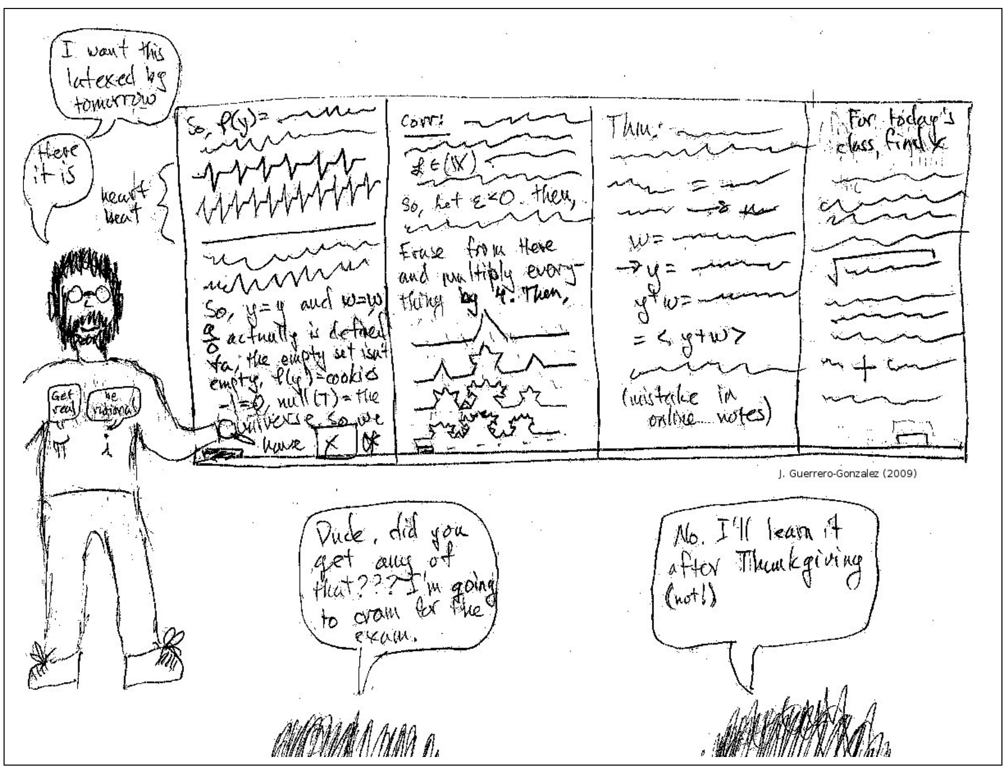
A characteristic (typical) Linear Algebra teacher.

Symbols Used

$(v_1, \ldots, v_n)$	List containing $v_1, \ldots, v_n$
$(v_1, \ldots, v_n \| p_1, \ldots, p_m)$	$(v_1, \ldots)$ with $(p_1, \ldots)$ removed
$\langle v, w \rangle$	inner product of two vectors
$\\|v\\|$	norm of a vector
$\mathbb{V} \oplus \mathbb{W}$	Direct Sum of $\mathbb{V}$ and $\mathbb{W}$
$\mathbb{C}$	Field of Complex Numbers
$\deg(p)$	Degree of polynomial p
$\det(A)$	Determinant of A
$\text{diagonal}(x_1, \ldots, x_n)$	Diagonal matrix
$\dim(\mathbb{V})$	Dimension of a Vector Space
$\mathbb{F}$	Either $\mathbb{R}$ or $\mathbb{C}$
$\mathbb{F}^n$	Set of tuples $(x_1, \ldots, x_n)$, $x_i \in \mathbb{F}$
$\mathbb{F}^\infty$	Set of sequences over $\mathbb{F}$
i	$\sqrt{-1}$
$\text{length}(B)$	Length of a list B
$\mathcal{L}(\mathbb{V})$	Set of linear operators $T : \mathbb{V} \mapsto \mathbb{V}$
$\mathcal{L}(\mathbb{V}, \mathbb{W})$	Set of linear maps $T : \mathbb{V} \mapsto \mathbb{W}$
$\mathcal{M}(T, B)$	Matrix of a linear map T
	with respect to a basis B

$\mathbb{P}(\mathbb{F})$	Set of polynomials over $\mathbb{F}$
$\mathbb{P}_m(\mathbb{F})$	Polynomials of degree m
$\mathbb{R}$	Field of Real Numbers
$\text{span}(v_1, \dots, v_n)$	Span of a list of vectors
$\mathbb{U}$, $\mathbb{V}$, $\mathbb{W}$	Vector Spaces
$\mathbb{V}(\mathbb{F})$	Vector Space over $\mathbb{F}$
z^* or $\overline{z}$	Complex conjugate of z (scalar or vector)
T^*	Adjoint of T

Working across the chalk board from right to left, an intrepid linear algebra teacher enters a fractal dimensioned vector space in search of the elusive x, braving Koch curves, PQ waves, and buggy lecture notes, demonstrating Bruce's law of learning: No learning takes place after Thanksgiving.[1]



[1]The American holiday of Thanksgiving, (a national celebration of televised football), is the 4th Thursday of November. It provides many college students with a 5-day weekend (the day after Thanksgiving, Black Friday, is the national day of shopping, and pretty much everyone cuts classes the day before). Final exams begin about a week or two later.The first week of December thereby degenerates into an educational singularity as instructors vainly try to make up for lost time.

Topic 1

Review of Elementary Linear Algebra

In this section we will recall a few of the concepts you should have learned in your elementary linear algebra class and calculus courses. This material is typically covered in Math 262, 150A/B and 250 at CSUN. If anything in this section is new to you, you should review your old textbooks and notes. This material is NOT covered in our textbook

Definition 1.1 A **Euclidean 3-vector v** is object with a **magnitude** and **direction** which we will denote by the ordered triple

$$\mathbf{v} = (x, y, z) \tag{1.1}$$

The **magnitude** or **absolute value** or **length** of the v is denoted by the postitive square root

$$v = |\mathbf{v}| = \sqrt{x^2 + y^2 + z^2} \tag{1.2}$$

This definition is motivated by the fact that v is the length of the line segment from the origin to the point $P = (x, y, z)$ in Euclidean 3-space.

A vector is sometimes represented geometrically by an arrow from the origin to the point $P = (x, y, z)$, and we will sometimes use the notation (x, y, z) to refer either to the point P or the vector $\mathbf{v}$ from the origin to the point P. Usually it will be clear from the context which we mean. This works because of the following theorem.

Definition 1.2 The set of all Euclidean 3-vectors is isomorphic to the Euclidean 3-space (which we typically refer to as $\mathbb{R}^3$).

If you are unfamiliar with the term *isomorphic*, don't worry about it; just take it to mean "in one-to-one correspondence with," and that will be sufficient for our purposes.

Definition 1.3 Let $\mathbf{v} = (x, y, z)$ and $\mathbf{w} = (x', y', z')$ be Euclidean 3-vectors. Then the **angle between v and w** is defined as the angle between the line segments joining the origin and the points $P = (x, y, z)$ and $P' = (x', y', z')$.

We can define **vector addition** or **vector subtraction** by

$$\mathbf{v} + \mathbf{w} = (x, y, z) + (x', y', z') = (x + x', y + y', z + z') \tag{1.3}$$

where $\mathbf{v} = (x, y, z)$ and $\mathbf{w} = (x', y', z')$, and **scalar multiplcation** (multiplication by a real number) by

$$k\mathbf{v} = (kx, ky, kz) \tag{1.4}$$

Theorem 1.4 The set of all Euclidean vectors is closed under vector addition and scalar multiplication.

Definition 1.5 Let $\mathbf{v} = (x, y, z), \mathbf{w} = (x', y', z')$ be Euclidean 3-vectors. Their **dot**

product is defined as

$$\mathbf{v} \cdot \mathbf{w} = xx' + yy' + zz' \tag{1.5}$$

Theorem 1.6 Let θ be the angle between the line segments from the origin to the points (x, y, z) and (x', y', z') in Euclidean 3-space. Then

$$\mathbf{v} \cdot \mathbf{w} = |\mathbf{v}||\mathbf{w}| \cos \theta \tag{1.6}$$

Definition 1.7 The **standard basis vectors** for Euclidean 3-space are the vectors

$$\mathbf{i} = (1, 0, 0) \tag{1.7}$$

$$\mathbf{j} = (0, 1, 0) \tag{1.8}$$

$$\mathbf{k} = (0, 0, 1) \tag{1.9}$$

Theorem 1.8 Let $\mathbf{v} = (x, y, z)$ be any Euclidean 3-vector. Then

$$\mathbf{v} = \mathbf{i}x + \mathbf{j}y + \mathbf{k}z \tag{1.10}$$

Definition 1.9 The vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are said to be **linearly dependent** if there exist numbers $a_1, a_2, \dots, a_n$, not all zero, such that

$$a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_n\mathbf{v}_n = 0 \tag{1.11}$$

If no such numbers exist the vectors are said to be **linearly independent**.

Definition 1.10 An $m \times n$ (or m by n) **matrix** A is a rectangular array of number with m rows and n columns. We will denote the number in the i^{th} row and j^{th} column as a_{ij}

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \tag{1.12}$$

We will sometimes denote the matrix A by $[a_{ij}]$.

The **transpose** of the matrix A is the matrix obtained by interchanging the row and column indices,

$$(A^{\mathrm{T}})_{ij} = a_{ji} \tag{1.13}$$

or

$$[a_{ij}]^{\mathrm{T}} = [a_{ji}] \tag{1.14}$$

The transpose of an $m \times n$ matrix is an $n \times m$ matrix. We will sometimes represent the vector $\mathbf{v} = (x, y, z)$ by its 3×1 **column-vector** representation

$$\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{1.15}$$

or its 1×3 **row-vector** representation

$$\mathbf{v}^{\mathrm{T}} = \begin{pmatrix} x & y & z \end{pmatrix} \tag{1.16}$$

Definition 1.11 Matrix Addition is defined between two matrices of the same size, by adding corresponding elemnts.

$$\begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & & \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots \\ b_{21} & b_{22} & \cdots \\ \vdots & & \end{pmatrix} = \begin{pmatrix} a_{11}+b_{11} & b_{22}+b_{12} & \cdots \\ a_{21}+b_{21} & a_{22}+b_{22} & \cdots \\ \vdots & & \end{pmatrix} \tag{1.17}$$

Matrices that have different sizes cannot be added.

Definition 1.12 A **square matrix** is any matrix with the same number of rows as columns. The **order** of the square matrix is the number of rows (or columns).

Definition 1.13 Let A be a square matrix. A **submatrix** of A is the matrix A with one (or more) rows and/or one (or more) columns deleted.

Definition 1.14 The **determinant of a square matrix** is defined as follows. Let A be a square matrix and let n be the order of A. Then

1. If $n = 1$ then $A = [a]$ and $\det A = a$.

2. If $n \geq 2$ then

$$\det A = \sum_{i=1}^{n} a_{ki}(-1)^{i+k} \det(A'_{ik}) \tag{1.18}$$

for any $k = 1, .., n$, where by A'_{ik} we mean the submatrix of A with the i^{th} row and k^{th} column deleted. (The choice of which k does not matter because the result will be the same.)

We denote the determinant by the notation

$$detA = \begin{vmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & & \end{vmatrix} \tag{1.19}$$

In particular,

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc \tag{1.20}$$

and

$$\begin{vmatrix} A & B & C \\ D & E & F \\ G & H & I \end{vmatrix} = A \begin{vmatrix} E & F \\ H & I \end{vmatrix} - B \begin{vmatrix} D & F \\ G & I \end{vmatrix} + C \begin{vmatrix} D & E \\ G & H \end{vmatrix} \tag{1.21}$$

Definition 1.15 Let $v = (x, y, z)$ and $w = (x', y', z')$ be Euclidean 3-vectors. Their **cross product** is

$$\mathbf{v} \times \mathbf{w} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ x & y & z \\ x' & y' & z \end{vmatrix} = (yz' - y'z)\mathbf{i} - (xz' - x'z)\mathbf{j} + (xy' - x'y)\mathbf{k} \tag{1.22}$$

Theorem 1.16 Let $v = (x, y, z)$ and $w = (x', y', z')$ be Euclidean 3-vectors, and let θ be the angle between them. Then

$$|\mathbf{v} \times \mathbf{v}| = |\mathbf{v}||\mathbf{w}| \sin \theta \tag{1.23}$$

Definition 1.17 A square matrix A is said to be **singular** if $\det A = 0$, and **non-singular** if $\det A \neq 0$.

Theorem 1.18 The n columns (or rows) of an $n \times n$ square matrix A are linearly independent if and only if $\det A \neq 0$.

Definition 1.19 Matrix Multiplication. Let $A = [a_{ij}]$ be an $m \times r$ matrix and let

$B = [b_{ij}]$ be an $r \times n$ matrix. Then the matrix product is defined by

$$[AB]_{ij} = \sum_{k=1}^{r} a_{ik} b_{kr} = \text{row}_i(A) \cdot \text{column}_j B \tag{1.24}$$

i.e., the ij^{th} element of the product is the dot product between the i^{th} row of A and the j^{th} column of B.

Example 1.1

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 8 & 9 \\ 10 & 11 \\ 12 & 13 \end{pmatrix} = \begin{pmatrix} (1,2,3)\cdot(8,10,12) & (1,2,3)\cdot(9,11,13) \\ (4,5,6)\cdot(8,10,12) & (4,5,6)\cdot(9,11,13) \end{pmatrix} \tag{1.25}$$

$$= \begin{pmatrix} 64 & 70 \\ 156 & 169 \end{pmatrix} \tag{1.26}$$

Note that the product of an $[n \times r]$ matrix and an $[r \times m]$ matrix is always an $[n \times m]$ matrix. The product of an $[n \times r]$ matrix and and $[s \times n]$ is undefined unless $r = s$.

Theorem 1.20 If A and B are both $n \times n$ square matrices then

$$\det AB = (\det A)(\det B) \tag{1.27}$$

Definition 1.21 Identity Matrix. The $n \times n$ matrix I is defined as the matrix with 1's in the **main diagonal** $a_{11}, a_{22}, \dots, a_{mm}$ and zeroes everywhere else.

Theorem 1.22 I is the identity under matrix multiplication. Let A be any $n \times n$ matrix and I the $n \times n$ Identity matrix. Then $AI = IA = A$.

Definition 1.23 A square matrix A is said to be **invertible** if there exists a matrix

A^{-1}, called the **inverse** of A, such that

$$AA^{-1} = A^{-1}A = I \tag{1.28}$$

Theorem 1.24 A square matrix is invertible if and only if it is nonsingular, i.,e, $\det A \neq 0$.

Definition 1.25 Let $A = [a_{ij}]$ be any square matrix of order n. Then the **cofactor** of a_{ij}, denoted by cof a_{ij}, is the $(-1)^{i+j} \det M_{ij}$ where M_{ij} is the submatrix of A with row i and column j removed.

Example 1.2 Let

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \tag{1.29}$$

Then

$$\text{cof } a_{12} = (-1)^{1+2} \begin{vmatrix} 4 & 6 \\ 7 & 9 \end{vmatrix} = (-1)(36 - 42) = 6 \tag{1.30}$$

Definition 1.26 Let A be a square matrix of order n. The **Clasical Adjoint** of A, denoted adj A, is the transopose of the matrix that results when every element of A is replaced by its cofactor.

Example 1.3 Let

$$A = \begin{pmatrix} 1 & 0 & 3 \\ 4 & 5 & 0 \\ 0 & 3 & 1 \end{pmatrix} \tag{1.31}$$

The classical adjoint is

$$\text{adj } A = \text{Transpose} \begin{pmatrix} (1)[(1)(5)-(0)(3)] & (-1)[(4)(1)-(0)(0)] & (1)[(4)(3)-(5)(0)] \\ (-1)[(0)(1)-(3)(3)] & (1)[(1)(1)-(3)(0)] & (-1)[(1)(3)-(0)(0)] \\ (1)[(0)(0)-(3)(5)] & (-1)[(1)(0)-(3)(4)] & (1)[(1)(5)-(0)(4)] \end{pmatrix} \tag{1.32}$$

$$= \text{Transpose} \begin{pmatrix} 5 & -4 & 12 \\ 9 & 1 & -3 \\ -15 & 12 & 5 \end{pmatrix} = \begin{pmatrix} 5 & 9 & -15 \\ -4 & 1 & 12 \\ 12 & -3 & 5 \end{pmatrix} \tag{1.33}$$

Theorem 1.27 Let A be a non-singular square matrix. Then

$$A^{-1} = \frac{1}{\det A} \text{adj } A \tag{1.34}$$

Example 1.4 Let A be the square matrix defined in equation 1.31. Then

$$\det A = 1(5-0) - 0 + 3(12-0) = 41 \tag{1.35}$$

Hence

$$A^{-1} = \frac{1}{41} \begin{pmatrix} 5 & 9 & -15 \\ -4 & 1 & 12 \\ 12 & -3 & 5 \end{pmatrix} \tag{1.36}$$

In practical terms, computation of the determinant is computationally inefficient, and there are faster ways to calculate the inverse, such as via Gaussian Elimination. In fact, determinants and matrix inverses are very rarely used computationally because there is almost always a better way to solve the problem, where by better we mean the total number of computations as measure by number of required multiplications and additions.

Definition 1.28 Let A be a square matrix. Then the **eigenvalues** of A are the numbers λ and **eigenvectors** $\mathbf{v}$ such that

$$A\mathbf{v} = \lambda\mathbf{v} \tag{1.37}$$

Definition 1.29 The **characteristic equation** of a square matrix of order n is the n^{th} order (or possibly lower order) polynomial

$$\det(A - \lambda I) = 0 \tag{1.38}$$

Example 1.5 Let A be the square matrix defined in equation 1.31. Then its characteristic equation is

$$0 = \begin{vmatrix} 1-\lambda & 0 & 3 \\ 4 & 5-\lambda & 0 \\ 0 & 3 & 1-\lambda \end{vmatrix} \tag{1.39}$$

$$= (1-\lambda)(5-\lambda)(1-\lambda) - 0 + 3(4)(3) \tag{1.40}$$

$$= 41 - 11\lambda + 7\lambda^2 - \lambda^3 \tag{1.41}$$

Theorem 1.30 The eigenvalues of a square matrix A are the roots of its characteristic polynomial.

Example 1.6 Let A be the square matrix defined in equation 1.31. Then its eigenvalues are the roots of the cubic equation

$$41 - 11\lambda + 7\lambda^2 - \lambda^3 = 0 \tag{1.42}$$

The only real root of this equation is approximately $\lambda \approx 6.28761$. There are two additional complex roots, $\lambda \approx 0.356196 - 2.52861i$ and $\lambda \approx 0.356196 + 2.52861i$.

Example 1.7 Let

$$A = \begin{pmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix} \tag{1.43}$$

Its characteristic equation is

$$0 = \begin{vmatrix} 2-\lambda & -2 & 3 \\ 1 & 1-\lambda & 1 \\ 1 & 3 & -1-\lambda \end{vmatrix} \tag{1.44}$$

$$= (2-\lambda)[(1-\lambda)(-1-\lambda) - 3] + 2[(-1-\lambda) - 1] \tag{1.45}$$

$$+ 3[3 - (1-\lambda)] \tag{1.46}$$

$$= (2-\lambda)(-1+\lambda^2 - 3) + 2(-2-\lambda) + 3(2+\lambda) \tag{1.47}$$

$$= (2-\lambda)(\lambda^2 - 4) - 2(\lambda+2) + 3(\lambda+2) \tag{1.48}$$

$$= (2-\lambda)(\lambda+2)(\lambda-2) + (\lambda+2) \tag{1.49}$$

$$= (\lambda+2)[(2-\lambda)(\lambda-2) + 1] \tag{1.50}$$

$$= (\lambda+2)(-\lambda^2 + 4\lambda - 3) \tag{1.51}$$

$$= -(\lambda+2)(\lambda^2 - 4\lambda + 3) \tag{1.52}$$

$$= -(\lambda+2)(\lambda-3)(\lambda-1) \tag{1.53}$$

Therefore the eigenvalues are -2, 3, 1. To find the eigenvector corresponding to -2 we would solve the system of

$$\begin{pmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = -2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{1.54}$$

for x, y, z. One way to do this is to multiply out the matrix on the left and solve the system of three equations in three unknowns:

$$2x - 2y + 3z = -2x \tag{1.55}$$

$$x + y + z = -2y \tag{1.56}$$

$$x + 3y - z = -2z \tag{1.57}$$

However, we should observe that the eigenvector is never unique. For example, if v is an eigenvector of A with eigenvalue λ then

$$A(k\mathbf{v}) = kA\mathbf{v} = k\lambda\mathbf{v} \tag{1.58}$$

i.e., $k\mathbf{v}$ is also an eigenvalue of A. So the problem is simplified: we can try to fix one of the elements of the eigenvalue. Say we try to find an eigenvector of A corresponding to $\lambda = -2$ with $y = 1$. Then we solve the system

$$2x - 2 + 3z = -2x \tag{1.59}$$

$$x + 1 + z = -2 \tag{1.60}$$

$$x + 3 - z = -2z \tag{1.61}$$

Simplifying

$$4x - 2 + 3z = 0 \tag{1.62}$$

$$x + 3 + z = 0 \tag{1.63}$$

$$x + 3 + z = 0 \tag{1.64}$$

The second and third equations are now the same because we have fixed one of the values. The remaining two equations give two equations in two unknowns:

$$4x + 3z = 2 \tag{1.65}$$

$$x + z = -3 \tag{1.66}$$

The solution is $x = 11, z = -14$. Therefore an eigenvalue of A corresponding to $\lambda = -2$ is $\mathbf{v} = (11, 1, -14)$, as is any constant multiple of this vector.

Definition 1.31 The **main diagonal** of a square matrix A is the list $(a_{11}, a_{22}, \ldots, a_{nn})$.

Definition 1.32 A **diagonal matrix** is a square matrix that only has non-zero entries on the main diagonal.

Theorem 1.33 The eigenvalues of a diagonal matrix are the elements of the diagonal.

Definition 1.34 An **upper (lower) triangular matrix** is a square matrix that only has nonzero entries on or above (below) the main diagonal.

Theorem 1.35 The eigenvalues of an upper (lower) triangular matrix line on the main diagonal.

Topic 2

Complex Numbers

Definition 2.1 Let $a, b \in \mathbb{R}$. Then a **Complex Number** is an ordered pair

$$z = (a, b) \tag{2.1}$$

with the following properties:

Complex Addition:

$$z + w = (a + c, b + d) \tag{2.2}$$

Complex Multiplication:

$$z \times w = (ac - bd, ad + bc) \tag{2.3}$$

The **set of all complex numbers** is denoted by $\mathbb{C}$. We will sometimes call this set the **complex plane** because we can think of the point $(a, b) \in \mathbb{C}$ as a point plotted with coordinates (a, b) in the Euclidean plane.

The **Real Part** of $z = (a, b)$ is defined by

$$\text{Re } z = \text{Re } (a, b) = a \tag{2.4}$$

and the **Imaginary Part** of $z = (a, b)$ is defined by

$$\text{Im } z = \text{Im } (a, b) = b. \tag{2.5}$$

The **Real Axis** is defined as the set

$$\{z = (x, 0) | x \in \mathbb{R}\} \tag{2.6}$$

We can see that there is a one-to-relationship between the real numbers and the set of complex numbers $(x, 0)$.

The imaginary axis is the set of complex numbers

$$\{z = (0, y) | y \in \mathbb{R}\} \tag{2.7}$$

Equations 2.2 and 2.3 tell us that it is possible to multiply a complex number by a real number in the way we would expect; if $a, b, c \in \mathbb{R}$ then we define

$$c(a, b) = (ca, cb) \tag{2.8}$$

To see why this works, let $u = (x, 0)$ be any point on the real axis. Then

$$uz = (x, 0) \times (a, b) = (ax - 0b, bx - 0a) = (ax, bx) = x(a, b) \tag{2.9}$$

The motivation for equation 2.3 is the following. Suppose $z = (0, 1)$. Then by 2.3,

$$z^2 = (0, 1) \times (0, 1) = (-1, 0) \tag{2.10}$$

We use the special symbol i to represent the complex number $i = (0, 1)$. Then we can write any complex number $z = (a, b)$ as

$$z = (a, b) = (a, 0) + (b, 0) = a(1, 0) + b(0, 1) \tag{2.11}$$

Since $i = (0, 1)$ multiplication by $(1, 0)$ is identical to multiplication by 1 we have

$$z = (a, b) = a + bi \tag{2.12}$$

and hence from 2.10

$$i^2 = -1 \tag{2.13}$$

The common notation is to represent complex numbers as $z = a + bi$ where $a, b \in \mathbb{R}$, where i represents the square root of -1. We will hencforth drop the ordered-pair notation and use the standard notation.

Theorem 2.2 Properties of Complex Numbers

1. Closure The set $\mathbb{C}$ is closed under addition and multiplication.

2. Commutivity

$$\left.\begin{aligned} w + z &= z + w \\ wz &= zw \end{aligned}\right\} \forall w, z \in \mathbb{C} \tag{2.14}$$

3. Associativity

$$\left.\begin{aligned}(u+v)+w &= u+(v+w)\\ (uv)w &= u(vw)\end{aligned}\right\} \forall u,v,w\in\mathbb{C} \tag{2.15}$$

4. Identities

$$\left.\begin{aligned}z+0=0+z=z\\ z1=1z=z\end{aligned}\right\} \forall z\in\mathbb{C} \tag{2.16}$$

5. Additive Inverse. $\forall z\in\mathbb{C},\ \exists$ unique $w\in\mathbb{C}\ \ni\ z+w=0$.

$$z+(-z)=(-z)+z=0 \tag{2.17}$$

6. Multiplicative Ineverse. $\forall z\in\mathbb{C},\ \exists$ unique $w\in\mathbb{C}\ \ni\ zw=wz=1$.

$$z(z^{-1})=(z^{-1})z=1 \text{ or } z(1/z)=(1/z)z=1 \tag{2.18}$$

7. Distributivity

$$u(w+z)=uw+uz\ \forall u,w,z\in\mathbb{C} \tag{2.19}$$

Definition 2.3 A **Field** is a set, together with two operations, that we will call **addition** (or $(+)$) and **multiplication** or $(\cdot)$, that satisfy the properites of closure; associativity and commutativity of both $(+)$ and $(\cdot)$; existence of identities and inverses for both $(+)$ and $(\cdot)$; and distributivity of $(\cdot)$ over $(+)$.

Theorem 2.4 $\mathbb{R}$ is a field.

Theorem 2.5 $\mathbb{C}$ is a field.

Notation. Throughout this course we will use the notation $\mathbb{F}$ to represent a general scalar field (we can take the term scalar to mean "not vector" because, as we will see, there are such things as vector fields as well). We will only be interested in two fields: $\mathbb{R}$ and $\mathbb{C}$. You can think of $\mathbb{F}$ as representating either $\mathbb{R}$ or $\mathbb{C}$; i.e., whenever we make a statement about a field $\mathbb{F}$ we mean that the statement refers to either $\mathbb{R}$ or $\mathbb{C}$.

Definition 2.6 Let $z = a + bi \in \mathbb{C}$. Then the **complex conjugate of** z is defined as

$$z^* = \overline{z} = a - bi \tag{2.20}$$

Note that there are two different but equivalent notations that we will use for the complex conjugate; you whould become comfortable with both since different texts use different notations and both are pretty standard.

Using the complex conjugate gives us a way to extend the way we factor the difference of two squares. Recall from algebra that if $a, b \in \mathbb{R}$ then

$$a^2 - b^2 = (a - b)(a + b) \tag{2.21}$$

We now observe that if $z = a + bi$ then

$$zz^* = (a + bi)(a - bi) = a^2 + b^2 \tag{2.22}$$

Definition 2.7 Let $z = a + bi \in \mathbb{C}$ where $a, b \in \mathbb{R}$. Then the **Absolute Value of** of z is defined as the positive square root,

$$|z| = \sqrt{zz^*} = \sqrt{a^2 + b^2} \tag{2.23}$$

Theorem 2.8 Properties of the Complex Conjugate

$$z + z^* = 2\text{Re } z, \ \forall z \in \mathbb{C} \tag{2.24}$$

$$z - z^* = 2i\text{Im } z, \ \forall z \in \mathbb{C} \tag{2.25}$$

$$\overline{z + w} = z^* + w^*, \ \forall z, w \in \mathbb{C} \tag{2.26}$$

$$\overline{zw} = (z^*)(w^*), \ \forall z, w \in \mathbb{C} \tag{2.27}$$

$$(z^*)^* = z, \ \forall z \in \mathbb{C} \tag{2.28}$$

$$|wz| = |w||z|, \ \forall w, z \in \mathbb{C} \tag{2.29}$$

We can represent the complex number $z = a+bi$ as a point (a, b) in the complex plane with x-coordinate a and y-coordinate b. The **phase** or **argument** of z is defined to be the angle between the line segment from $(0, 0)$ to (a, b) and the x-axis (the real axis).

Definition 2.9 Let $z = a + bi$. Then the **phase** of z is defined as

$$\text{Ph}(z) = \tan^{-1}(a, b) \tag{2.30}$$

where the quadrant is determined by the location of the point (a, b) in the complex plane.

Since the distance between the origin and (a, b) is $\sqrt{a^2 + b^2} = |z|$ we have

Theorem 2.10 Euler's Formula

$$a + bi = |z|(\cos\theta + i\sin\theta) = |z|e^{i\theta} \tag{2.31}$$

where $\theta = \text{Ph}(z)$.

The second equality in equation 2.31 can be proven by expanding $\sin\theta$, $\cos\theta$ and $e^{i\theta}$ in MacLaurin series.

Theorem 2.11 Euler's Equation

$$e^{i\pi} = -1 \tag{2.32}$$

By taking the n^{th} root of 2.31 we have the following.

Theorem 2.12 Every complex number z has a total of n distinct n^{th} roots given by

$$r_k = |z|^{1/n}\left[\cos\left(\frac{\theta + 2k\pi}{n}\right) + i\sin\left(\frac{\theta + 2k\pi}{n}\right)\right],\ k = 0, 1, .., n-1 \tag{2.33}$$

Example 2.1 Find $\sqrt{i}$.

From 2.31 we have $i = e^{i\pi/2} = e^{i\pi/2 + 2k\pi}$ hence

$$\sqrt{i} = \left(e^{i\pi/2+2k\pi}\right)^{1/2} \tag{2.34}$$

$$= e^{i\pi/4+k\pi} \tag{2.35}$$

$$= \cos\left(\frac{\pi}{4} + k\pi\right) + i\sin\left(\frac{\pi}{4} + k\pi\right),\ k = 0, 1 \tag{2.36}$$

$$= \frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2},\ \frac{-\sqrt{2}}{2} - i\frac{\sqrt{2}}{2} \quad \square \tag{2.37}$$

Topic 3

Vector Spaces

Definition 3.1 A **list of length n** is a finite ordered collection of n objects, e.g.,

$$(x_1, x_2, \ldots, x_n) \tag{3.1}$$

The expression

$$(x_1, x_2, \ldots) \tag{3.2}$$

refers to a list of some finite unspecified length (it is not an infinite list). The **list of length 0** is written as ().

A list is similar to set except that the order is critical. For example,

$$\{1, 3, 4, 2, 3, 5\} \text{ and } \{1, 2, 3, 3, 4, 5\} \tag{3.3}$$

are the same set but

$$(1, 3, 4, 2, 3, 5) \text{ and } (1, 2, 3, 3, 4, 5) \tag{3.4}$$

are different lists.

We will sometimes use the term **ordered tuple** and list interchangeably.

Definition 3.2 An **ordered pair** is a list of length 2. An **ordered triple** is a list of length 3.

Definition 3.3 Two lists are said to be **equal** if they contain the same elements in the same locations.

We are already familiar of the use of a list to represent a point in space; the list (x, y) represents a point in $\mathbb{R}^2$, etc.

Definition 3.4 The set $\mathbb{F}^n$ is defined as the set of lists

$$\mathbb{F}^n = \{(x_1, x_2, \ldots, x_n)|x_1, x_2, \ldots, x_n \in F\} \tag{3.5}$$

We will use the simplified notation

$$x = (x_1, x_2, \ldots, x_n) \tag{3.6}$$

to represent a list. For example, if x and y are two lists of the same length n then we can defined **list addition** as

$$x + y = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \tag{3.7}$$

It is not possible to add lists of different lengths, but this will not be a problem since we will in general be interested in using lists to represent points in $\mathbb{F}^n$. For example, if $x, y \in \mathbb{F}^n$ we can define the point

$$z = x + y \in \mathbb{F}^n \tag{3.8}$$

by equation 3.7.

Definition 3.5 Let $\mathbb{F}$ be a field. A **vector space** $\mathbb{V}$ **over** $\mathbb{F}$ is a set $\mathbb{V}$ along with two operations **addition** and **scalar multiplication** with the following properties[1] Elements of vector spaces are called **vectors** or **points**.

1. Closure. $\forall v, w \in \mathbb{V}$ and $\forall c \in \mathbb{F}$,

$$\left.\begin{aligned} & v + w \in \mathbb{V} \\ & cv \in \mathbb{V} \end{aligned}\right\} \tag{3.9}$$

2. Commutivity. $\forall u, v \in \mathbb{V}$,

$$u + v = v + u \tag{3.10}$$

3. Associativity. $\forall u, v, w \in \mathbb{V}$ and $\forall a, b \in \mathbb{F}$,

$$\left.\begin{aligned} & (u + v) + w = u + (v + w) \\ & (ab)v = a(bv) \end{aligned}\right\} \tag{3.11}$$

4. Additive Identity. $(\exists 0 \in \mathbb{V}) \ni$

$$v + 0 = 0 + v = v \tag{3.12}$$

$\forall v \in \mathbb{V}$.

[1]Although this looks deceptively like the definition of a field, it is not, because we only require scalar multiplication and not multiplication between elements of the set. Also one should be careful to avoid confusing the terms **vector space** and **vector field**, which you might see in your studies. A **vector field** is actually a function $f : \mathbb{F}^n \mapsto \mathbb{V}(\mathbb{F})$ that associates an element in the vector space $\mathbb{V}$ over $\mathbb{F}$ with every point in $\mathbb{F}^n$. Vector fields occur frequently in physics.

5. Additive Inverse. $(\forall v \in \mathbb{V}), (\exists w \in \mathbb{V}) \ni$

$$v + w = w + v = 0 \tag{3.13}$$

6. Multiplicative Identity. $1v = v1 = v$ for all $v \in \mathbb{V}$, where 1 is the multipicative identity in $\mathbb{F}$.

7. Distributive Property. $(\forall a, b \in \mathbb{F})$ and $(\forall u, v \in \mathbb{V})$

$$a(u + v) = au + av, \qquad (a + b)u = au + bu \tag{3.14}$$

A vector space over $\mathbb{R}$ is called a **Real vector space**. For example, Euclidean 3-space $\mathbb{R}^3$ defined by

$$\mathbb{R}^3 = \{(x, y, z)|x, y, z \in \mathbb{R}\} \tag{3.15}$$

is a real vector space.

A vector space over $\mathbb{C}$ is called a **Complex vector space**. The set of all points

$$\mathbb{C}^4 = \{(z_1, z_2, z_3, z_4)|z_1, z_2, z_3, z_4 \in \mathbb{C}\} \tag{3.16}$$

is a complex vector field, where each element is a 4-vector with complex values.

Some texts will use the term **linear vector space** or **linear space** instead of vector space.

Example 3.1 Let $\mathbb{V}$ be set of all polynomials $p(f) : \mathbb{F} \mapsto \mathbb{F}$, with coefficients in $\mathbb{F}$. The elements of $\mathbb{V}$ are then functions, such as

$$v = a_0 + a_1 f + a_2 f^2 + \cdots + a_m f^m \tag{3.17}$$

where $a_0, a_1, \dots, a_m \in \mathbb{F}$. The $\mathbb{V}$ is a vector space over $\mathbb{F}$ with the normal definition of addition and multiplication over $\mathbb{F}$. If

$$v = \sum_{k=0}^{m} v_k f^k, \qquad w = \sum_{k=0}^{n} w_k f^k, \qquad u = \sum_{k=0}^{p} u_k f^k, \tag{3.18}$$

are polynomials over $\mathbb{F}$, then

$$v + w = \sum_{k=0}^{m} v_k f^k + \sum_{k=0}^{n} w_k f^k \tag{3.19}$$

$$= \sum_{k=0}^{\max(m,n)} (v_k + w_k) f^k \in \mathbb{V} \tag{3.20}$$

where if $m < n$ the we define $v_k = 0$; and if $m > m = n$ we define $w_k = 0$. proving closure, and, since $v_k + w_k = w_k + v_k$ by the commutivity of $\mathbb{F}$, commutivity also follows,

$$v + w = \sum_{k=0}^{\max(m,n)} (v_k + w_k) f^k = \sum_{k=0}^{\max(m,n)} (w_k + v_k) f^k = w + v \tag{3.21}$$

To see associativity over addition,

$$(u + v) + w = \sum_{k=0}^{\max(m,n,p)} ((u_k + v_k) + w_k) f^k \tag{3.22}$$

$$= \sum_{k=0}^{\max(m,n,p)} (u_k + (v_k + w_k)) f^k \tag{3.23}$$

$$= u + (v + w) \tag{3.24}$$

where the coefficients u_k, v_k, w_k are all suitably extended to $\max(m, n, p)$ by setting them equal to zero in the summation.

The addititive identity is the polynomial $g(f) = 0$, since $g(f) + v(f) = v(f) + g(f) =$

$v(f)$ over $\mathbb{F}$; the additive inverse if v is

$$-v = \sum_{k=0}^{m}(-v_k)f^k \tag{3.25}$$

Multiplying any polynomial by $1 \in \mathbb{F}$ returns the original polynomial:

$$1v = \sum_{k=0}^{m}(1)(v_k)f^k = \sum_{k=0}^{m}(v_k)f^k = v \tag{3.26}$$

becuase $1v_k = v_k$. Finally, we see distributivity. Let $a, b \in \mathbb{F}$. Then

$$a(u+v) = a\left(\sum_{k=0}^{p} u_k f^k + \sum_{k=0}^{m} v_k f^k\right) \tag{3.27}$$

$$= a\sum_{k=0}^{p} u_k f^k + a\sum_{k=0}^{m} v_k f^k = au + av \tag{3.28}$$

$$(a+b)u = (a+b)\sum_{k=0}^{p} u_k f^k \tag{3.29}$$

$$= a\sum_{k=0}^{p} u_k f^k + b\sum_{k=0}^{p} u_k f^k = av + bv \quad \square \tag{3.30}$$

Example 3.2 Let $\mathbb{V}$ be the set of all polynomials on $\mathbb{F}$ of degree at most n. Then $\mathbb{V}$ is a vector space.

Example 3.3 Let $\mathbb{V}$ be the set of all 2×3 vector matrices over $\mathbb{F}$:

$$v = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \tag{3.31}$$

Then $\mathbb{V}$ is a vector space over $\mathbb{F}$ with matrix addition as $(+)$ and scalar multiplication of a matrix as $(\cdot)$. For example, the sum of any two 2×3 matrix is a 2×3 matrix; matrix addition is commutative and associative; the additive inverse of v is the matrix

$-v$ and the additive identity is the matrix of all zeros; the multiplicative identity $1 \in \mathbb{F}$ is the multiplicative identity; and scalar multiplication of matrices distributes over matrix summation. □

Example 3.4 Let $\mathbb{V}$ be the set of all functions $f : \mathbb{F} \mapsto \mathbb{F}$. Then $\mathbb{V}$ is a vector space.

Theorem 3.6 Let $\mathbb{V}$ be an vector space over $\mathbb{F}$. Then $\mathbb{V}$ has a unique additive identity.

Proof. Suppose that $0, 0' \in \mathbb{V}$ are both additive identities. Then

$$0' = 0' + 0 = 0 \tag{3.32}$$

Hence any additive identity equals 0, proving uniqueness. □

Theorem 3.7 Let $\mathbb{V}$ be a vector space. Then every element in $\mathbb{V}$ has a unique additive inverse, i.e., $\forall v \in \mathbb{V} \exists$ a unique $w \ni v + w = w + v = 0$.

Proof. Let $v \in \mathbb{V}$ have two different additive inverses w and w'. Then since each is an additive inverse,

$$w = w + 0 = w + (v + w') = (w + v) + w' = 0 + w' = w' \tag{3.33}$$

which proves uniqueness. □

Notation: Because, given any $v \in \mathbb{V}$, its additive inverse is unique, we write it as $-v$, i.e.,

$$v + (-v) = (-v) + v = 0 \tag{3.34}$$

Furthermore, given any $v, w \in \mathbb{V}$, we define the notation

$$w - v = w + (-v) \tag{3.35}$$

Theorem 3.8 Let $1 \in \mathbb{F}$ be the multiplicative identity in $\mathbb{F}$ and let $v \in \mathbb{V}$. Then

$$(-1)v = -v \tag{3.36}$$

Proof.

$$v + (-1)v = (1)v + (-1)v = (1 + -1)v = 0v = 0 \tag{3.37}$$

Hence $(-1)v$ is the additive inverse of v, which we call $-v$. Hence $(-1)v = -v$. □

Notation. Because of associativity we will dispense with parenthesis, i.e., given any $u, v, w \in \mathbb{V}$ and any $a, b \in \mathbb{F}$, we define the notations uvw and $u + v + w$ to mean the following:

$$u + v + w = (u + v) + w = u + (v + w) \tag{3.38}$$

$$abv = a(bv) = (ab)v \tag{3.39}$$

Remark. There are different $0's$, namely sometimes we will use $0 \in \mathbb{F}$ and other times $0 \in \mathbb{V}$ and sometimes they will be used in the same equation. We will use the same symbol for these zeroes. Be careful.

Theorem 3.9 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then $\forall v \in \mathbb{V}$,

$$0v = 0 \tag{3.40}$$

This is one of those cases where we have two different zeroes in the same equation. The 0 on the left is the scalar zero in $\mathbb{F}$, while the 0 on the right is the vector $0 \in \mathbb{V}$.

Proof.

$$0v = (0 + 0)v = 0v + 0v \;\Rightarrow\; 0v - 0v = 0v \;\Rightarrow\; 0 = 0v \tag{3.41}$$

□

Theorem 3.10 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let $0 \in \mathbb{V}$ be the additive indentity in $\mathbb{V}$. Then

$$a0 = 0 \tag{3.42}$$

Proof.

$$a0 = a(0+0) = a0 + a0 \ \Rightarrow 0 = a0 - a0 = a0 + a0 - a0 = a0 \tag{3.43}$$

□

Topic 4

Subspaces

Definition 4.1 A subset $\mathbb{U} \subset \mathbb{V}$ is called a **subspace** of $\mathbb{V}$ if $\mathbb{U}$ is also a vector space. Specifically, the following properties need to hold:

1. Additive identity:

$$0 \in \mathbb{U} \tag{4.1}$$

2. Closure under addition:

$$\forall u, v \in \mathbb{U}, \quad u + v \in \mathbb{U} \tag{4.2}$$

3. Closure under scalar multiplication:

$$\forall a \in \mathbb{F} \wedge \forall u \in \mathbb{U}, \quad au \in \mathbb{U} \tag{4.3}$$

For example $\mathbb{V} = \{(x, y, 0)|x, y \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^3$, and $\mathbb{W} = \{(x, 0)|x \in \mathbb{R}\}$ is a subspace of $\mathbb{R}^2$.

Definition 4.2 Let $\mathbb{U}_1, \dots, \mathbb{U}_n$ be subspaces of $\mathbb{V}$. Then the **sum** of $\mathbb{U}_1, \dots \mathbb{U}_n$ is

$$\mathbb{U}_+ \cdots + \mathbb{U}_n = \{u_1 + \cdots + u_n | u_1 \in \mathbb{U}_1 \dots, u_n \in \mathbb{U}_n\} \tag{4.4}$$

Theorem 4.3 Let $\mathbb{U}$ and $\mathbb{W}$ be subspaces of $\mathbb{V}$. The $\mathbb{U} \cap \mathbb{W}$ is a subspace of $\mathbb{V}$.

Proof. (a) Since $\mathbb{U}$ and $\mathbb{W}$ are subspaces of $\mathbb{V}$ they each contain 0. Hence their intersection contains 0.

(b) Let $x, y \in \mathbb{U} \cap \mathbb{W}$. Then $x, y \in \mathbb{U}$ and $x, y \in \mathbb{W}$. Hence by closure of, $\mathbb{U}$, $x + y \in \mathbb{U}$, and by closure of $\mathbb{W}$, $x + y \in \mathbb{W}$. Hence $x + y \in \mathbb{U} \cap \mathbb{W}$.

(c) Let $a \in \mathbb{F}$ and $x \in \mathbb{U} \cap \mathbb{W}$. Hence $x \in \mathbb{U}$ and $x \in \mathbb{W}$. Since $\mathbb{U}$ is a subspace, it is closed under scalar multiplication, hence $ax \in \mathbb{U}$. Similarly, $\mathbb{V}$ is a subspace, so it is also closed under scalar multiplication, hence $ax \in \mathbb{W}$. Thus $ax \in \mathbb{U} \cap \mathbb{W}$. □

Theorem 4.4 If $\mathbb{U}_1, \dots, \mathbb{U}_n$ are subspaces of $\mathbb{V}$ then $\mathbb{U}_1 + \cdots + \mathbb{U}_n$ is a subspace of $\mathbb{V}$.

Proof. (exercise.) □

Example 4.1 Let $\mathbb{V} = \mathbb{F}^3$, and

$$\mathbb{U} = \{(x, 0, 0) \in \mathbb{F}^3 | x \in \mathbb{F}\} \tag{4.5}$$

$$\mathbb{W} = \{(0, y, 0) \in \mathbb{F}^3 | x \in \mathbb{F}\} \tag{4.6}$$

The $\mathbb{U}$ and $\mathbb{W}$ are subspaces of $\mathbb{V}$, and

$$\mathbb{U} + \mathbb{W} = \{(x, y, 0) \in \mathbb{F}^3 | x, y \in \mathbb{F}\} \tag{4.7}$$

is also a subspace of $\mathbb{V}$.

Theorem 4.5 Let $\mathbb{U}_1, \dots, \mathbb{U}_n$ be substaces of $\mathbb{V}$. Then $\mathbb{U}_1 + \cdots + \mathbb{U}_n$ is the smallest subspace of $\mathbb{V}$ that contains all of $\mathbb{U}_1, \dots, \mathbb{U}_n$.

Proof. We need to prove (a) that $\mathbb{U}_1 + \cdots + \mathbb{U}_n$ contains $\mathbb{U}_1, \dots, \mathbb{U}_n$, and (b) that any subspace that contains $\mathbb{U}_1, \dots, \mathbb{U}_n$ also contains $\mathbb{U}_1 + \cdots + \mathbb{U}_n$.

To see (a), let $u_1 \in \mathbb{U}_1$. Since $0 \in \mathbb{U}_i$ for all i, we can let $u_2 = 0 \in \mathbb{U}_2, u_3 = 0 \in \mathbb{U}_3, \dots, u_n = 0 \in \mathbb{U}_n$. Then

$$u_1 = u_1 + 0 + \cdots + 0 \tag{4.8}$$

$$= u_1 + u_2 + \cdots + u_n \in \mathbb{U}_1 + \cdots + \mathbb{U}_n \tag{4.9}$$

Hence $\mathbb{U}_1 \subset \mathbb{U}_1 + \cdots + \mathbb{U}_n$. By a similar argument, each of $\mathbb{U}_2, \dots, \mathbb{U}_n \subset \mathbb{U}_1 + \cdots + \mathbb{U}_n$. This proves assertion (a).

To see (b), Let $\mathbb{W}$ be a subspace of $\mathbb{V}$ such that $\mathbb{U}_1 \subset \mathbb{W}, \dots, \mathbb{U}_n \subset \mathbb{W}$. Let $u_1 \in \mathbb{U}_1, \dots, u_n \in \mathbb{U}_n$. By definition of the sum,

$$u_1 + u_2 + \cdots + u_n \in \mathbb{U}_1 + \mathbb{U}_2 + \cdots + \mathbb{U}_n \tag{4.10}$$

But since each $\mathbb{U}_i \subset \mathbb{W}$, then $u_i \in \mathbb{W}$. Since $\mathbb{W}$ is a vector space, it is closed under addition, hence

$$u_1 + \cdots + u_n \in \mathbb{W} \tag{4.11}$$

Since this is true for every element of $\mathbb{U}_1 + \cdots + \mathbb{U}_n$, then

$$\mathbb{U}_1 + \cdots + \mathbb{U}_n \subset \mathbb{W} \tag{4.12}$$

which proves assertion (b). □

Definition 4.6 Let $\mathbb{U}_1, \dots, \mathbb{U}_n$ be subspaces of $\mathbb{V}$ such that

$$\mathbb{V} = \mathbb{U}_1 + \cdots + \mathbb{U}_n \tag{4.13}$$

We say that the $\mathbb{V}$ is the **direct sum of** $\mathbb{U}_1, \dots, \mathbb{U}_n$ if each element of $\mathbb{V}$ can be written uniquely as a sum $u_1 + \cdots + u_n$ where $u_1 \in \mathbb{U}_1, \dots, u_n \in \mathbb{U}_n$, and we write

$$\mathbb{V} = \mathbb{U}_1 \oplus \mathbb{U}_2 \oplus \cdots \oplus \mathbb{U}_n \tag{4.14}$$

Example 4.2 Let $\mathbb{V} = \mathbb{R}^2$. Then $V = \mathbb{U} \oplus W$ where

$$\mathbb{U} = \{(x, 0) | x \in \mathbb{R}\} \tag{4.15}$$

$$\mathbb{W} = \{(0, y) | y \in \mathbb{R}\} \tag{4.16}$$

More generally, if

$$\mathbb{V} = \mathbb{F}^n = \{(v_1, v_2, \dots, v_n) | v_i \in \mathbb{F}\} \tag{4.17}$$

and

$$\left.\begin{aligned} &\mathbb{V}_1 = \{(v, 0, \dots, 0) | v \in \mathbb{F}\} \\ &\mathbb{V}_2 = \{(0, v, 0, \dots, 0) | v \in \mathbb{F}\} \\ &\vdots \\ &\mathbb{V}_n = \{(0, \dots, 0, v) | v \in \mathbb{F}\} \end{aligned}\right\} \tag{4.18}$$

Then

$$\mathbb{V} = \mathbb{V}_1 \oplus \mathbb{V}_2 \oplus \cdots \oplus \mathbb{V}_n \tag{4.19}$$

Example 4.3 Let $\mathbb{V} = \mathbb{R}^3$ and suppose that

$$\left.\begin{aligned} \mathbb{U} &= \{(x, y, 0)|x \in \mathbb{R}\} \\ \mathbb{W} &= \{(0, y, y)|y \in \mathbb{R}\} \\ \mathbb{Z} &= \{(0, 0, z)|z \in \mathbb{R}\} \end{aligned}\right\} \tag{4.20}$$

Then $\mathbb{U}$, $\mathbb{W}$, and $\mathbb{Z}$ are subspaces of $\mathbb{V}$, and

$$\mathbb{V} = \mathbb{U} + \mathbb{W} + \mathbb{Z} \tag{4.21}$$

but

$$\mathbb{V} \neq \mathbb{U} \oplus \mathbb{W} \oplus \mathbb{Z} \tag{4.22}$$

(1) $\underline{\mathbb{U} \text{ is a subspace of } \mathbb{V}}$: $\mathbb{U}$ contains $(0, 0, 0) = 0$;

$$(x, y, 0) + (x', y', 0) = (x + x', y + y', 0) \in \mathbb{U} \tag{4.23}$$

hence $\mathbb{U}$ is closed under addition; and

$$a(x, y, 0) = (ax, ay, 0) \in \mathbb{U} \tag{4.24}$$

hence $\mathbb{U}$ is closed under scalar multiplication.

(2) $\underline{\mathbb{W} \text{ is a subspace of } \mathbb{V}}$: Since $y = 0 \in \mathbb{R}$, we know that $(0, 0, 0) \in \mathbb{W}$. Since

$$(0, y, y) + (0, y', y') = (0, y + y', y + y') \in \mathbb{W} \tag{4.25}$$

then $\mathbb{W}$ is closed under addition; and since

$$a(0, y, y) = (0, ay, ay) \in \mathbb{W} \tag{4.26}$$

then $\mathbb{W}$ is closed under scalar multiplication.

(3) $\underline{\mathbb{Z} \text{ is a subspace of } \mathbb{V}}$: Since $z = 0 \in \mathbb{R}$ then $(0, 0, 0) \in \mathbb{Z}$; since

$$(0, 0, z) + (0, 0, z') = (0, 0, z + z') \in \mathbb{Z} \tag{4.27}$$

we have $\mathbb{Z}$ cloased under addition; and since

$$a(0, 0, z) = (0, 0, az) \in \mathbb{Z} \tag{4.28}$$

it is closed under scalar multiplication.

(4) $\underline{\mathbb{V} = \mathbb{U} + \mathbb{W} + \mathbb{Z}}$: Two sets are equal if each is a subset of the other. Hence if we can show that (a) $\mathbb{V} \subset \mathbb{U} + \mathbb{W} + \mathbb{Z}$ and (b) $\mathbb{U} + \mathbb{W} + \mathbb{Z} \subset \mathbb{V}$ then the two sets must be identical.

To see (a), Let $(x, y, z) \in \mathbb{V}$. Then

$$(x, y, z) = (x, y/2, 0) + (0, y/2, y/2) + (0, 0, z - y/2) \tag{4.29}$$

Since $(x, y/2, 0) \in \mathbb{U}$, $(0, y/2, y/2) \in \mathbb{W}$, and $(0, 0, z - y/2) \in \mathbb{Z}$ then

$$(x, y, z) \in \mathbb{U} + \mathbb{W} + \mathbb{Z} \tag{4.30}$$

and therefore $\mathbb{V} \subset \mathbb{U} + \mathbb{W} + \mathbb{Z}$.

(5) $\underline{\mathbb{V} \neq \mathbb{U} \oplus \mathbb{W} \oplus \mathbb{Z}}$: Consider the element $(0,0,0)$ which is in each of the subspaces as well as $\mathbb{V}$. Then

$$(0,0,0)_V = (0,0,0)_U + (0,0,0)_W + (0,0,0)_Z \tag{4.31}$$

But we also can write

$$(0,0,0)_V = (0,17,0)_U + (0,-17,-17)_W + (0,0,17)_Z \tag{4.32}$$

This means we can express the vector $(0,0,0)$ as two different sums of the form $u+w+z$, and hence the method of defining u, w, z is not unique.

Going back to equation 4.29 we see that that sum is also not unique. For example, we could also write

$$(x,y,z) = (x, y/4, 0)_U + (0, 3y/4, 3y/4)_W + (0, 0, z - 3y/4)_Z \tag{4.33}$$

as well as any number of other combinations! In fact, we only have to check the zero vector to make sure it can only be formed of individual zero vectors, from the following theorem.

Theorem 4.7 Let $\mathbb{U}_1, \dots, \mathbb{U}_n$ be subspaces of $\mathbb{V}$. Then $V = \mathbb{U}_1 \oplus \cdots \oplus \mathbb{U}_n$ if and only if both of the following are true:

(1) $\mathbb{V} = \mathbb{U}_1 + \cdots + \mathbb{U}_n$

(2) The only way to write 0 as a sum $u_1 + \cdots + u_n$, with each $u_j \in \mathbb{U}_j$, is if $u_j = 0$ for all the j.

Proof. ($\Longrightarrow$) Suppose that

$$V = \mathbb{U}_1 \oplus \cdots \oplus \mathbb{U}_n \tag{4.34}$$

Then by definition of direct sum,

$$V = \mathbb{U}_1 + \cdots + \mathbb{U}_n \tag{4.35}$$

which proves (1). Now suppose that there are $u_i \in \mathbb{U}_i, i = 1, \dots, n$, such that

$$u_1 + \cdots + u_n = 0 \tag{4.36}$$

By the uniqueness part of the defintion of direct sums, there must be only one way to choose these u_i. Since we know that the choice $u_i = 0, i = 1, \dots, n$ works, this choice must be unique. Thus (2) follows.

($\Longleftarrow$) Suppose that (1) and (2) both hold. Let $v \in \mathbb{V}$. By (1) we can find $u_1 \in \mathbb{U}_1, \dots, u_n \in \mathbb{U}_n$ so that

$$v = u_1 + u_2 + \cdots + u_n \tag{4.37}$$

Suppose that there is some other representation

$$v = v_1 + v_2 + \cdots + v_n \tag{4.38}$$

Then

$$0 = v - v = (u_1 - v_1) + (u_2 - v_2) + \cdots + (u_n - v_n) \tag{4.39}$$

where by closure of each $\mathbb{U}_i$ we have $u_i - v_i \in \mathbb{U}_i$. Since we have assumed that (2) is true, then we must have $u_i - v_i = 0$ for all i. Hence $u_i = v_i$ and therefore the representation is unique. □

Theorem 4.8 Let $\mathbb{V}$ be a vector space with subspaces $\mathbb{U}$ and $\mathbb{W}$. Then $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ if and only if $\mathbb{V} = \mathbb{U} + \mathbb{W}$ and $\mathbb{U} \cap \mathbb{W} = \{0\}$.

Proof. ($\Longrightarrow$) Suppose that $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$.

Then $\mathbb{V} = \mathbb{U} + \mathbb{W}$ by the definition of direct sum. Now suppose that $v \in \mathbb{U} \cap \mathbb{W}$. Then

$$0 = v + (-v) \tag{4.40}$$

where $v \in \mathbb{U}$ and $-v \in \mathbb{W}$ (since they are both in the intersection). Since the representation must be unique this means that $v = 0$ (in $\mathbb{U}$) and $-v = 0$ (in $\mathbb{W}$). Thus $\mathbb{U} \cap \mathbb{W} = \{0\}$.

($\Longleftarrow$) Suppose that $\mathbb{V} = \mathbb{U} + \mathbb{W}$ and $\mathbb{U} \cap \mathbb{W} = \{0\}$. Suppose that there exist $u \in \mathbb{U}$ and $w \in \mathbb{W}$ such that

$$0 = u + w \tag{4.41}$$

But by equation 4.41

$$-w = u + w - w = u \tag{4.42}$$

Since $-w \in \mathbb{W}$ then $u \in \mathbb{W}$. Hence $u \in \mathbb{U} \cap \mathbb{W}$. But $\mathbb{U} \cap \mathbb{W} = \{0\}$ and therefore $u = 0$. Since $w = -u$ we also have $w = 0$. Hence the only way to construct $0 = u + w$ is with $u = w = 0$, from which by theorem 4.7 we conclude that $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$. $\square$

Topic 5

Polynomials

Definition 5.1 A function $p(x) : \mathbb{F} \mapsto \mathbb{F}$ is called a **polynomial (with coefficients) in** $\mathbb{F}$ if there exists $a_0, \ldots, a_n \in \mathbb{F}$ such that

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n \tag{5.1}$$

If $a_n \neq 0$ then we say that the **degree** of p is n and we write $\deg(p) = n$.

If all the coefficients $a_0 = \cdots = a_n = 0$ we say that $\deg(p) = -\infty$.

We denote the vector space of all polynomials with coefficents in $\mathbb{F}$ as $\mathbb{P}(\mathbb{F})$ and the vector space of all polynomials with coefficients in $\mathbb{F}$ and **degree at most** m as $\mathbb{P}_m(\mathbb{F})$ (thus, e.g., $3x^2 + 4$ is in $\mathbb{P}_3$ as well as $\mathbb{P}_2$).

A number $\lambda \in \mathbb{F}$ is called **a root of** p if

$$p(\lambda) = 0 \tag{5.2}$$

Lemma 5.2 Let $\lambda \in \mathbb{F}$. Then for $j = 2, .3, \ldots,$

$$z^j - \lambda^j = (z - \lambda)(z^{j-1} + z^{j-2}\lambda + \cdots + z\lambda^{j-2} + \lambda^{j-1}) \tag{5.3}$$

Proof. (Exercise.) □

Theorem 5.3 Let $p \in \mathbb{P}(\mathbb{F})$ be a polynomial with degree $m \geq 1$. Then $\lambda \in \mathbb{F}$ is a root of p if and only if there is a polynomial $q \in \mathbb{P}(\mathbb{F})$ with degree $m - 1$ such that

$$p(z) = (z - \lambda)q(z) \tag{5.4}$$

for all $z \in \mathbb{F}$.

Proof. ($\Longrightarrow$) Suppose that there exists a $q \in \mathbb{P}(\mathbb{F})$ such that 5.4 holds. Then

$$p(\lambda) = (\lambda - \lambda)q(\lambda) = 0 \tag{5.5}$$

hence λ is a root of p.

($\Longleftarrow$)Let λ be a root of p, where

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n \tag{5.6}$$

Then

$$0 = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_n\lambda^n \tag{5.7}$$

Subtracting,

$$p(z) = a_1(\lambda - z) + a_2(z^2 - \lambda^2) + \cdots + a_n(z^n - \lambda^n) \tag{5.8}$$

By the lemma

$$z^j - \lambda^j = (z - \lambda)q_{j-1}(z) \tag{5.9}$$

where

$$q_{j-1}(z) = z^{j-1} + z^{j-2}\lambda + \cdots + z\lambda^{j-2} + \lambda^{j-1} \tag{5.10}$$

and therefore

$$p(z) = a_1(\lambda - z) + a_2(z - \lambda)q_1(z) + a_3(z - \lambda)q_2(z) + \cdots + a_n(z - \lambda)q_n(z) \tag{5.11}$$

$$= (z - \lambda)(a_1 + a_2 q_1(z) + a_3 q_2(z) + \cdots + a_n q_{n-1}(z)) \tag{5.12}$$

$$= (z - \lambda)q(z) \tag{5.13}$$

where

$$q(z) = a_1 + a_2 q_1(z) + a_3 q_2(z) + \cdots + a_n q_{n-1}(z) \tag{5.14}$$

is a polynomial of degree $n - 1$ (because the sum of polynomials of degree at most $n-1$ is a polynomial of degree at most $n-1$, and because there is a nonzero-coefficient to the z^{n-1} term). □

Corollary 5.4 Let $p \in \mathbb{P}(\mathbb{F})$ and that $\deg(p) = m \geq 0$. Then p has at most m distinct roots in $\mathbb{F}$.

Proof. If $m = 0$, then $p(z) = a_0 \neq 0$ so p has no roots (and $0 \leq 0 = m$).

If $m = 1$ then $p(z) = a_0 + a_1 z$ where $a_1 \neq 0$, and p has precisely one root at $z = -a_0/a_1$.

Suppose $m > 1$ and assume that the theorem is true for degree $m - 1$ (inductive hypothesis).

Either p has a root $\lambda \in \mathbb{F}$ or it does not.

If p does not have a root then the number of roots is 0 which is less than m and the theorem is proven.

Now assume that p does have a root λ. By theorem 5.3 there is a polynomial $q(z)$ with $\deg(q) = m - 1$ such that

$$p(z) = (z - \lambda)q(z) \tag{5.15}$$

By the inductive hypothesis, q has at most $m - 1$ distinct roots. Either λ is one of these roots, in which case p has precisely $m - 1$ roots, or λ is not one of these roots, in which case p has at most $m - 1 + 1 = m$ distinct roots. In either case, p has $\leq m$ distinct roots, proving the theorem. □

Corollary 5.5 Let $a_0, \dots, a_n \in \mathbb{F}$ and

$$p(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n = 0 \tag{5.16}$$

for all $z \in \mathbb{F}$. Then $a_0 = a_1 = \cdots = a_n = 0$.

Proof. Suppose that equation 5.16 holds with at least one $a_j \neq 0$. Then by corollary 5.4 $p(z)$ has at most m roots. But by equaton 5.16 every value of z is a root, which means that p has an infinite number of roots. This is a contradiction. Hence all the $a_j = 0$. □

Theorem 5.6 Polynomial Division. Let $p(z), q(z) \in \mathbb{P}(\mathbb{F})$ with $p(z) \neq 0$ (the zero polynomial). Then there exist polynomials $s(z), r(z) \in \mathbb{P}(\mathbb{F})$ such that

$$q(z) = s(z)p(z) + r(z) \tag{5.17}$$

such that $\deg(r) < \deg(p)$.

Proof. [1] Let $n = \deg(q)$ and $m = \deg(p)$. We need to show that there exist s and r with $\deg(r) < m$.

Case 1. Suppose that $n < m$ ($\deg(q) < \deg(p)$). Let $s = 0$ and $r = q(z)$. Then $\deg(r(z)) = \deg(q(z)) = n < m$ as required.

Case 2. $0 \leq m \leq n$. Prove by induction.

If $n = 0$: Since $0 \leq m \leq n = 0$ we must have $m = 0$. Hence there exists non-zero constants $q_0 \in \mathbb{F}$ and $p_0 \in \mathbb{F}$ such that

$$\left.\begin{aligned} q(z) &= q_0 \\ p(z) &= p_0 \end{aligned}\right\} \tag{5.18}$$

we need to find $s(z)$ and $r(z)$ such that

$$q_0 = p_0 s(z) + r(z) \tag{5.19}$$

If we choose $r(z) = 0$ and

$$s(z) = \frac{q_0}{p_0} \tag{5.20}$$

then $\deg(r) = -\infty < 0 = m$ as required.

If $n > 0$: (Inductive hypothesis): Let n be fixed and assume that the result holds for all polynomials of degree less than n. Then we can assume that there exist some constants $p_0, \dots, p_m \in \mathbb{F}$ and $q_0, \dots, q_n \in \mathbb{F}$, with $p_n \neq 0$ and $q_m \neq 0$, such that

$$q(z) = q_0 + q_1 z + q_2 z^2 + \cdots + q_n z^n \tag{5.21}$$

[1] This proof is based on the one given in Friedberg et al., *Linear Algebra, 4th edition*, Pearson (2003).

and

$$p(z) = p_0 + p_1 z + p_2 z^2 + \cdots + p_m z^m \tag{5.22}$$

with $0 \leq m \leq n$. Then define

$$h(z) = q(z) - \frac{q_n}{p_m} z^{n-m} p(z) \tag{5.23}$$

$$= \left(q_0 + q_1 z + q_2 z^2 + \cdots + q_n z^n\right) - \frac{q_n}{p_m} z^{n-m} \left(p_0 + p_1 z + p_2 z^2 + \cdots + p_m z^m\right) \tag{5.24}$$

Then since the z^n terms subtract out,

$$\deg(h(z)) < n \tag{5.25}$$

Then either $\deg(h) < m = \deg(p)$ or $m \leq \deg(h) < n$. In the first instance, case 1 applies, and in the second instance the inductive hypothesis applies to h and p, i.e., there exist polynomials $s'(z)$ and $r(z)$ such that

$$h(z) = s'(z)p(z) + r(z) \tag{5.26}$$

with

$$\deg(r) < \deg(p) = m \tag{5.27}$$

Substituting equation 5.26 into equation 5.23 gives

$$q(z) - \frac{q_n}{p_m} z^{n-m} p(z) = s'(z)p(z) + r(z) \tag{5.28}$$

Solving for $q(z)$,

$$q(z) = \frac{q_n}{p_m} z^{n-m} p(z) + s'(z)p(z) + r(z) \tag{5.29}$$

$$= \left(\frac{q_n}{p_m} z^{n-m} + s'(z)\right) p(z) + r(z) \tag{5.30}$$

$$= s(z)p(z) + r(z) \tag{5.31}$$

for some polynomial $s(z)$, as required. □

Theorem 5.7 The factorization of theorem 5.6 is unique.

Proof. (Exercise.) □

Theorem 5.8 (**Fundamental Theorem of Algebra**) Every polynomial of degree n over $\mathbb{C}$ has precisely n roots.

Descarte proposed the fundamental theorem of algebra in 1637 but did not prove it. Albert Girard had earlier (1629) proposed that an n^{th} order polynomial has n roots but that they may exist in a field larger than the complex numbers. The first published proof of the fundamental theorem of algebra was by DAlembert in 1746, but his proof was based on an earlier theorem that itself used the theorem, and hence is circular. At about the same time Euler proved it for polynomials with real coefficients up to 6th order. Between 1799 (in his doctoral dissertation) and 1816 Gauss published three different proofs for polynomials with with real coefficients, and in 1849 he proved the general case for polynomials with complex coefficients.

Theorem 5.9 Let $p(z)$ be a polynomial with real coefficients. Then if $\lambda \in \mathbb{C}$ is a root of p then λ^* is also a root of $p(z)$.

Proof. Define

$$p(z) = a_0 + a_1 z + \cdots + a_n z^m \tag{5.32}$$

Since λ is a root,

$$0 = a_0 + a_1 \lambda + \cdots + a_n \lambda^m \tag{5.33}$$

Taking the complex conjugate of this equation proves the theorem:

$$0^* = (a_0 + a_1 \lambda + \cdots + a_n \lambda^m)^* \tag{5.34}$$

$$a_0 + a_1 \lambda^* + \cdots + a_n (\lambda^*)^m \tag{5.35}$$

Since $0 = 0^*$, we have the result that λ^* is a root of p. □

Theorem 5.10 Let $\alpha, \beta \in \mathbb{R}$. Then there exist $\lambda_1, \lambda_2 \in \mathbb{C}$ such that

$$z^2 + \alpha z + \beta = (z - \lambda_1)(z - \lambda_2) \tag{5.36}$$

and $\lambda_1, \lambda_2 \in \mathbb{R} \iff \alpha^2 > 4\beta$.

Proof. We "complete the squares" in the quadratic,

$$z^2 + \alpha z + \beta = z^2 + \alpha z + \left(\frac{\alpha}{2}\right)^2 - \left(\frac{\alpha}{2}\right)^2 + \beta \tag{5.37}$$

$$= \left(z + \frac{\alpha}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right) \tag{5.38}$$

If $\alpha^2 \geq 4\beta$, then we define $c \in \mathbb{R}$ by

$$c^2 = \frac{\alpha^2}{4} - \beta \tag{5.39}$$

and therefore

$$z^2 + \alpha z + \beta = \left(z + \frac{\alpha}{2}\right)^2 - c^2 \tag{5.40}$$

$$= \left(z + \frac{\alpha}{2} - c\right)\left(z + \frac{\alpha}{2} + c\right) \tag{5.41}$$

hence

$$\lambda_{1,2} = \frac{1}{2}\left(-\alpha \pm \sqrt{\alpha^2 - 4\beta}\right) \tag{5.42}$$

which is the familiar quadratic equation.

If $\alpha^2 < 4\beta$ then the right hand side of the equation 5.37 is always positive, and hence there is no real value of z that gives $p(z) = 0$. Hence there can be no real roots; if any root exists, it must be complex. We can solve for these two roots using the quadratic formula:

$$\lambda_{1,2} = \frac{1}{2}\left(-\alpha \pm i\sqrt{4\beta - \alpha^2}\right) \tag{5.43}$$

(substitution proves that these are roots). This is a complex conjugate pair. □

Theorem 5.11 If $p \in \mathbb{P}(\mathbb{C})$ is a non-constant polynomial then it has a unique factorization

$$p(z) = c(z - \lambda_1)(z - \lambda_2)\cdots(z - \lambda_n) \tag{5.44}$$

where $c, \lambda_1, \dots, \lambda_n \in \mathbb{C}$ and $\lambda_1, \dots, \lambda_n$ are the roots (note that some of the roots might be repeated, if they are real).

If some of the roots are complex, then let m be the number of real roots and $k = n - m$ be the number of complex roots, where k is even because the complex roots come in pairs. Then the complex roots can written as

$$\lambda_{m+2j-1} = a_j + ib_j, \quad \lambda_{2j} = a_j - ib_j, \quad j = 1, \dots, k/2 \tag{5.45}$$

for some real numbers $a_1, b_1, \ldots, a_n, b_n$. Then

$$\begin{aligned} p(x) = c(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)\times \\ [(x-\lambda_{m+1})(x-\lambda_{m+2}]\cdots[(x-\lambda_{m+k-1})(x-\lambda_{m+k})] \quad (5.46) \\ = c(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)(x^2+\alpha_1 x+\beta_1)\cdots(x^2+\alpha_k x+\beta_k) \quad (5.47) \end{aligned}$$

where

$$x^2+\alpha_j x+\beta_j = (x-\lambda_{m+2j-1})(x-\lambda_{m+2j}) \quad (5.48)$$

$$= (x-a_j-ib_j)(x-a_j+ib_j) \quad (5.49)$$

$$= (x-a_j)^2+b_j^2 \quad (5.50)$$

$$= x^2-2a_jx+a_j^2+b_j^2 \quad (5.51)$$

Thus $\alpha_j = -2a_j$ and $\beta_j = a_j^2+b_j^2 = (\alpha_j/2)^2+b_j^2$. Hence we have the following result.

Theorem 5.12 Let $p(x)$ be a non-constant polynomial of order n with real coefficients. Then $p(x)$ may be uniquely factored as

$$p(x) = c(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)(x^2+\alpha_1 x+\beta_1)\cdots(x^2+\alpha_k x+\beta_k) \quad (5.52)$$

where $n = m+2k$, $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, $\alpha_i, \beta_i \in \mathbb{R}$, and $\alpha_i^2 < 4\beta_j$. If $k > 0$ then the complex roots are

$$\Lambda_{j1,j2} = -\frac{\alpha_j}{2} \pm i\sqrt{\beta_j-(\alpha_j/2)^2} \quad (5.53)$$

Topic 6

Span and Linear Independence

Definition 6.1 Let $\mathbb{V}$ be a vector field over $\mathbb{F}$ and let $s = (v_1, v_2, \ldots, v_n)$ be a list of vectors in $\mathbb{V}$. Then a **linear combination of** s is any vector of the form

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \tag{6.1}$$

where $a_1, \ldots, a_n \in \mathbb{F}$. The **span of** $(v_1, \ldots, v_n)$ is the set of all linear combinations of $(v_1, \ldots, v_n)$,

$$\text{span}(v_1, \ldots, v_n) = \{a_1 v_1 + \cdots + a_n v_n | a_1, \ldots, a_n \in \mathbb{F}\} \tag{6.2}$$

We define the **span of the empty list** as span() $= \{0\}$.

We say $(v_1, \ldots, v_n)$ **spans** $\mathbb{V}$ if $\mathbb{V} = \text{span}(v_1, \ldots, v_n)$.

A vector space is **finite dimensional** if $\mathbb{V} = \text{span}(v_1, \ldots, v_n)$ for some $v_1, \ldots, v_n \in \mathbb{V}$.[1] A vector space that is not finite dimensional is called **infinite dimensional.**

[1] Recall that a list has a finite length, by definition.

Example 6.1 $\text{span}(1, z, z^2, \ldots, z^m) = \mathbb{P}_m(\mathbb{F}))$.

Example 6.2 $\mathbb{P}(F)$ is infinite dimensional.

Theorem 6.2 The span of any list of vectors in $\mathbb{V}$ is a subspace of $\mathbb{V}$.

Proof. (Exercise.) □

Theorem 6.3 $\text{span}(v_1, \ldots, v_n)$ is the smallest subspace of $\mathbb{V}$ that contains all the vectors in the list $(v_1, \ldots, v_n)$.

Proof. (Exercise.) □

Theorem 6.4 $\mathbb{P}_m(\mathbb{F})$ is a subspace of $\mathbb{P}(\mathbb{F})$

Proof. (Exercise.) □

Definition 6.5 A list of vectors $(v_1, \ldots, v_n)$ in $\mathbb{V}$ is said to be **linearly independent** if

$$a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0 \iff a_1 = a_2 = \cdots = a_n = 0 \tag{6.3}$$

for $a_i \in \mathbb{F}$, and is called **linearly dependent** iff there exists $a_1, \ldots, a_n$ not all zero such that

$$a_1v_1 + a_2v_2 + \cdots + a_nv_n = 0 \tag{6.4}$$

Example 6.3 The list $((1, 5), (2, 2))$ is linearly independent because the only solution to

$$a(1, 5) + b(2, 2) = 0 \tag{6.5}$$

is $a = b = 0$. To see this, suppose that there is a solution; then

$$\left.\begin{aligned} a + 2b &= 0 \\ 5a + 2b &= 0 \end{aligned}\right\} \tag{6.6}$$

Subtracting the first equation from the second gives $4a = 0 \implies a = 0$; hence (from either equation), $b = 0$.

Example 6.4 The list $((1, 2), (-1, 3), (0, -1))$ is linear dependent because

$$(1)(1, 2) + (1)(-1, 3) + (5)(0, -1) = 0 \tag{6.7}$$

Theorem 6.6 Linear Dependence Lemma. Let $(v_1, \dots, v_n)$ be linearly dependent in $\mathbb{V}$ with $v_1 \neq 0$. Then for some integer j, $2 \leq j \leq n$,

(a) $v_j \in \text{span}(v_1, \dots, v_{j-1})$

(b) If the j^{th} term is removed from $(v_1, \dots, v_n)$ then the span of the remaining lists equals $\text{span}(v_1, \dots, v_n)$.

Proof. (a) Let $(v_1, \dots, v_n)$ be linearly dependent with $v_1 \neq 0$. Then $\exists a_1, \dots, a_n$, not all 0, such that

$$0 = a_1v_1 + a_2v_2 + \cdots + a_nv_n \tag{6.8}$$

Let j be the largest integer $2 \leq j \leq n$ such that $a_j \neq 0$. At least one $a_j, j > 1$ must be nonzero because $a_1 \neq 0$. Then, since $a_{j+1} = a_{j+2} = \cdots = a_n = 0$,

$$0 = a_1v_1 + a_2v_2 + \cdots + a_jv_j \tag{6.9}$$

hence (since $a_j \neq 0$),

$$v_j = -\frac{a_1}{a_j}v_1 - \cdots - \frac{a_{j-1}}{a_j}v_{j-1} \tag{6.10}$$

This proves (a).

To prove (b), let $u \in \text{span}(v_1, \dots, v_n)$. Then for some numbers $c_1, \dots, c_n \in \mathbb{F}$,

$$u = c_1 v_1 + \cdots + c_j v_j + \cdots + c_n v_n \tag{6.11}$$

$$= c_1 v_1 + \cdots + c_j \left(-\frac{a_1}{a_j} v_1 - \cdots - \frac{a_{j-1}}{a_j} v_{j-1} \right) + \cdots + c_n v_n \tag{6.12}$$

which does not depend on v_j, proving (b). □

Theorem 6.7 Let $\mathbb{V}$ be a finite-dimnsional vector space. Then the length of every linearly independent list of vectors in $\mathbb{V}$ is less than or equal to the length of every spanning list of vectors in $\mathbb{V}$.

Proof. Let $(u_1, \dots, u_m)$ be linearly independant in $\mathbb{V}$ and let $(w_1, \dots, w_n)$ be any spanning list of vectors of $\mathbb{V}$. We need to show that $m \leq n$.

Since $(w_1, \dots, w_n)$ spans $\mathbb{V}$ then there are some numbers $a_1, \dots, a_n$ such that

$$u_1 = a_1 w_1 + \cdots + a_n w_n \tag{6.13}$$

hence

$$0 = -(1)u_1 + w_1 + \cdots + a_n w_n \tag{6.14}$$

which tells us that the list

$$b_1 = (u_1, w_1, w_2, \dots, w_n) \tag{6.15}$$

is linearly dependent.

By the linear dependence theorem, we can remove one of the w_i so that the remaining elements in 6.15 spans $\mathbb{V}$. Call the vector removed w_{i_1}, and define

$$B_1 = (u_1, w_1, w_2, \dots, w_n | w_{i_1}) \tag{6.16}$$

where by $(a, b, ...|p, q, ...)$ we mean the list $(a, b, ..)$ with the elements $(p, q, ..)$ removed.

Since B spans $\mathbb{V}$, if we add any vector to it from $\mathbb{V}$, it becomes linearly dependent. For example

$$b_2 = (u_1, u_2, w_1, w_2, \dots, w_n | w_{i_1}) \tag{6.17}$$

is linearly dependent. Applying the linear dendence theorem a second time, we can remove one of the elements and the list still spans $\mathbb{V}$. Since the u_i are linearly independent, it must be one of the w_j, call it w_{i_2}. So the list

$$B_2 = (u_1, u_2, w_1, \dots, w_n | w_{i_1}, w_{i_2}) \tag{6.18}$$

spans $\mathbb{V}$.

We keep repeating this process. In each step, we add one u_k and remove one w_j. If at some point we do not have any w's to remove this must mean that we have created a list that only contains the u's but is linearly dependent. This is a contradition. Hence there must always be at least one w to remove. Thus there must be at least as many w's as there are u's. Hence $m \leq n$. □

Theorem 6.8 Every subspace of a finite dimensional vector space is finite dimensional.

Proof. Let $\mathbb{V}$ be finite dimensional, and let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Since $\mathbb{V}$ is finite dimensional, for some m, there exists $w_1, \dots, w_m \in \mathbb{V}$ such that

$$L = (w_1, \dots, w_m) \tag{6.19}$$

where $\mathbb{V} = \text{span}(L)$.

If $\mathbb{U} = \{0\}$ then we are done.

If $\mathbb{U} \neq \{0\}$ then there exists at least one $v_1 \in \mathbb{U}$ such that $v_1 \neq 0$. Define

$$B = (v_1) \tag{6.20}$$

Either $\mathbb{U} = \text{span}(B)$ or $\mathbb{U} \neq \text{span}(B)$

If $\mathbb{U} = \text{span}(B)$ then $\mathbb{U}$ is finite dimensional and the proof is complete.

If $\mathbb{U} \neq \text{span}(B)$, and there is some $v_2 \in \mathbb{U}$ such that

$$v_2 \notin \text{span}(B) \tag{6.21}$$

Append this vector to B so that

$$B = (v_1, v_2) \tag{6.22}$$

We then keep repeating this process. If $U = \text{span}(B)$ we stop, having completed the proof; or, because $U \neq \text{span}(B)$, there is some vector in $\mathbb{U}$ that is not in the span(B)

that we can append to B. After n steps we will have

$$B = (v_1, v_2, \ldots, v_n) \tag{6.23}$$

Either the process stops with $n < m$ or it does not. If it does, we are done.

If it does not then when there are $n = m$ vectors in B. It is not possible to find any other vector $u \in \mathbb{U}$ at this point such that

$$B' = (v_1, \ldots, v_m, u) \tag{6.24}$$

is linearly independent. If there was such a vector, then we would have a linearly independent list that has length $m + 1 > m$, and we already have a list of length m that spans $\mathbb{U}$, namely L. But it is not possible to have a linearly independent set with length greater than m, by theorem 6.7.

Hence when $m = n$ every vector in $\mathbb{U}$ can be expressed as a linear combination of B. Thus B spans $\mathbb{U}$ and since the length of B is finite, then $\mathbb{U}$ is finite dimensional. □

Topic 7

Bases and Dimension

Definition 7.1 A **basis** of $\mathbb{V}$ is a list of linearly indpendent vectors in $\mathbb{V}$ that spans $\mathbb{V}$.

The **standard basis in** $\mathbb{F}^n$ is

$$((1, 0, \ldots, 0), (0, 1, 0, \ldots, 0), (0, 0, 1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)) \tag{7.1}$$

Example 7.1 $(1, z, z^2, \ldots, z^m)$ is a basis of $\mathbb{P}_m$.

Theorem 7.2 Let $v_1, \ldots, v_n \in \mathbb{V}(\mathbb{F})$. Then $(v_1, \ldots, v_n)$ is a basis of $\mathbb{V}$ if and only if every $v \in \mathbb{V}$ can be written uniquely in the form

$$v = a_1 v_1 + \cdots + a_n v_n \tag{7.2}$$

for some $a_1, \ldots, a_n \in \mathbb{F}$.

Proof. ($\Longrightarrow$) Suppose that $B = (v_1, \ldots, v_n)$ is a basis of $\mathbb{V}$.

Let $v \in \mathbb{V}$. Then because B spans $\mathbb{V}$ there exists some $a_1, \dots, a_n \in \mathbb{F}$ such that

$$v = a_1 v_1 + \cdots + a_n v_n \tag{7.3}$$

We must show that the numbers $a_1, \dots, a_n$ are unique. To do those, suppose that there is a second set of numbers $b_1, \dots, b_n \in \mathbb{F}$ such that

$$v = b_1 v_1 + \cdots + b_n v_n \tag{7.4}$$

Subtracting,

$$0 = (a_1 - b_1)v_1 + (a_2 - b_2)v_2 + \cdots + (a_n - b_n)v_n \tag{7.5}$$

Since B is a basis, it is linearly independent. Hence every coefficent in equation 7.5 must be zero, i.e., $a_i = b_i, i = 1, \dots, n$. Thus the representation is unique.

($\Longleftarrow$) Suppose that every $v \in \mathbb{V}$ can written uniquely in the form of equation 7.3.

Then by definition of a spanning list, $(v_1, \dots, v_n)$ spans $\mathbb{V}$. To show that B is a basis of $\mathbb{V}$ we need to also show that it is linearly independent.

Suppose that B is linearly dependent. Then there exist $b_1, \dots, b_n \in \mathbb{F}$ such that

$$0 = b_1 v_1 + \cdots + b_n v_n \tag{7.6}$$

By uniqueness (which we are assuming), this is the only set of b_i for which this is true. But we also know that

$$0 = (0)v_1 + \cdots + (0)v_n \tag{7.7}$$

as a different representation of the vector 0. Hence $b_i = 0, i = 1, \dots, n$. Thus the list

B is linearly independent, and hence a basis of $\mathbb{V}$. □

Theorem 7.3 Let S be a list of vectors that spans $\mathbb{V}$. Then either S is a basis of $\mathbb{V}$ or it can be reduced to a basis of S by removing some elements of S.

Proof. Let $S = (v_1, \ldots, v_n)$ span $\mathbb{V}$. Let $B = S$.

For each element in B, starting with $i = 1$,

(1) If $v_i = 0$, remove it from B and reindex B.

(2) If $v_i \in \text{span}(v_1, \ldots, v_i - 1)$, remove v_i from B and re-index.

At each step, we only remove a vector if it is in the span of vectors to the left of it in B. Hence what remains still spans $\mathbb{V}$.

Let $m = \text{length}(B)$ after the process is finished.

Since no vector is in the span of any vector to the left of it, v_m is not in the span of $(v_1, \ldots, v_{m-1})$. Hence there is no way to write v_m as a linear combination of the list $(v_1, \ldots, v_{m-1})$ and therefore $(v_1, \ldots, v_m)$ is linearly independent.

Hence B spans $\mathbb{V}$ and is linearly independent, hence it is a basis of $\mathbb{V}$. □

Theorem 7.4 Every finite dimensional vector space has a basis.

Proof. Let $\mathbb{V}$ be finite dimensional. Then it has a spanning list. This list can be reduced to a basis by theorem 7.3. Thus $\mathbb{V}$ has a basis. □

Theorem 7.5 Let $S = (v_1, \ldots, v_m)$ be a linearly independent list of vectors in a finite dimensional vector space $\mathbb{V}$. Then S can be extended to a basis of $\mathbb{V}$.

Proof. Let $W = (w_1, \ldots, w_n)$ be any list of vectors that spans $\mathbb{V}$.

If $w_1 \in \text{span}(S)$, let $B = S$. Otherwise, let $B = (v_1, \ldots, v_m, w_1)$.

For each $j = 2, \ldots, n$, repeat this process: if $w_j \in \text{span}(B)$, then ignore w_j; otherwise append w_j to B.

At each step, B is still linearly indepenent, because we are adding a vector that is not in the previous span of B.

After the n^{th} step, every w_i is either in B or in the span of B. Thus every $w_j \in \text{span}(B)$.

Since W spans $\mathbb{V}$, every vector $v \in \mathbb{V}$ can be expressed as a linear combination of the w_i. Since every $w_i \in \text{span}(B)$, then every w_i can be expressed as a linear combination of B. Hence every vector $v \in \mathbb{V}$ can be expressed as a linear combination of B. Hence B spans $\mathbb{V}$.

Since B spans $\mathbb{V}$ and B is linearly independent, it is a basis if $\mathbb{V}$. □

Theorem 7.6 Let $\mathbb{V}$ be a finite dimensional vector space, and let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Then there is a subspace $\mathbb{W}$ of $\mathbb{V}$ such that $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$.

Proof. Since any subset of a finite-dimensional vector space is finite-dimensional, then $\mathbb{U}$ is finite-dimensional (Theorem 6.8).

Since every finite-dimensional vector space has a basis, then $\mathbb{U}$ has a basis $B = (u_1, \ldots, u_m)$ (Theorem 7.4).

Since B is a linearly independent list of vectors in $\mathbb{V}$, it can be extended to a basis $B' = (u_1, \ldots, u_m, w_1, \ldots, w_n)$ of $\mathbb{V}$ (Theorem 7.5).

Let $\mathbb{W} = \text{span}(w_1, \ldots, w_n)$. Then $\mathbb{W}$ is a subspace of $\mathbb{V}$.

To prove that $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ we must show that (a) $V = U + W$; and (b) $\mathbb{U} \cap \mathbb{W} = \{0\}$.

To prove (a): let $v \in \mathbb{V}$. Since B' spans $\mathbb{V}$ there exist some numbers $a_1, \ldots, a_m, b_1, \ldots, b_n$

such that

$$v = \underbrace{a_1u_1 + \cdots + a_mu_m}_{\in \mathbb{U}} + \underbrace{b_1w_1 + \cdots + b_nw_n}_{\in \mathbb{W}} = u + w \tag{7.8}$$

where

$$u = a_1u_1 + \cdots + a_mu_m \in \mathbb{U} \tag{7.9}$$

$$w = b_1w_1 + \cdots + b_nw_n \in \mathbb{W} \tag{7.10}$$

Hence $v \in \mathbb{U} + \mathbb{W}$, which proves (a).

To prove (b): let $v \in \mathbb{U} \cap \mathbb{W}$. Then $v \in \mathbb{U}$ and $v \in \mathbb{W}$. Hence there are some numbers $a_1, \dots, a_n$ such that

$$v = a_1u_1 + \cdots + a_mu_m \tag{7.11}$$

and some numbers $b_1, \dots, b_n$ such that

$$v = b_1w_1 + \cdots + b_nw_n \tag{7.12}$$

Setting the two expression for v equal to one another,

$$a_1u_1 + \cdots + a_mu_m = b_1w_1 + \cdots + b_nw_n \tag{7.13}$$

and by rearrangment,

$$a_1u_1 + \cdots + a_mu_m - b_1w_1 - \cdots - b_nw_n = 0 \tag{7.14}$$

which means that a linear combination of B' is equal to zero. Since B' is a basis and is hence linearly independent, the only way that this is possible if if all the coefficients are zero, namely

$$a_1 = \cdots = a_m = b_1 = \cdots = b_n = 0 \tag{7.15}$$

Thus (by substitution of either all the a_i or all the b_i),

$$v = 0 \tag{7.16}$$

But we chose v as an arbitrary element of $U \cap \mathbb{W}$. Hence $\mathbb{U} \cap \mathbb{W} = \{0\}$, which proves (b).

Since $\mathbb{V} = \mathbb{U} + \mathbb{W}$ and $\mathbb{U} \cap \mathbb{W} = \{0\}$, we conclude by theorem 4.8 that $\mathbb{V} = \mathbb{U} + \mathbb{W}$. □

Theorem 7.7 Let $\mathbb{V}$ be any finite dimensional vector space, and let B_1 and B_2 be any two bases of $\mathbb{V}$. Then

$$\text{length}(B_1) = \text{length}(B_2) \tag{7.17}$$

Proof. Since B_1 is linearly independent; and since B_2 spans $\mathbb{V}$, we have by theorem 6.7

$$\text{length}(B_1) \leq \text{length}(B_2) \tag{7.18}$$

Similarly, since B_2 is linearly independent and B_1 spans $\mathbb{V}$,

$$\text{length}(B_2) \leq \text{length}(B_1) \tag{7.19}$$

Equation 7.17 follows. □

Definition 7.8 The **dimension** of a finite-dimensional vector space $\mathbb{V}$, denoted by $\dim(V)$ is the length of any basis $\mathbb{V}$. More specifically, if B is any basis of $\mathbb{V}$, then

$$\dim(\mathbb{V}) = \text{length}(B) \tag{7.20}$$

Theorem 7.9 Let $\mathbb{V}$ be any finite dimensional vector space and let $\mathbb{U}$ be any subspace of $\mathbb{V}$. Then

$$\dim(\mathbb{U}) \leq \dim(\mathbb{V}) \tag{7.21}$$

Proof. Let B be a basis of $\mathbb{U}$. Then B is a linearly independent list in $\mathbb{V}$. By theorem 7.5 B can be extended to a basis in $\mathbb{V}$. Let B' be any basis of $\mathbb{V}$ obtained by extending B. Then

$$\dim(\mathbb{U}) = \text{length}(B) \leq \text{length}(B') = \dim(\mathbb{V}) \tag{7.22}$$

□

Theorem 7.10 Let $\mathbb{V}$ be a finite-dimensional vector space, and let B be a list of vectors in $\mathbb{V}$ such that (a) $\text{span}(B) = \mathbb{V}$ and (b) $\text{length}(B) = \dim(\mathbb{V})$. Then B is a basis of $\mathbb{V}$.

Proof. Since $\text{span}(B) = \mathbb{V}$, then either B is already a basis of $\mathbb{V}$ or B can be reduced to a basis of $\mathbb{V}$ by theorem 7.3. But since every basis of $\mathbb{V}$ has length $\dim(\mathbb{V})$, no elements of B can be removed, else the basis produced by removing elements would be shorter than $\dim(\mathbb{V})$. Hence B is already a basis of $\mathbb{V}$. □

Theorem 7.11 Let $\mathbb{V}$ be a finite-dimensional vector space, and let B be a linearly independent list of vectors in $\mathbb{V}$ such that $\text{length}(B) = \dim(\mathbb{V})$. Then B is a basis of $\mathbb{V}$.

Proof. Since B is linearly independent it can be extended to a basis of $\mathbb{V}$ by theorem

7.5. Let B' be any such basis.

Since every basis has length $\dim(\mathbb{V})$ then

$$\text{length}(B') = \dim(\mathbb{V}) = \text{length}(B) \tag{7.23}$$

Hence it is not necessary to add any vectors to B to make it a basis. Hence B is already a basis of $\mathbb{V}$. □

Theorem 7.12 Let $\mathbb{V}$ be a finite-dimensional vector space, and let $\mathbb{U}$ and $\mathbb{W}$ be subspaces of $\mathbb{V}$. Then

$$\dim(\mathbb{U} + \mathbb{W}) = \dim(\mathbb{U}) + \dim(\mathbb{W}) - \dim(\mathbb{U} \cap \mathbb{W}) \tag{7.24}$$

Proof. Let

$$B = (v_1, \ldots, v_m) \tag{7.25}$$

be a basis of $\mathbb{U} \cap \mathbb{W}$, hence

$$\dim(\mathbb{U} \cap \mathbb{W}) = m \tag{7.26}$$

By theorem 7.5, B can be extended to a basis B_U of $\mathbb{U}$,

$$B_U = (v_1, \ldots, v_m, u_1, \ldots, u_j) \tag{7.27}$$

and to a basis B_W of W,

$$B_W = (v_1, \ldots, v_m, w_1, \ldots, w_k) \tag{7.28}$$

so that

$$\left.\begin{aligned} \dim(\mathbb{U}) &= m + j \\ \dim(\mathbb{W}) &= m + k \end{aligned}\right\} \tag{7.29}$$

for some integers m, j, and k. Let

$$B' = (v_1, \ldots, v_m, u_1, \ldots, u_j, w_1, \ldots, w_k) \tag{7.30}$$

Then $\mathbb{U} \subset \text{span}(B')$ and $\mathbb{W} \subset \text{span}(B')$ hence $\mathbb{U} + \mathbb{W} \subset \text{span}(B')$.

Furthermore, B' is linearly independent. [To see this, suppose that there exist scalars $a_1, \ldots, a_m, b_1, \ldots, b_j, c_1, \ldots, c_k$ such that

$$a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_j u_j + c_1 w_1 + \cdots + c_k w_k = 0 \tag{7.31}$$

By rearrangement,

$$\underbrace{a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_j u_j}_{\in \mathbb{U}} = \underbrace{-c_1 w_1 - \cdots - c_k w_k}_{\in \mathbb{W}} \tag{7.32}$$

Thus since the left hand side is a linear combination of B_U,

$$w = c_1 w_1 + \cdots + c_k w_k \in \mathbb{U} \tag{7.33}$$

But all the $w_i \in \mathbb{W}$; hence $w \in \mathbb{U} \cap \mathbb{W}$.

Thus there exist scalars $d_1, \ldots, d_m$ such that

$$w = c_1 w_1 + \cdots + c_k w_k = d_1 v_1 + \cdots + d_m v_m \tag{7.34}$$

because $B = (v_1, \dots, v_m)$ is a basis of $\mathbb{U} \cap \mathbb{W}$. Rearranging,

$$c_1 w_1 + \cdots + c_k w_k - d_1 v_1 - \cdots - d_m v_m = 0 \tag{7.35}$$

Since $B_W = (v_1, \dots, v_m, w_1, \dots, w_k)$ is a basis of W it is a linearly independent set, hence

$$c_1 = \cdots = c_k = d_1 = \cdots = d_m = 0 \tag{7.36}$$

Substituting back into equation 7.31,

$$a_1 v_1 + \cdots + a_m v_m + b_1 u_1 + \cdots + b_j u_j = 0 \tag{7.37}$$

Now since $B_U = (v_1, \dots, v_m, u_1, \dots, u_j)$ is a basis of W, it is a linearly independent set, which tells us that

$$a_1 = \cdots = a_m = b_1 = \cdots = b_j = 0 \tag{7.38}$$

Hence there is no linear combination of B' that gives the zero vector except for the one in which all the coefficients are zero. This means the B' is a linearly independent list.]

Since B' is linearly independent and spans $\mathbb{U} + \mathbb{W}$, it is a basis of $\mathbb{U} + \mathbb{W}$. Hence

$$\dim(\mathbb{U} + \mathbb{W}) = \text{length}(B') \tag{7.39}$$

$$= m + j + k \tag{7.40}$$

$$= (m + j) + (m + k) - m \tag{7.41}$$

$$= \dim(\mathbb{U}) + \dim(\mathbb{W}) - \dim(\mathbb{U} \cap \mathbb{W}) \quad \square \tag{7.42}$$

Theorem 7.13 Let $\mathbb{V}$ be a finite dimensional vector space, and suppose that $\mathbb{U}_1, \dots, \mathbb{U}_m$ are subspaces of $\mathbb{V}$ such that

$$\mathbb{V} = \mathbb{U}_1 + \cdots + \mathbb{U}_m \tag{7.43}$$

and

$$\dim(\mathbb{V}) = \dim(\mathbb{U}_1) + \cdots + \dim(\mathbb{U}_m) \tag{7.44}$$

Then

$$\mathbb{V} = \mathbb{U}_1 \oplus \cdots \oplus \mathbb{U}_m \tag{7.45}$$

Proof. Proof. Define bases $B_1, \dots, B_m$ for each of the $\mathbb{U}_i$. Let[1]

$$B = (B_1, B_2, \dots, B_n) \tag{7.46}$$

Then B spans $\mathbb{V}$ (by equation 7.43), and

$$\begin{aligned} \text{length}(B) &= \text{length}(B_1) + \cdots + \text{length}(B_m) && (7.47)\\ &= \dim(\mathbb{U}_1) + \cdots + \dim(\mathbb{U}_m) && (7.48)\\ &= \dim(\mathbb{V}) && (7.49) \end{aligned}$$

by equation 7.44. Hence B is a basis of $\mathbb{V}$.

Let $u_1 \in \mathbb{U}_1, \dots, u_m \in \mathbb{U}_m$ be chosen such that

$$0 = u_1 + \cdots + u_m \tag{7.50}$$

We can express each u_i in terms of the basis vectors of U_i. Suppose that $B_i =$

[1] By $((a_1, a_2, \dots), (b_1, b_2, \dots))$ we will mean $(a_1, a_2, \dots, b_1, b_2, \dots)$.

$(v_{i1}, v_{i2}, \dots, v_{i\,\dim(\mathbb{U}_i)})$. Then for some scalars $a_{i1}, \dots, a_{i\,\dim(U_i)})$

$$u_i = \sum_{k=1}^{\dim(\mathbb{U}_i)} a_{ik} v_{ik} \tag{7.51}$$

Hence

$$0 = \sum_{k=1}^{\dim(\mathbb{U}_1)} a_{1k} v_{1k} + \sum_{k=1}^{\dim(\mathbb{U}_2)} a_{2k} v_{2k} + \cdots + \sum_{k=1}^{\dim(\mathbb{U}_m)} a_{mk} v_{mk} \tag{7.52}$$

But the right hand side is a linear combination of elements of B, which is a basis of $\mathbb{V}$. Hence every coefficient is zero, which means every $u_j = 0$ in equation 7.50. Thus the only way to write zero as a sum of vectors in each of the U_i is if each u_i is zero. By theorem 4.7 this means that Then

$$\mathbb{V} = \mathbb{U}_1 \oplus \cdots \oplus \mathbb{U}_m \tag{7.53}$$

□

Topic 8

Linear Maps

Definition 8.1 Let $\mathbb{V}$, W be vector spaces over a field $\mathbb{F}$. Then a **linear map from $\mathbb{V}$ to $\mathbb{W}$** is a function $T : \mathbb{V} \mapsto \mathbb{W}$ with the properties:

(1) Additivity: for all $u, v \in \mathbb{V}$,

$$T(u + v) = T(u) + T(v) \tag{8.1}$$

(2) Homogeneity: for $a \in \mathbb{F}$ and for all $v \in \mathbb{V}$,

$$T(av) = aT(v) \tag{8.2}$$

It is common notation to omit the parenthesis when expressing maps, writing $T(v)$ as Tv. The reason for this will become clear when we study the matrix representation of linear maps.

The properties of additivity and homogeneity can be combined into a **linearity prop-**

erty expressed as

$$T(au + bv) = aTu + bTv \tag{8.3}$$

where $a, b \in \mathbb{F}$ and $u, v \in \mathbb{V}$.

Definition 8.2 The **set of all linear maps from** $\mathbb{V}$ **to** $\mathbb{W}$ is denoted by $\mathcal{L}(\mathbb{V}, \mathbb{W})$

Definition 8.3 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ be a linear map $T : \mathbb{V} \mapsto \mathbb{W}$. The **range of** T is the subset of $\mathbb{W}$ that are mapped to by T:

$$\text{range}(T) = \{w \in \mathbb{W} | w = Tv, v \in \mathbb{V}\} \tag{8.4}$$

Example 8.1 Define $T : \mathbb{V} \mapsto \mathbb{V}$ by $Tv = 0$; here $T \in \mathcal{L}(V, V)$. The range of T is $\{0\}$.

Example 8.2 Define $T : \mathbb{V} \mapsto \mathbb{V}$ by $Tv = v$. This is called the **Identity Map** and has the special symbol I reserved for it:

$$Iv = v \tag{8.5}$$

The range of I is $\mathbb{V}$.

Example 8.3 Differentiation: Define $\mathcal{D} \in \mathcal{L}(\mathbb{P}(\mathbb{R}), \mathbb{P}(\mathbb{R}))$ by $\mathcal{D}p = p'$ where $p'(x) = dp/dx$.

Example 8.4 Integration: Define $\mathcal{I} \in \mathcal{L}(\mathbb{P}(\mathbb{R}), \mathbb{R}))$ by

$$\mathcal{I}p = \int_0^1 p(x)dx \tag{8.6}$$

Remark 8.4 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces and $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ be a linear map, and let $B = (v_1, \ldots, v_n)$ be a basis of $\mathbb{V}$. Then Tv is completely determinded by the values of Tv_i, because $\forall v \in \mathbb{V}$, there exist $a_i, i = 1, \dots, n$ such that

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \tag{8.7}$$

Hence by linearity

$$Tv = a_1 Tv_1 + a_2 Tv_2 + \cdots + a_n Tv_n \tag{8.8}$$

Remark 8.5 Linear maps can be made to take on arbitrary values on a basis. Let $B = (v_1, \dots, v_n)$ be a basis of $\mathbb{V}$ and let $w_1, \dots, w_n \in \mathbb{W}$. Then we can construct a linear map such that $Tv_1 = w_1, Tv_2 = w_2, \dots, Tv_n = w_n$ by

$$T(a_1 v_1 + \cdots + a_n v_n) = a_1 w_1 + \cdots + a_n w_n \tag{8.9}$$

Definition 8.6 Addition of Linear Maps. Let $\mathbb{V}, \mathbb{W}$ be vector spaces and let $S, T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. The we define the map $(S + T) : \mathbb{V} \mapsto \mathbb{W}$ by

$$(S + T)v = Sv + Tv \tag{8.10}$$

which must hold for all $v \in \mathbb{V}$.

Definition 8.7 Scalar Multiplication of Linear Maps. Let $\mathbb{V}, \mathbb{W}$ be vector spaces let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. The we define the map $aT : \mathbb{V} \mapsto \mathbb{W}$ by

$$(aT)v = a(Tv) \tag{8.11}$$

which must hold for all $a \in \mathbb{F}$ and all $v \in \mathbb{V}$.

Theorem 8.8 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces. Then $\mathcal{L}(\mathbb{V}, \mathbb{W})$ is a vector space under addition and scalar multiplication of linear maps.

Proof. Exercise. You need to prove closure; commutativity and associativity of addition; existence of additive identity and inverse; existence of identity for scalar multiplication; and distributivity. □

Definition 8.9 Product of Linear Maps. Let $\mathbb{U}, \mathbb{V}, \mathbb{W}$ be vector spaces, $T \in \mathcal{L}(\mathbb{U}, \mathbb{V})$ and $S \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then we define $ST \in \mathcal{L}(U, W)$ by

$$(ST)v = S(Tv) \tag{8.12}$$

We normally drop the parenthesis and write this as ST.

Remark 8.10 Multiplication of linear maps is associative

$$(ST)U = S(TU) = STU = S(T(U(v))) \tag{8.13}$$

and distributive

$$\left.\begin{aligned} (S+T)U &= SU + TU \\ S(T+U) &= ST + SU \end{aligned}\right\} \tag{8.14}$$

but is not commutative. In fact, even if ST is defined, TS might be undefined.

Remark 8.11 The product ST only makes sense if domain(S) = range(T), i.e, if

$$S : \mathbb{U} \mapsto \mathbb{V} \text{ and } T : \mathbb{V} \mapsto \mathbb{W} \implies ST : \mathbb{U} \mapsto \mathbb{W} \tag{8.15}$$

The product ST is analogous to the matrix product ST which only makes sense if

$$S \text{ is } [m \times p] \text{ and } T \text{ is } [p \times n] \implies ST \text{ is } [m \times n] \tag{8.16}$$

When written in this way (ordered as $\mathbb{U}, \mathbb{V}, \mathbb{W}$ (or m, p, n)), the "middle" vector space (or middle dimension) must be the same, and the final mapping is between the "outer" vector space (or dimensions).

Remark 8.12 Sometimes the product ST is written as the composition of the linear operators, $S \circ T$.

Definition 8.13 Let $\mathbb{V}, \mathbb{W}$ be vector spaces and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then the **null space of** T is defined as the subset of $\mathbb{V}$ that T maps to zero:

$$\text{null}(T) = \{v \in \mathbb{V} | Tv = 0\} \tag{8.17}$$

Theorem 8.14 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then null$(T)$ is a subspace of $\mathbb{V}$.

Proof. By additivity,

$$T(0) = T(0 + 0) = T(0) + T(0) \tag{8.18}$$

$$T(0) = 0 \tag{8.19}$$

hence $0 \in \text{null}(T)$.

Now let $u, v \in \text{null}(T)$.

$$T(u + v) = Tu + Tv = 0 + 0 = 0 \tag{8.20}$$

Hence $u, \in \text{null}(T) \implies u + v \in \text{null}(T)$.

Finally, let $u \in \text{null}(T)$ and $a \in \mathbb{F}$.

$$T(au) = aTu = a(0) = 0 \tag{8.21}$$

so that $u \in \text{null}(T) \implies au \in \text{null}(T)$.

Thus null(T) contains 0 and is closed under addition and scalar multiplication. Hence it is a subspace of $\mathbb{V}$. □

Definition 8.15 Let $\mathbb{V}$ and $\mathbb{W}$ be vectors spaces and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then T is called **injective** or **one-to-one** if whenever $u, v \in \mathbb{V}$,

$$Tu = Tv \implies u = v \tag{8.22}$$

Definition 8.16 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then T is called **surjective** or **onto** if range(T) = $\mathbb{W}$, i.e, if $\forall w \in \mathbb{W},\ \exists v \in \mathbb{V} \ \ni\ w = Tv$.

Theorem 8.17 Let $\mathbb{V}$ and $\mathbb{W}$ be vectors spaces and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then T is injective if and only if null(T) = $\{0\}$.

Proof. ($\implies$) Suppose that T is injective. By equation 8.19 we known that $T(0) = 0$ hence

$$\{0\} \subset \text{null}(T) \tag{8.23}$$

Let $v \in \text{null}(T)$; then $T(v) = 0 = T(0)$. Since T is injective, this means $v = 0$. Hence

$$\text{null}(T) \subset \{0\} \tag{8.24}$$

Thus

$$\text{null}(T) = \{0\} \tag{8.25}$$

($\impliedby$) Assume that equation 8.25 is true.

Let $u, v \in \mathbb{V}$ and suppose that $Tu = Tv$. Then

$$0 = Tu - Tv = T(u - v) \tag{8.26}$$

Hence

$$u - v \in \text{null}(T) = \{0\}. \tag{8.27}$$

Therefore $u - v = 0$ or $u = v$ which proves that T is injective (because $Tu = Tv \implies u = v$). □

Theorem 8.18 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then range$(T)$ is a subspace of $\mathbb{W}$.

Proof. By definition $R = \text{range}(T) \subset \mathbb{W}$. To show that R is a subspace of $\mathbb{W}$ we need to show that (a) $0 \in \mathbb{R}$; (b) R is closed under addition; and (c) R is closed under scalar multiplication.

By equation 8.19, $T(0) = 0$, hence $0 \in R$.

Let $w, z \in \text{range}(T)$. Then there exist some $u, v \in \mathbb{V}$ such that $T(u) = w$ and $T(v) = z$. Then

$$T(u + v) = Tu + Tv = w + z \tag{8.28}$$

so that $w + z \in \text{range}(T)$, proving (b).

Now let $w \in R$ and pick any $a \in \mathbb{F}$. Then since $w \in= R$ there is some $v \in \mathbb{V}$ such that $Tv = w$, so that

$$T(av) = aTv = aw \tag{8.29}$$

and hence $aw \in \text{range}(T)$ proving (c). □

Theorem 8.19 Let $\mathbb{V}$ be a finite dimensional vector space over $\mathbb{F}$, let $\mathbb{W}$ be a vector space over $\mathbb{F}$, and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then range$(T)$ is a finite-dimensional subspace of $\mathbb{W}$ and

$$\dim \mathbb{V} = \dim \operatorname{null}(T) + \dim \operatorname{range}(T) \tag{8.30}$$

Proof. Let $B = (u_1, \dots, u_m)$ be a basis of null(T). Hence

$$\dim \operatorname{null}(T) = m \tag{8.31}$$

We can extend B to a basis B' of $\mathbb{V}$, i.e., for some integer n,

$$B' = (u_1, \dots, u_m, w_1, \dots, w_n) \tag{8.32}$$

Hence

$$\dim \mathbb{V} = m + n \tag{8.33}$$

Let $v \in \mathbb{V}$. Because B' spans $\mathbb{V}$, there are scalars $a_1, \dots, b_n \in \mathbb{F}$ such that

$$v = a_1u_1 + \cdots + a_mu_m + b_1w_1 + \cdots + b_nw_n \tag{8.34}$$

Hence

$$\begin{aligned} Tv &= Ta_1u_1 + \cdots + Ta_mu_m + Tb_1w_1 + \cdots + Tb_nw_n && (8.35)\\ &= a_1Tu_1 + \cdots + a_mTu_m + b_1Tw_1 + \cdots + b_nTw_n && (8.36)\\ &= b_1Tw_1 + \cdots + b_nTw_n && (8.37) \end{aligned}$$

where the last line follows because the $u_i \in$ null(T).

Thus $\forall v \in \mathbb{V}$, Tv can be expressed as a linear combination of

$$B'' = (Tw_1, \dots, Tw_n) \tag{8.38}$$

Hence

$$\text{span}(B'') = \text{range}(T) \tag{8.39}$$

Since it is spanned by a finite list, B'', this tells us that range(T) is finite dimensional, since $\dim \text{range}(T)$ is no larger than the length of any spanning list.

Suppose that there exist $c_1, \dots, c_n \in \mathbb{F}$ such that

$$0 = c_1 T w_1 + \cdots + c_n T w_n \tag{8.40}$$

$$= T(c_1 w_1 + \cdots + c_n w_n) \tag{8.41}$$

This implies that

$$c_1 w_1 + \cdots + c_n w_n \in \text{null}(T) \tag{8.42}$$

But B is a basis of null(T) so any vector in null(T) is a linear combination of the elements of B. So there exist some scalars $d_1, \dots, d_m \in \mathbb{F}$ such that

$$c_1 w_1 + \cdots + c_n w_n = d_1 u_1 + \cdots + d_m u_m \tag{8.43}$$

or by rearrangement

$$c_1 w_1 + \cdots + c_n w_n - d_1 u_1 - \cdots + d_m u_m = 0 \tag{8.44}$$

But this is a linear combination of the elements of B', which is a basis of $\mathbb{V}$ and hence linearly independent. Thus

$$c_1 = \cdots = c_n = d_1 = \cdots = d_m = 0 \tag{8.45}$$

and thus (see equation 8.40) the only linear combination of the B'' that gives the zero vector is on in which all the coefficients are zero. This means that B'' is linearly independent.

Since B'' is linearly indpendent and spans range(T), it is a basis of range(T) and hence

$$\dim \operatorname{range}(T) = \operatorname{length}(B'') = n \tag{8.46}$$

Combining equations 8.46 with 8.31 and 8.33

$$\dim \mathbb{V} = m + n = \dim \operatorname{null}(T) + \dim \operatorname{range}(T) \tag{8.47}$$

□

Corollary 8.20 Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector spaces with

$$\dim \mathbb{V} > \dim \mathbb{W} \tag{8.48}$$

and let $T \in \mathcal{L}(V, W)$. Then T is not injective (one-to-one).

Proof. First, observe that since range$(T) \subset \mathbb{W}$,

$$\dim \operatorname{range}(T) \leq \dim \mathbb{W} \tag{8.49}$$

Then by theorem 8.19

$$\dim \operatorname{null}(T) = \dim \mathbb{V} - \dim \operatorname{range}(T) \geq \dim \mathbb{V} - \dim \mathbb{W} > 0 \tag{8.50}$$

Since $\dim \text{null}(T) > 0$ then $\text{null}(T)$ must contain vectors other than 0.

Hence (see theorem 8.17) T is not injective. □

Corollary 8.21 Let $\mathbb{V}$ and $\mathbb{W}$ be finite-dimensional vector spaces with

$$\dim \mathbb{V} < \dim \mathbb{W} \tag{8.51}$$

and let $T \in \mathcal{L}(V, W)$. Then T is not surjective (onto).

Proof. By theorem 8.19

$$\dim \text{range}(T) = \dim \mathbb{V} - \dim \text{null}(T) \leq \dim \mathbb{V} < \dim \mathbb{W} \tag{8.52}$$

Hence there are vectors in $\mathbb{W}$ that are not mapped to by T, and thus T is not surjective. □

Topic 9

Matrices of Linear Maps

We will denote the set of all $m \times n$ matrices with entries in $\mathbb{F}$ by $\mathrm{Mat}(m, n, \mathbb{F})$. Under the standard definitions of matrix addition and scalar multiplication, $\mathrm{Mat}(m, n, \mathbb{F})$ is a vector space.

Definition 9.1 We define the **Matrix of a Linear Map** for $T \in \mathcal{L}(V, W)$ as follows. Let $(v_1, \ldots, v_n)$ be a basis of $\mathbb{V}$ and let $(w_1, \ldots, w_m)$ be a basis of $\mathbb{W}$. Then

$$\mathcal{M}(T, (v_1, \ldots, v_n), (w_1, \ldots, w_m)) = [a_{ij}] \tag{9.1}$$

where

$$Tv_k = a_{1,k}w_1 + \cdots + a_{m,k}w_m \tag{9.2}$$

If the choice of bases for $\mathbb{V}$ and $\mathbb{W}$ are clear from the context we use the notation $\mathcal{M}(T)$.

The following may help you remember the structure of this matrix:

$$\begin{array}{c} \\ w_1 \\ \vdots \\ w_m \end{array} \begin{array}{c} \begin{array}{ccccc} v_1 & \cdots & v_k & \cdots & v_n \end{array} \\ \begin{bmatrix} & & a_{1,k} & & \\ & \cdots & \vdots & \cdots & \\ & & a_{m,k} & & \end{bmatrix} \end{array} \tag{9.3}$$

Theorem 9.2 Properties of $\mathcal{M}(T)$. Let $\mathbb{V}, \mathbb{W}$ be vector spaces over $\mathbb{F}$. Then

(1) Scalar Multiplication: Let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ and $c \in \mathbb{F}$.

$$\mathcal{M}(cT) = c\mathcal{M}(T) \tag{9.4}$$

(2) Matrix Addition: Let $T, S \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then

$$\mathcal{M}(T + S) = \mathcal{M}(T) + \mathcal{M}(S) \tag{9.5}$$

Proof. (sketch) Express the left hand side of each formula as a matrix and then apply the properties of matrices as reviewed in Chapter 1. □

To derive a rule for matrix multiplication, suppose that $S \in \mathcal{L}(\mathbb{U}, \mathbb{V})$ and $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. The composition $T \circ S$ is a map $TS : \mathbb{U} \mapsto \mathbb{W}$, i.e., $TS \in \mathcal{L}(\mathbb{U}, \mathbb{W})$:

$$\left.\begin{array}{c} S : \mathbb{U} \mapsto \mathbb{V} \quad \text{and} \quad T : \mathbb{V} \mapsto \mathbb{W} \\ \implies ST : \mathbb{U} \mapsto \mathbb{W} \end{array}\right\} \tag{9.6}$$

Let $(v_1, \dots, v_n)$ be a basis of $\mathbb{V}$, $(w_1, \dots, w_m)$ be a basis of $\mathbb{W}$, and $(U_1, \dots, u_p)$ be a basis of $\mathbb{U}$. Suppose that

$$\mathcal{M}(T) = [a_{i,j}]_{i\in\{1\dots m\},j\in\{1\dots n\}} \tag{9.7}$$

$$\mathcal{M}(S) = [b_{j,k}]_{j\in\{1\dots n\},k\in\{1\dots p\}} \tag{9.8}$$

Then

$$Su_k = \sum_j b_{j,k} v_j \tag{9.9}$$

$$Tv_j = \sum_i a_{i,j} w_i \tag{9.10}$$

and therefore

$$\begin{aligned} TSu_k &= T\sum_j b_{j,k} v_j &(9.11)\\ &= \sum_j b_{j,k} T v_j &(9.12)\\ &= \sum_j b_{j,k} \sum_i a_{i,j} w_i &(9.13)\\ &= \sum_i w_i \left(\sum_j a_{i,j} b_{j,k}\right) &(9.14) \end{aligned}$$

Therefore if we identify

$$\mathcal{M}(TS)_{i,k} = \sum_j a_{i,j} b_{j,k} \tag{9.15}$$

we can define matrix multiplication by

$$\mathcal{M}(TS) = \mathcal{M}(T)\mathcal{M}(S) \tag{9.16}$$

This is in fact the same definition of matrix multiplication with which you are already aquainted.

Definition 9.3 Matrix of a Vector. Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let $v \in \mathbb{V}$. If $(v_1, \cdots, v_n)$ is a basis of $\mathbb{V}$ then for some some $a_1, \ldots, a_n \in \mathbb{F}$

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \tag{9.17}$$

We **define the matrix of the vector** v as

$$\mathcal{M}(v) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \tag{9.18}$$

Theorem 9.4 Let $\mathbb{V}, \mathbb{W}$ be vector spaces over $\mathbb{F}$ with bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then for every $v \in \mathbb{V}$,

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v) \tag{9.19}$$

Proof. Let

$$\mathcal{M}(T) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \tag{9.20}$$

Then (see equation 9.2)

$$Tv_k = \sum_{j=1}^{m} a_{j,k} w_j \tag{9.21}$$

For any vector $v \in \mathbb{V}$ there are some scalars $b_1, \ldots, b_n \in \mathbb{F}$ such that

$$v = \sum_{k=1}^{n} b_k v_k \tag{9.22}$$

Hence

$$Tv = T\left(\sum_{k=1}^{n} b_k v_k\right) \tag{9.23}$$
$$= \sum_{k=1}^{n} b_k T v_k \tag{9.24}$$
$$= \sum_{k=1}^{n} b_k \sum_{j=1}^{m} a_{j,k} w_j \tag{9.25}$$
$$= \sum_{j=1}^{m} w_j \left(\sum_{k=1}^{n} a_{j,k} b_k\right) \tag{9.26}$$

Therefore

$$[\mathcal{M}(Tv)]_j = \sum_{k=1}^{n} a_{j,k} b_k \tag{9.27}$$

Similarly, since

$$\mathcal{M}(T)\mathcal{M}(v) = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} a_{1,k} b_k \\ \vdots \\ \sum_{k=1}^{n} a_{m,k} b_k \end{pmatrix} \tag{9.28}$$

we conclude that

$$[\mathcal{M}(T)\mathcal{M}(v)]_j = \sum_{k=1}^{n} a_{j,k} b_k = [\mathcal{M}(Tv)]_j \tag{9.29}$$

and therefore

$$\mathcal{M}(Tv) = \mathcal{M}(T)\mathcal{M}(v) \tag{9.30}$$

□

Topic 10

Invertibility of Linear Maps

Definition 10.1 Let $\mathbb{V}, \mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(V, W)$. Then T is called **invertible** if there exists a linear map $S \in \mathcal{L}(W, V)$ such that

$$\left.\begin{aligned} ST = I \in \mathbb{V} \\ TS = I \in \mathbb{W} \end{aligned}\right\} \tag{10.1}$$

The linear map S is called the **inverse** of T.

Theorem 10.2 The inverse is unique

Proof. Let T be a linear map with inverses S and S'. Then

$$S = SI = S(TS') = (ST)S' = IS' = S' \tag{10.2}$$

□

Notation: Since the index is unique we denote the inverse of T by T^{-1}:

$$TT^{-1} = T^{-1}T = I \tag{10.3}$$

Theorem 10.3 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$ and let $T \in \mathcal{L}(V, W)$. Then T is invertible if and only if T is both injective (one-to-one) and surjective (onto).

Proof. ($\Longrightarrow$) Assume that T is invertible.

Suppose that $u, v \in \mathbb{V}$ and that $Tu = Tv$. Then

$$u = T^{-1}(Tu) = T^{-1}Tu = T^{-1}Tv = v \tag{10.4}$$

Since $Tu = Tv \implies u = v$ we conclude that T is one-to-one (injective).

Now suppose that $w \in \mathbb{W}$. Then

$$w = TT^{-1}w = T\left(T^{-1}w\right) \tag{10.5}$$

Hence $w \in \text{range}(T) \implies \mathbb{W} \subset \text{range}(T)$. Since by definition of T, $\text{range}(T) \subset W$ we conclude that $\text{range}(T) = W$. Hence T is onto (surjective).

($\Longleftarrow$) Suppose that T is both injective and surjective. We must show that it is invertible. This requires showing that there exists a map S with three properties: (a)TS is the indentity; (b) ST is the identity; (c) S is linear.

Let $w \in \mathbb{W}$. Then since T is onto there is some $v \in \mathbb{V}$ such that

$$w = Tv \tag{10.6}$$

Since T is injective, if there is any number $u \in \mathbb{V}$ such that $Tu = Tv$ then $u = v$. Hence for any $w \in \mathbb{W}$ there is a unique element $v \in \mathbb{V}$ such that $w = Tv$. Define the map $S : \mathbb{W} \mapsto \mathbb{V}$ such that $v = Sw$ is that number. Hence

$$w = TSw \tag{10.7}$$

Thus TS is the identity map on $\mathbb{W}$ (since it maps $w \in \mathbb{W}$ to itself).

Next, suppose that $v \in \mathbb{V}$. Then

$$T(STv) = (TS)Tv = I_{\mathbb{W}}Tv = Tv \tag{10.8}$$

Since T is injective (one-to-one), the fact that $T(STv) = v$ implies that

$$STv = v \tag{10.9}$$

Since ST maps every element of v to itself, ST is the identity map on $\mathbb{V}$.

To show that S is linear, let $w_1, w_2 \in \mathbb{W}$. Then

$$T(Sw_1 + Sw_2) = TSw_1 + TSw_2 \text{ because } T \text{ is linear} \tag{10.10}$$

$$= w_1 + w_2 \text{ because } TS \text{ is the identity map} \tag{10.11}$$

For homogeneity, Let $a \in \mathbb{F}$ and $w \in \mathbb{W}$. Then

$$T(aSw) = aT(Sw) = aTSw = aw \tag{10.12}$$

$$ST(aSw) = Saw \tag{10.13}$$

$$aSw = Saw \tag{10.14}$$

i.e., $S(aw) = aS(w)$. Hence S is a linear map; it has the properties that $ST = I$ and $TS = I$ in their respective domains hence it is the inverse of T. Hence T is invertible. □

Definition 10.4 Let $\mathbb{V}$ and $\mathbb{W}$ be vector spaces over $\mathbb{F}$. Then $\mathbb{V}$ and $\mathbb{W}$ are said to be **isomorphic vector spaces** if there exists an invertible linear map $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ (i.e., there also exists a linear map $S = T^{-1} \in \mathcal{L}(\mathbb{W}, \mathbb{V})$).

Definition 10.5 An **operator** is a linear map from a vector space to itself. We denote the set of operators on $\mathbb{V}$ by $\mathcal{L}(\mathbb{V})$ (instead of $\mathcal{L}(\mathbb{V}, \mathbb{V})$).

Theorem 10.6 Let $\mathbb{V}$ and $\mathbb{W}$ be finite-dimensional vector spaces. Then they are isomorphic if and only if they have the same dimension.

Proof. ($\Longrightarrow$) Assume that $\mathbb{V}$ and $\mathbb{W}$ are isomorphic.

Then there exists an invertible linear map $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$.

Since T is invertible, it is injective (one-to-one), hence by theorem 8.17, $\text{null}(T) = \{0\}$. Hence

$$\dim \text{null}(T) = 0 \tag{10.15}$$

Because T is invertible, it is surjective (onto) and thus

$$\text{range}(T) = \mathbb{W} \tag{10.16}$$

By equation 8.30

$$\dim(\mathbb{V}) = \dim \text{null}(T) + \dim \text{range}(T) = \dim \text{range}(T) = \dim(\mathbb{W}) \tag{10.17}$$

($\Longleftarrow$) Assume that $\dim(\mathbb{V}) = \dim(\mathbb{W})$.

Let $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_n)$ be bases of $\mathbb{V}$ and $\mathbb{W}$.

Let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ be defined such that

$$T(a_1 v_1 + \cdots + a_n v_n) = a_1 w_1 + \cdots + a_n w_n \tag{10.18}$$

Since $(w_1, \ldots, w_n)$ is a basis of $\mathbb{W}$, T spans $\mathbb{W}$. Hence range$(T) = \mathbb{W}$, i.e., T is onto.

To show that T is one-to-one, suppose that $T\alpha = T\beta$ where

$$\left.\begin{aligned} \alpha &= a_1 v_1 + \cdots + a_n v_n \\ \beta &= b_1 v_1 + \cdots + a_n v_n \end{aligned}\right\} \tag{10.19}$$

Then

$$T(a_1 v_1 + \cdots + a_n v_n) = T(b_1 v_1 + \cdots + b_n v_n) \tag{10.20}$$

i.e.,

$$a_1 w_1 + \cdots + a_n w_n = b_1 w_1 + \cdots + b_n w_n \tag{10.21}$$

Since $(w_1, \ldots, w_n)$ is linearly independent, $a_i = b_i, i = 1, \ldots, n$, hence $\alpha = \beta$. Hence T is one-to-one.

Since T is one-to-one and onto, it is invertible, and therefore $\mathbb{V}$ and $\mathbb{W}$ are isomorphic.

$\square$

Corollary 10.7 Let $\mathbb{V}$ be a finite dimensional vector space of dimension n. Then $\mathbb{V}$ is isomorphic to $\mathbb{F}^n$.

Theorem 10.8 Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector spaces with bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$, and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then $\mathcal{M}$ is an invertible linear map

$$\mathcal{M}(T) : \mathcal{L}(\mathbb{V}, \mathbb{W}) \mapsto \text{Mat}(m, n, \mathbb{F}) \tag{10.22}$$

i.e., $\mathcal{M}(T) \in \mathcal{L}(\mathcal{L}(\mathbb{V}, \mathbb{W}), \text{Mat}(m, n, \mathbb{F}))$ is invertible.

Proof. We have already shown that $\mathcal{M}(T)$ is linear (theorem 9.2). To show invertibility, we need to show that $\mathcal{M}(T)$ is one-to-one and onto.

Suppose that $\mathcal{M}(T) = 0$. Then $Tv_k = 0$ for all $k = 1, \ldots, n$. Because $(v_1, \ldots, v_n)$ is a basis, every $v \in \mathbb{V}$ is a linear combination. Therefore

$$T(a_1 v_1 + \cdots + a_n v_n) = 0 \tag{10.23}$$

which must hold for all values of $a_i \in \mathbb{F}$. Thus $T = 0$. Thus $\text{null}(T) = \{0\}$. By 8.17 $\mathcal{M}$ is one-to-one (injective).

Now suppose that $A \in \text{Mat}(m, n, \mathbb{F})$ is any matrix, given by

$$A = \begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & & \vdots \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \tag{10.24}$$

If we define $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ by

$$Tv_k = \sum_{j=1}^{m} a_{j,k} w_j \tag{10.25}$$

then $\mathcal{M}(T) = A$. Hence $\text{range}(\mathcal{M}(T)) = \text{Mat}(m, n\mathbb{F})$, meaning $\mathcal{M}$ is onto. □

Theorem 10.9 $\dim(\text{Mat}(n, m, \mathbb{F})) = mn$.

Proof. Use as a basis the set of all matrices with a 1 in one entry and zeros everywhere else. □

Corollary 10.10 Let $\mathbb{V}$ and $\mathbb{W}$ be finite dimensional vector spaces over $\mathbb{F}$. Then

$$\dim(\mathcal{L}(\mathbb{V}, \mathbb{W})) = (\dim \mathbb{V})\,(\dim \mathbb{W}) \tag{10.26}$$

Proof. $\mathcal{L}(\mathbb{V}, \mathbb{W})$ is isomorphic to $\text{Mat}(m, n, \mathbb{F})$, which has dimension mn, where $m = \dim \mathbb{V}$ and $n = \dim \mathbb{W}$. □

Theorem 10.11 Let $\mathbb{V}$ be a finite dimensional vector space and let $T \in \mathcal{L}(\mathbb{V})$. Then the following are equivalent:(a) T is invertible; (b) T is one-to-one; (c) T is onto.

Proof. $((a) \implies (b))$ This follows because T invertible $\implies$ T is both onto and one-to-one.

$((b) \implies (c)))$ Assume that T is one-to-one. Then $\text{null}(T) = \{0\}$. Hence

$$\dim \text{range}(T) = \dim(\mathbb{V}) - \dim(\text{null}(T)) = \dim(\mathbb{V}) \tag{10.27}$$

i.e., $\text{range}(T) = \mathbb{V}$ (this follows from exercise 2.11). Hence T is onto, and (b) is true.

$((c) \implies (a))$. Assume that T is onto.

Then $\text{range}(T) = \mathbb{V}$, and so $\dim \text{range}(T) = \dim(\mathbb{V})$. Hence

$$\dim \text{null}(T) = \dim \mathbb{V} - \dim \text{range}(T) = 0 \tag{10.28}$$

Therefore T is one-to-one. Since T is one-to-one it is invertible, so (a) is true.

□

Topic 11

Operators and Eigenvalues

Definition 11.1 An **operator** is a linear map from a vector space to itself. We denote the set of operators on $\mathbb{V}$ by $\mathcal{L}(\mathbb{V})$ (instead of $\mathcal{L}(\mathbb{V}, \mathbb{V})$).

If $T \in \mathcal{L}(\mathbb{V})$ then $T^n \in \mathcal{L}(\mathbb{V})$ for any positive integer n. We use the notation T^2 to denote the product TT, $T^3 = TTT$, etc.

If T is invertible, then we define $T^{-m} = (T^{-1})^m$. Furthermore,

$$T^m T^n = T^{m+n}, \qquad (T^m)^n = T^{mn} \tag{11.1}$$

where m, n are any integers.

If T is not invertible then equation 11.1 still holds for integers $n, m \geq 0$.

Definition 11.2 If $p(x) = a_0 + a_1 x + \cdots + a_n x^n$ is a polynomial and T is an operator then we define the **polynomial operator** as

$$p(T) = a_0 I + a_1 T + \cdots + a_n T^n \tag{11.2}$$

If p and q are polynomials then we define

$$(pq)(T) = p(T)q(T) \tag{11.3}$$

Definition 11.3 Let $T \in \mathcal{L}(\mathbb{V})$ be an operator on a finite dimensional non-zero vector space $\mathbb{V}$ over $\mathbb{F}$, and let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Then the **restriction of T to $\mathbb{U}$**, denoted by $T|_{\mathbb{U}}$ is T operating on $\mathbb{U}$.

Definition 11.4 Let $\mathbb{V}$ be a finite dimensional non-zero vector space over $\mathbb{V}$, let $\mathbb{U}$ be a subspce of $\mathbb{V}$, and let $T \in \mathcal{L}(\mathbb{V})$ be an operator on $\mathbb{V}$. Then U **is invariant under T** if

$$\text{range}(T|_{\mathbb{U}}) = \mathbb{U} \tag{11.4}$$

i.e, if for ever $u \in \mathbb{U}$ then $Tu \in \mathbb{U}$. In this case $T|_{\mathbb{U}}$ is an operator on $\mathbb{U}$.

Definition 11.5 Let $\mathbb{V}$ be a vector space and $T \in \mathcal{L}(\mathbb{V})$. If there exists a vector $v \in \mathbb{V}, v \neq 0$, and a scalar $\lambda \in \mathbb{F}$ such that

$$Tv = \lambda v \tag{11.5}$$

Then λ is called an **eigenvalue** of T with **eigenvector** v.

From equation 11.5 we see that (λ, v) are an eigenvalue-eigenvector pair if and only if

$$(T - \lambda I)v = 0 \tag{11.6}$$

Remark 11.6 The set of all eigenvectors of T is null$(T - \lambda I)$.

Remark 11.7 Let T be an operator on $\mathbb{V}$, and let λ be an eigenvalue of T. Then the set of all eigenvectors of T with eigenvalue λ is a subspace of $\mathbb{V}$.

Theorem 11.8 The following are equivalent:

(1) λ is an eigenvalue of T

(2) $T - \lambda I$ is not injective.

(3) $T - \lambda I$ is not invertible.

(4) $T - \lambda I$ is not surjective.

Proof. ((1) $\iff$ (2)) λ is an eigenvalue of T iff there is some $u \neq 0$ such that $(T - \lambda I)u = 0$. Hence $u \in \text{null}(T)$. Hence $\text{null}(T) \neq \{0\}$. Hence T is not injective by theorem 8.17.

((2) $\iff$ (3)) An operator is injective iff it is invertible. Since T is not injective, it is not invertible. This follows from theorem 10.11.

((2) $\iff$ (4)) An operator is invertible iff it is surjective, which also follows from theorem 10.11. □

Theorem 11.9 Let $\mathbb{V}$ be a finite-dimensional vector space over $\mathbb{F}$ and $T \in \mathcal{L}(\mathbb{V})$. Let $\lambda_1 \neq \lambda_2 \neq \cdots \neq \lambda_m$ be eigenvalues of T with corresponding non-zero eigenvectors $v_1, \ldots, v_m$. Then $(v_1, \ldots, v_m)$ is linearly independent.

Proof. Suppose $(v_1, \ldots, v_m)$ is linearly dependent. Since the eigenvectors are all nonzero then $v_1 \neq 0$. Then by theorem 6.6, there is some k such that

$$v_k \in \text{span}(v_1, \ldots, v_{k-1}) \tag{11.7}$$

and if v_k is removed from $(v_1, \ldots, v_m)$ then the span of the remaining list equals the span of the original list. Let k designate the smallest integer such that this is true.

Since k is the smallest integer for which this is true, the list $(v_1, \ldots, v_{k-1})$ is linearly independent.

Hence there exists constants $a_1, \ldots, a_k \in \mathbb{F}$ such that

$$v_k = a_1 v_1 + \cdots + a_{k-1} v_{k-1} \tag{11.8}$$

$$Tv_k = Ta_1 v_1 + \cdots + Ta_{k-1} v_{k-1} \tag{11.9}$$

Since $Tv_j = \lambda_j v_j$,

$$\lambda_k v_k = a_1 \lambda_1 v_1 + \cdots + a_{k-1} \lambda_{k-1} v_{k-1} \tag{11.10}$$

From equation 11.8,

$$\lambda_k v_k = a_1 \lambda_k v_1 + \cdots + a_{k-1} \lambda_k v_{k-1} \tag{11.11}$$

Subtracting equation 11.10 from 11.11,

$$0 = a_1(\lambda_k - \lambda_1) v_1 + \cdots + a_{k-1}(\lambda_k - \lambda_{k-1}) v_{k-1} \tag{11.12}$$

Since $(v_1, \ldots, v_k)$ is linearly independent, and since the λ_j are all distinct,

$$a_1 = a_2 = \cdots = a_{k-1} = 0 \tag{11.13}$$

From equation 11.8, $v_k = 0$. This contradicts the assumption that all the $v_k \neq 0$.

Therefore $(v_1, \ldots, v_m)$ must be linearly independent. □

Corollary 11.10 Let $T \in \mathcal{L}(\mathbb{V})$ be an operator on $\mathbb{V}$. Then T has at most $\dim(\mathbb{V})$ distinct eigenvalues.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the distinct eigenvalues with corresponding non-zero eigenvectors $(v_1, \ldots, v_m)$.

By theorem 11.9, the list $(v_1, \ldots, v_m)$ are linearly independent.

By theorem 6.7, the length of every linearly independent list is at most the length of every spanning list of $\mathbb{V}$. By definition of dimension, the length of any basis of $\mathbb{V}$ is $\dim \mathbb{V}$, Hence

$$m = \text{length}(v_1, \ldots, v_m) \leq \dim \mathbb{V} \tag{11.14}$$

□

Theorem 11.11 Let $\mathbb{V}$ be a finite-dimensional non-zero complex vector space, and let $T \in \mathcal{L}(V)$. Then T has an eigenvalue.

Proof. Let $n = \dim(V) > 0$, an pick any nonzero $v \in \mathbb{V}$, and define

$$\ell = (v, Tv, T^2v, \ldots, T^nv) \tag{11.15}$$

Since

$$\text{length}(\ell) = n + 1 > n = \dim(V) \tag{11.16}$$

the list ℓ is linearly-dependent. Hence there exists $a_0, \ldots, a_n$, not all zero, such that

$$0 = a_0v + a_1Tv + a_2T^2v + \cdots + a_nT^nv \tag{11.17}$$

Define $m \leq n$ as the largest integer such that $a_m \neq 0$. Then

$$0 = a_0v + a_1Tv + a_2T^2v + \cdots + a_mT^mv \tag{11.18}$$

Define the polynomial

$$p(z) = a_0 + a_1z + \cdots + a_mz^m = c(z - \lambda_1)(z - \lambda_2) \cdots (z - \lambda_m) \tag{11.19}$$

for some roots $\lambda_1, \lambda_2, \ldots$. Hence

$$0 = p(T)v = c(T - \lambda_1 I) \cdots (T - \lambda_m I)v \tag{11.20}$$

which holds for all nonzero vectors $v \in \mathbb{V}$. Hence for some j

$$(T - \lambda_j I)w = 0 \tag{11.21}$$

where

$$w = (T - \lambda_{j+1} I) \cdots (T - \lambda_m I)v \neq 0 \tag{11.22}$$

Thus

$$\text{null}(T - \lambda_j I) \neq \{0\} \tag{11.23}$$

Hence $T - \lambda_j I$ is not injective. By theorem 11.8, λ is an eigenvalue of T.

Hence an eigenvalue exists. □

Topic 12

Matrices of Operators

Definition 12.1 Matrix of an Operator. Let $T \in \mathcal{L}(\mathbb{V})$ and let $(v_1, \dots, v_n)$ be a basis of $\mathbb{V}$. Then there are some numbers $a_{i,j}$ such that

$$\left.\begin{aligned} Tv_1 &= a_{11}v_1 + \cdots + a_{n1}v_n \\ Tv_2 &= a_{12}v_1 + \cdots + a_{n2}v_n \\ &\vdots \\ Tv_n &= a_{1n}v_1 + \cdots + a_{nn}v_n \end{aligned}\right\} \tag{12.1}$$

Then we define the **matrix of T with respect to the basis** $(v_1, \dots, v_n)$ as

$$\mathcal{M}(T) = \mathcal{M}(T, (v_1, \dots, v_n)) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \tag{12.2}$$

The j^{th} column of $\mathcal{M}$ represents Tv_j. One way to remember this is by the matrix product

$$(Tv_1, \dots, Tv_n) = (v_1, \dots, v_n)\mathcal{M}(T) \tag{12.3}$$

or

$$\mathcal{M}T = A^{-1}TA = (v_1, \dots, v_n)^{-1}T(v_1, \dots, v_n) \tag{12.4}$$

where $A = (v_1, \dots, v_n)$.

Theorem 12.2 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ with basis $(v_1, \dots, v_n)$ and $T \in \mathcal{L}(\mathbb{V})$ an operator on $\mathbb{V}$. The the following are equivalent:

(1) $\mathcal{M}(T, (v_1, \dots, v_n)$ is upper triangular.

(2) $Tv_k \in \text{span}(v_1, \dots, v_k)$ for $k = 1, \dots, n$.

(3) $\text{span}(v_1, \dots, v_k)$ is invariant under T for each $k = 1, \dots, n$.

Proof. $\underline{((1) \implies (2))}$ Let $\mathcal{M}(T)$ be upper triangular,

$$\mathcal{M}(T) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} \\ 0 & a_{22} & \cdots & & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & a_{nn} \end{pmatrix} \tag{12.5}$$

then from equation 12.1 i ndefinintion 12.1

$$\left.\begin{aligned}
Tv_1 &= a_{11}v_1 \\
Tv_2 &= a_{12}v_1 + a_{22}v_2 \\
Tv_3 &= a_{13}v_1 + a_{23}v_2 + a_{33}v_3 \\
&\vdots \\
Tv_n &= a_{1n}v_1 + \cdots + a_{nn}v_n
\end{aligned}\right\} \tag{12.6}$$

Thus each Tv_i is a linear combination of the $(v_1, \dots, v_i)$, i.e.,

$$Tv_i \in \text{span}(v_1, \dots, v_i) \tag{12.7}$$

which is statement (2). Hence $((1) \implies (2))$.

$\underline{((2) \implies (1))}$ Assume each $Tv_i \in \text{span}(v_1, \dots, v_i)$.

Then there exist a_{ij} such that equation 12.6 holds.

Then $\mathcal{M}(T)$ is defined as in equation 12.5.

Hence $\mathcal{M}(T)$ is upper triangular, and $((2) \implies (1))$.

$\underline{((3) \implies (2))}$ Assume $\text{span}(v_1, \dots, v_k)$ is invariant under T.

Then

$$\left.\begin{aligned}
&Tv_1 \in \text{span}(v_1) \\
&Tv_2 \in \text{span}(v_1, v_2) \\
&\vdots \\
&Tv_k \in \text{span}(v_1, \dots, v_k) \\
&\vdots
\end{aligned}\right\} \tag{12.8}$$

which is precisely statement (2).

((2) $\implies$ (3)). Fix k and assume that $Tv_k \in \text{span}(v_1, \dots, v_k)$. Hence

$$\left.\begin{aligned} &Tv_1 \in \text{span}(v_1) \subset \text{span}(v_1, \dots, v_k) \\ &Tv_2 \in \text{span}(v_1, v_2) \subset \text{span}(v_1, \dots, v_k) \\ &\vdots \\ &Tv_k \in \text{span}(v_1, \dots, v_k) \subset \text{span}(v_1, \dots, v_k) \end{aligned}\right\} \tag{12.9}$$

Let $v \in \text{span}(v_1, \dots, v_k)$. Then v is a linear linear combination of $(v_1, \dots, v_k)$. Then by equation 12.9,

$$Tv = a_1v_1 + a_2v_2 + \cdots + a_nv_n \in \text{span}(v_1, \dots, v_k) \tag{12.10}$$

Hence $\text{span}(v_1, \dots, v_k)$ is invariant under T. Hence ((2) $\implies$ (3)). □

Theorem 12.3 Let $\mathbb{V}$ be a complex vector space and $T \in \mathcal{L}(\mathbb{V})$. Then there exists a basis under which $\mathcal{M}(T)$ is upper triangular.

Proof. Prove by induction. Let $n = \dim \mathbb{V}$.

Since any 1×1 matrix is upper-triangular, the result holds for $n = 1$.

Assume $n > 1$ and that the result holds for $n - 1$.

Let λ be any eigenvalue of T (see theorem 11.11). Define

$$\mathbb{U} = \text{range}(T - \lambda I) \tag{12.11}$$

Since λ is an eigenvalue of T, $T - \lambda I$ is not surjective (theorem 11.8)

$$\dim \mathbb{U} < \dim \mathbb{V} \tag{12.12}$$

Let $u \in \mathbb{U}$. Then

$$Tu = Tu - \lambda Iu + \lambda u = (T - \lambda I)u + \lambda u \tag{12.13}$$

Since $u \in \mathbb{U}$ and $(T - \lambda I)u \in \mathbb{U}$ (by definition of $\mathbb{U}$), then $Tu \in \mathbb{U}$. Hence $\mathbb{U}$ is invariant under T.

Since U is invariant under T, $T|_{\mathbb{U}}$ is an operator on $\mathbb{U}$.

Hence $T|_{\mathbb{U}}$ has an upper triangular basis $(u_1, \dots, u_m)$ (by the inductive hypothesis and equation 12.12).

By theorem 12.2

$$Tu_j = (T|_{\mathbb{U}})u_j \in \text{span}(u_1, \dots, u_j) \tag{12.14}$$

for each j.

We can extend $(u_1, \dots, u_m)$ to a basis

$$B = (u_1, \dots, u_m, v_1, \dots, v_{n-m}) \tag{12.15}$$

of $\mathbb{V}$. For each k,

$$Tv_k = Tv_k - \lambda v_k + \lambda v_k = (T - \lambda I)v_k + \lambda v_k \tag{12.16}$$

Since the first term is in $\mathbb{U}$ (definition of $\mathbb{U}$) and the second term is in $\text{span}(v_1, \dots, v_{n-m})$, then

$$Tv_k \in \text{span}(B) \tag{12.17}$$

Hence theorem 12.2 applies again and T has an upper triangular matrix with respect to the basis B. $\square$

Theorem 12.4 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let $T \in \mathcal{L}(V)$ be such that $\mathcal{M}(T)$ is upper triangular with resepct to some basis $B = (v_1, \dots, v_n)$ of $\mathbb{V}$. Then T is invertible if and only if all the entries on the diagonal of $\mathcal{M}(T)$ are nonzero.

Proof. ($\Longrightarrow$) Let $\mathcal{M}(T)$ be upper triangular, and write

$$\mathcal{M}(T, B) = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \lambda_1 & * & * \\ 0 & 0 & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} \tag{12.18}$$

where "$*$" means "anything."

If $\lambda_1 = 0$ then $Tv_1 = 0$ and hence T is not invertible. (because $\text{null}(T) \neq \{0\}$, hence T is not injective, hence it is not invertible).

Supose $\lambda_k = 0$ for some k. Then for $i = 1, \dots, k-1$

$$Tv_i \in \text{span}(v_1, \dots, v_{k-1}) \tag{12.19}$$

because the matrix is upper triangular; and because $\lambda_k = a_{kk} = 0$,

$$\begin{aligned} Tv_k &= a_{1k}v_1 + a_{2k}v_2 + \cdots + a_{k-1,k}v_{k-1} + a_{kk}v_k && (12.20) \\ &= a_{1k}v_1 + a_{2k}v_2 + \cdots + a_{k-1,k}v_{k-1} && (12.21) \\ &\in \text{span}(v_1, \dots, v_{k-1}) && (12.22) \end{aligned}$$

Define $S : \text{span}(v_1, \dots, v_k) \mapsto \text{span}(v_1, \dots, v_{k-1})$ by

$$Sv = Tv = T|_{(v_1, \dots, v_k)}v \tag{12.23}$$

But

$$\dim(v_1, \dots, v_{k-1}) = k - 1 \tag{12.24}$$

$$\dim(v_1, \dots, v_k) = k \tag{12.25}$$

hence S cannot be injective (Corollary 8.20).

Hence there exists a vector v such that $Sv = 0$ (Theorem 8.17). Hence $Tv = 0$ and therefore T is not injective (also by Theorem 8.17).

Since T is not injective, it is not invertible (theorem 10.11).

Hence if any of the $\lambda_k = 0$ then T is not invertible. Hence invertibility requires that all the diagonal elements be nonzero.

($\Longleftarrow$) Suppose T is not invertible.

Then T is not injective (see theorem 10.11).

Thus there exists a $v \neq 0$ such that $Tv = 0$ (Theorem 8.17). Since B is a basis of $\mathbb{V}$, there exists $a_1, \dots, a_k$ with $a_k \neq 0$, $k \leq n$, such that

$$v = a_1 v_1 + \cdots + a_k v_k \tag{12.26}$$

Choose k to be the largest k such that equation 12.26 holds (with $a_k \neq 0$). Then

$$0 = Tv = a_1 T v_1 + \cdots + a_{k-1} T v_{k-1} + a_k T v_k \tag{12.27}$$

$$a_k T v_k = -a_1 T v_1 - \cdots - a_{k-1} T v_{k-1} \tag{12.28}$$

$$T v_k = -\frac{a_1}{a_k} T v_1 - \cdots - \frac{a_{k-1}}{a_k} T v_{k-1} \tag{12.29}$$

But because T is upper triangular,

$$\left.\begin{aligned} Tv_1 &= b_{1,1}v_1 \\ Tv_1 &= b_{1,2}v_1 + b_{2,2}v_2 \\ &\vdots \\ Tv_{k-1} &= b_{1,k-1}v_1 + b_{2,k-1}v_2 + \cdots + b_{k-1,k-1}v_{k-1} \end{aligned}\right\} \tag{12.30}$$

where the b_{ij} are the elements of $\mathcal{M}(T, B)$. Hence

$$\begin{aligned} Tv_k = -\frac{a_1}{a_k}b_{1,1}v_1 - \frac{a_2}{a_k}(b_{1,2}v_1 + b_{2,2}v_2) - \cdots - \frac{a_{k-1}}{a_k}(b_{1,k-1}v_1 \\ + b_{2,k-1}v_2 + \cdots + b_{k-1,k-1}v_{k-1}) \end{aligned} \tag{12.31}$$

and consequently $Tv_k \in \text{span}(v_1, \ldots, v_{k-1})$ (becasue v_k does not appear in the above expansion), i.e., for some numbers $c_1, \ldots, c_{k-1}$,

$$Tv_k = c_1v_1 + \cdots + c_{k-1}v_{k-1} \tag{12.32}$$

But because T is upper triangular,

$$Tv_k = b_{1,k}v_1 + b_{2,k}v_2 + \cdots + b_{k-1,k}v_{k-1} + b_{k,k}v_k$$

Comparing the last two expressions gives $c_i = b_{i,k}$, $i = 1, \ldots, k-1$, and $0 = b_{k,k} = \lambda_k$ (the last equality follows because $b_{k,k}$ is the diagonal entry. □

Theorem 12.5 Let $\mathbb{V}$ be a vector space and $T \in \mathcal{L}(\mathbb{V})$. Then $\mathcal{M}(T, B)$ is diagonal if and only if $\mathbb{V}$ has a basis consisting solely of eigenvectors of T.

Proof. Let $B = (v_1, \ldots, v_n)$ be a basis of $\mathbb{V}$.

But $\mathcal{M}(T, B)$ is diagonal if and only if (see equation 12.1),

$$\left.\begin{aligned} Tv_1 &= \lambda_1 v_1 \\ &\vdots \\ Tv_n &= \lambda_n v_n \end{aligned}\right\} \tag{12.33}$$

which is true if and only if $(v_1, \dots, v_n)$ are eigenvectors of T. □

Theorem 12.6 Let $\mathbb{V}$ be a vector space and $T \in \mathcal{L}(\mathbb{V})$ an operator with $\dim \mathbb{V}$ distinct eigenvalues. Then $\mathbb{V}$ has a basis B such that $\mathcal{M}(T, B)$ is diagonal.

Proof. Suppose that T has $n = \dim \mathbb{V}$ distinct eigenvalue $\lambda_1, \dots, \lambda_n$ with corresponding non-zero eigenvectors $v_1, \dots, v_n$.

By theorem 11.9, since the λ_i are distinct then $(v_1, \dots, v_n)$ is linearly independent.

Consequently, since

$$\text{length}(v_1, \dots, v_n) = n = \dim(\mathbb{V}) \tag{12.34}$$

then $B = (v_1, \dots, v_n)$ is a basis, because every list of linearly independent vectors with the same length as the dimension of the vector space is a basis (theorem 7.11).

Since B consists solely of eigenvectors of T, then $\mathcal{M}(T, B)$ is diagonal by theorem 12.5. □

Remark 12.7 The converse of theorem 12.6 is not true. It is possible to find operators that have diagonal matrices even though they do not have n distinct eigenvalues.

Theorem 12.8 Let $\mathbb{V}$ be a vectors space and $T \in \mathcal{L}(\mathbb{V})$; let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then the following are equivalent:

(1) $M(T, B)$ is diagonal with respect to some basis B of $\mathbb{V}$.

(2) $\mathbb{V}$ has a basis consisting only of eigenvectors of T.

(3) There exist one-dimensional subspaces $\mathbb{U}_1, \dots, \mathbb{U}_n$ of $\mathbb{V}$ that are each invariant under T such that

$$\mathbb{V} = \mathbb{U}_1 \oplus \mathbb{U}_2 \oplus \cdots \oplus \mathbb{U}_n \tag{12.35}$$

(4) $\mathbb{V} = \text{null}(T - \lambda_1 I) \oplus \cdots \oplus \text{null}(T - \lambda_m I)$

(5) $\dim \mathbb{V} = \dim \text{null}(T - \lambda_1 I) + \cdots + \dim \text{null}(T - \lambda_m I)$

Proof. ((1) $\iff$ (2)) This is theorem 12.5.

((2) $\implies$ (3)) Assume (2), that $\mathbb{V}$ has a basis $B = (v_1, \dots, v_n)$ consisting of eigenvectors of T.

Let

$$\mathbb{U}_j = \text{span}(v_j), \quad j = 1, 2, \dots, n \tag{12.36}$$

By definition each $\mathbb{U}_j$ is one-dimensional. Furthermore, since v_j is an eigenvector then

$$Tv_j = \lambda_j v_j \in \mathbb{U}_j \tag{12.37}$$

so each $\mathbb{U}_j$ is invariant under T.

Since B is a basis, each vector $v \in \mathbb{V}$ can be written as a linear combination

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \tag{12.38}$$

and since each $u_i = a_i v_i \in \mathbb{U}_j$

$$v = u_1 + \cdots + u_n \tag{12.39}$$

where $u_i \in \mathbb{U}_i$. These u_i are unique, hence

$$\mathbb{V} = \mathbb{U}_1 \oplus \cdots \oplus \mathbb{U}_n \tag{12.40}$$

hence ((2) $\implies$ (3).

((3) $\implies$ (2)) Next, assume that (3) holds. Then there are one-dimensional subspaces $\mathbb{U}_1, \dots, \mathbb{U}_n$, each invariant under T, such that equation 12.40 holds.

Let $v_j \in \mathbb{U}_j$ such that $v_j \neq 0$. Since U_j is invariant under T, $Tv_j \in \mathbb{U}$. Since $\mathbb{U}$ is one-dimensional, (v_j) is a basis, hence there exists some number λ_j such that $Tv_j = \lambda_j v_j$. Hence v_j is an eigenvector of T.

By definition of the direct sum each vector $v \in \mathbb{V}$ can be written as a sum as in 12.39, where, as we have just shown, each u_j is is a scalar multiple of the v_j we chose above. Hence $(v_1, \dots, v_n)$ is a basis of eigenvectors. Hence ((3) $\Longrightarrow$ (2)).

((2) $\implies$ (4)) Assume (2). Then $\mathbb{V}$ has a basis consisting of eigenvectors $(v_1, \dots, v_n)$ of T. Let $(\lambda_1, \dots, \lambda_n)$ be the eigenvalues.

Let $v \in \mathbb{V}$. Then there are constants a_i such that

$$v = a_1 v_1 + \cdots + a_n v_n \tag{12.41}$$

Since v_i is an eigenvector,

$$\text{span}(v_i) = \text{null}(T - \lambda_i I) \tag{12.42}$$

Combining equations 12.41 and 12.42

$$\mathbb{V} = \text{null}(T - \lambda_1 I) + \cdots + \text{null}(T - \lambda_n I) \tag{12.43}$$

Suppose that there exist $u_i \in \text{null}(T - \lambda_i I)$ such that

$$0 = u_1 + \cdots + u_n \tag{12.44}$$

Since each u_i is an eigenvector, the u_i are linearly independent. Thus each term in equation 12.44 is zero:

$$u_i = 0 \tag{12.45}$$

By theorem 4.7,

$$\mathbb{V} = \text{null}(T - \lambda_1 I) \oplus \cdots \oplus \text{null}(T - \lambda_n I) \tag{12.46}$$

and (3) $\implies$ (4).

<u>((4) $\implies$ (5))</u> follows from the text exercise 2.17 and hence is left as an exercise.

<u>((5) $\implies$ (2))</u> Assume that (5) is true. Then

$$\dim \mathbb{V} = \dim \text{null}(T - \lambda_1 I) + \cdots + \dim \text{null}(T - \lambda_m I) \tag{12.47}$$

Define $\mathbb{U}_i = \text{null}(T - \lambda_i I)$. Define a basis of each $\mathbb{U}_j$ and put all these bases together to form a list $B = (v_1, \dots, v_n)$, where $n = \dim \mathbb{V}$.

Each v_i is an eigenvector of T because $v_i \in \mathbb{U}_i = \text{null}(T - \lambda_i I)$.

Suppose that there exist $a_i \in \mathbb{F}$ such that

$$0 = a_1 v_1 + \cdots + a_n v_n \tag{12.48}$$

Define u_i as the sum of all the terms such that $v_k \in (T - \lambda_i I)$ (the number of eigenvalues might be smaller than the number of eigenvectors so there might be more than one linearly independent eigenvector is each set).

Hence each u_i is an eigenvalue of T with eigenvector λ_i. (To see this, suppose that

$$u_i = a_{i,1}v_{i,1} + \cdots + a_{i,k}v_{i,k} \tag{12.49}$$

Since each $v_{i,j}$ has eigenvalue λ_i,

$$Tu_i = a_{i,1}Tv_{i,1} + \cdots + a_{i,k}Tv_{i,k} = \lambda_i u_i \tag{12.50}$$

hence u_i is an eigenvector with eigenvalue λ_i.)

By equation 12.48 this means

$$u_1 + \cdots + u_m = 0 \tag{12.51}$$

Since eigenvectors of distinct eigenvalues are linearly independent, this is only true if each $u_i = 0$.

But each u_i is the sum as in 12.49 where the $v_{i,j}$ are a basis of $\mathbb{U}_i$, then the coefficients in equation 12.48 are also zero. This means that the v_i are linearly independent and hence form a basis.

Thus $\mathbb{V}$ has a basis consisting of eigenvectors, hence (5) $\Longrightarrow$ (2). $\square$

Topic 13

The Canonical Diagonal Form

Definition 13.1 A **linear transformation** on $\mathbb{R}^n$ is a linear map

$$T : \mathbb{R}^n \mapsto \mathbb{R}^n \tag{13.1}$$

given by

$$y = Tx \tag{13.2}$$

where $x, y \in \mathbb{R}^n$ and T is an $n \times n$ matrix.

Definition 13.2 Two square matrices A and B, both of dimension n, are said to be **similar** if there exists an $n \times n$ invertible matrix U such that

$$U^{-1}AU = B \tag{13.3}$$

Similar matrices represent the same linear transformation in different coordinate systems.

Let $E = (e_1, e_2, \dots, e_n)$ be a basis of $\mathbb{R}^n$. Then we can write

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n \tag{13.4}$$

$$y = \eta_1 e_1 + \eta_2 e_2 + \cdots + \eta_n e_n \tag{13.5}$$

for some $\xi_1, \dots, \xi_n$ and $\eta_1, \dots, \eta_n \in \mathbb{R}$. The numbers ξ_i, η_i are called the **coordinates** of the vectors x and y with respect to the basis E.

If we let E be the matrix whose columns are the vectors $e_1, \dots, e_n$, then

$$x = E\xi, \qquad y = E\eta \tag{13.6}$$

where ξ and η are column vectors, i.e.,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = E \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}, \qquad \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = E \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix} \tag{13.7}$$

If $y = Tx$ then

$$E\eta = TE\xi \implies \eta = (E^{-1}TE)\xi \tag{13.8}$$

In other words if $y = Tx$ in a coordinate system defined by basis $(b_1, \dots, b_n)$ then then $\eta = T'\xi$ in the coordinate system defined by basis $(e_1, \dots, e_n)$, where $T' = E^{-1}TE$.

Example 13.1 Let A represent the counter-clockwise rotation through 90 degrees:

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{13.9}$$

For example if $x = (0.5, 0.7)^T$ then

$$y = Ax = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix} = \begin{pmatrix} -.7 \\ 5 \end{pmatrix} \tag{13.10}$$

Consider the basis

$$e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{13.11}$$

The transformation matrix is

$$T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{13.12}$$

and its inverse is

$$T^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \tag{13.13}$$

so that the transformation of the matrix of the linear transformation given by $y = Ax$ is

$$B = T^{-1}AT = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{13.14}$$

$$= \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix} \tag{13.15}$$

$$= \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix} \tag{13.16}$$

Now consider the vector

$$\xi = T^{-1}x = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix} \tag{13.17}$$

Then ξ represents the same vector as x but in the coordinate system (e_1, e_2). In this coordinate system, the transform $y = Ax$ is represented by

$$\eta = B\xi = \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix} = \begin{pmatrix} -0.7 \\ 1.2 \end{pmatrix} \tag{13.18}$$

Then

$$T\eta = T = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} -0.7 \\ 1.2 \end{pmatrix} = \begin{pmatrix} -.7 \\ 0.5 \end{pmatrix} = y \tag{13.19}$$

In terms of the original basis,

$$\eta = -0.7e_1 + 1.2e_2 = -0.7 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1.2 * \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -0.7 \\ 0.5 \end{pmatrix} \tag{13.20}$$

$$\xi = 0.5e_1 + 0.2e_2 = 0.5 * \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 0.2 * \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix} \tag{13.21}$$

So we have in the standard basis $((1,0),(0,1))$:

$$A : \begin{pmatrix} 0.5 \\ 0.7 \end{pmatrix} \mapsto \begin{pmatrix} -.7 \\ 0.5 \end{pmatrix} \tag{13.22}$$

and in the basis (e_1, e_2)

$$B : \begin{pmatrix} 0.5 \\ 0.2 \end{pmatrix} \mapsto \begin{pmatrix} -0.7 \\ 1.2 \end{pmatrix} \tag{13.23}$$

representing the same transform in different coordinate systems.

Theorem 13.3 Similar matrices have the same eigenvalues with the same multiplicities.

Proof. Let $B = T^{-1}AT$. Then, using the property that $\det AB = (\det A)(\det B)$ gives

$$\det(B - \lambda I) = \det(T^{-1}AT - \lambda I) \tag{13.24}$$

$$= \det(T^{-1}(A - T\lambda I T^{-1})T) \tag{13.25}$$

$$= \det(T^{-1}(A - \lambda I)T) \tag{13.26}$$

$$= (\det T^{-1})(\det(A - \lambda I))(\det T) \tag{13.27}$$

$$= \det(A - \lambda I) \tag{13.28}$$

In the last step we used the fact that

$$1 = \det I = \det(TT^{-1}) = (\det T)(\det T^{-1}) \tag{13.29}$$

Since A and B have the same characteristic equation they have the same eigenvalues with the same multiplicities. □

Example 13.2 From the previous example, we had $B = T^{-1}AT$ where

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & -1 \\ 2 & 1 \end{pmatrix} \tag{13.30}$$

Each of these matrices have the same eigenvalues $\{i, -i\}$. To see this observe that

$$\det(A - \lambda I) = \lambda^2 + 1 \tag{13.31}$$

while

$$\det(B - \lambda I) = (-1 - \lambda)(1 - \lambda) + 2 = -1 - \lambda + \lambda + \lambda^2 + 2 = \lambda^2 + 1 \tag{13.32}$$

Definition 13.4 A diagonal matrix Λ is called the **Diagonal Canonical Form** of a square matarix A if it is similar to A.

Theorem 13.5 Let A be an $n \times n$ square matrix. Then A is similar to a diagonal matrix if and only if A has n linearly independent eigenvectors.

Proof. ($\Longrightarrow$) Suppose A is similar to a diagonal matrix. Then there exists some invertible matrix T with linearly independent column vectors $(e_1, \dots, e_n)$ such that

$$T^{-1}AT = \text{diagonal}(\lambda_1, \dots, \lambda_n) \tag{13.33}$$

Multiplying on the left by T,

$$AT = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix} \text{diagonal}(\lambda_1, \dots, \lambda_n) \tag{13.34}$$

$$\begin{pmatrix} Ae_1 & \cdots & Ae_n \end{pmatrix} = \begin{pmatrix} \lambda_1 e_1 & \cdots & \lambda_n e_n \end{pmatrix} \tag{13.35}$$

where we use the observation that the jth column vector of AT is Ae_j.Hence

$$Ae_j = \lambda e_j \tag{13.36}$$

Therefore the vectors e_j must be eigenvectors of A with eigenvalues λ_j. Hence A has n linearly independent eigenvectors.

$(\Longleftarrow)$ Suppose that A has n linearly independent eigenvectors $(e_1, \dots, e_n)$ with eigenvectors $\lambda_1, \dots, \lambda_n$. Then let T be the matrix whose columns are e_i. Then

$$AT = A\begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix} \tag{13.37}$$

$$= \begin{pmatrix} Ae_1 & \cdots & Ae_n \end{pmatrix} \tag{13.38}$$

$$= \begin{pmatrix} \lambda_1 e_1 & \cdots & \lambda_n e_n \end{pmatrix} \tag{13.39}$$

$$= \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix} \text{diagonal}(\lambda_1, \dots, \lambda_n) \tag{13.40}$$

$$= T\text{diagonal}(\lambda_1, \dots, \lambda_n) \tag{13.41}$$

$$T^{-1}AT = \text{diagonal}(\lambda_1, \dots, \lambda_n) \tag{13.42}$$

Hence A is similar to a diagonal matrix. □

Example 13.3 Find the diagonal canonical form of the matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \tag{13.43}$$

The characteristic equation is $0 = \lambda^2 + 1$ hence the eigenvalues are $\lambda = \pm i$.

The eigenvector of i is found by solving

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = i \begin{pmatrix} x \\ y \end{pmatrix} \tag{13.44}$$

This gives $y = -ix$ and $x = iy$. One parameter is arbitrary so we choose $x = 1$ to give $y = -i$. Hece $e_1 = (1, -i)^T$.

Similarly, the eigenvector of $-i$ is found by solving

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = -i \begin{pmatrix} x \\ y \end{pmatrix} \tag{13.45}$$

Hence $x = -iy$ and $y = ix$. Again choosing $x = 1$ give $y = i$; hence $e_2 = (1, i)^T$.

The transformation matrix is

$$T = \begin{pmatrix} e_1 & e_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \tag{13.46}$$

Its inverse is

$$T^{-1} = \frac{1}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \tag{13.47}$$

Hence

$$T^{-1}AT = \frac{1}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix} \tag{13.48}$$

$$= \frac{1}{2} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix} \tag{13.49}$$

$$= \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \tag{13.50}$$

which, as expected, is the diagonal matrix of eigenvalues of A.

Topic 14

Invariant Subspaces

In this section we will assume that $\mathbb{V}$ is a real, finite-dimensional, non-zero vector space.

Theorem 14.1 Let $\mathbb{V}$ be a finite dimensional non-zero real vector space. Then $\mathbb{V}$ has an invariant subspace of either dimension 1 or dimension 2.

Proof. Let $n = \dim(\mathbb{V}) > 0$, $T \in \mathcal{L}(\mathbb{V})$, and pick any $v \in \mathbb{V}$ such that $v \neq 0$. Then define the list

$$L = (v, Tv, \dots, Tv^n) \tag{14.1}$$

Since $\text{Length}(L) > n = \dim \mathbb{V}$, L cannot be linearly independent. Hence there exist real numbers (note that we are assuming that $\mathbb{V}$ is real, hence $\mathbb{F} = \mathbb{R}$) $a_0, \dots, a_n$ such that

$$0 = a_0 v + a_1 T v + a_2 T^2 v + \cdots + a_n T^n v \tag{14.2}$$

Define the polynomial $p(x)$ by

$$p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \tag{14.3}$$

Then $p(x)$ nas n complex roots, which may be grouped into m real roots $\lambda_1, \dots, \lambda_m$, and $k = (m-n)/2$ complex conjugate pairs of roots (see theorem 5.12) and can be factored

$$p(x) = c(x-\lambda_1)\cdots(x-\lambda_m)(x^2+\alpha_1 x+\beta_1)\cdots(x^2+\alpha_k x+\beta_k) \tag{14.4}$$

where c and all the $\alpha_j, \beta_j, \lambda_j$ are real. From equation 14.2

$$0 = a_0 I v + a_1 T v + a_2 T^2 v + \cdots + a_n T^n v \tag{14.5}$$

$$= (a_0 I + a_1 T + a_2 T^2 + \cdots + a_n T^n) v \tag{14.6}$$

$$= c(T-\lambda_1 I)\cdots(T-\lambda_m I)(T^2+\alpha_1 T+\beta_1 I)\cdots(T^2+\alpha_k T+\beta_k I)v \tag{14.7}$$

Since 0 is in the null space of the operator on the right (the entire product) it must be in the null space of at least one of the factors.

Hence at least one of the factors on the right is not injective.

Either $T-\lambda_j I$ is not injective for some j, or $T^2+\alpha_j T+\beta_j I$ is not not injective for some j.

If $T-\lambda_j I$ is not injective, then T has an eigenvalue λ_j. Then T has an invariant subspace because $Tv_j = \lambda_j v_j \in \text{span}(v_j)$. But this subspace has dimension 1, hence T has an invariant subspace of dimension 1.

If $T^2+\alpha_j T+\beta_j I$ is not injective, then there is a non-zero solution to

$$T^2 u + \alpha_j T u + \beta_j u = 0 \tag{14.8}$$

Let $\mathbb{U} = \text{span}(u, Tu)$. Then either $\dim \mathbb{U} = 1$ or $\dim \mathbb{U} = 2$. Let $v \in \mathbb{U}$. Then in general there are numbers a, b such that

$$v = au + bTu \tag{14.9}$$

because (u, Tu) is a basis of $\mathbb{U}$. Hence from equation 14.8,

$$Tv = Tau + bT^2u = Tau + b(-\alpha_j Tu - \beta_j u) \tag{14.10}$$

rearranging

$$Tv = (a - b\alpha_j)Tu - b\beta_j u \in \text{span}(u, Tu) \tag{14.11}$$

Thus $\mathbb{U}$ is invariant under T. Hence in this case T has an invariant subspace of either dimension 1 or dimension 2. □

Definition 14.2 (**Projection**.) Suppose that $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$. Then for any vector $v \in \mathbb{V}$ such that $v = u + w$, where $u \in \mathbb{U}$ and $v \in \mathbb{W}$, the **Projection of** v **onto** $\mathbb{U}$ is

$$P_{\mathbb{U},\mathbb{W}}v = u \tag{14.12}$$

and the **Projection of** v **onto** $\mathbb{W}$ is

$$P_{\mathbb{W},\mathbb{U}} = w \tag{14.13}$$

Remark 14.3 Properties of Projections. Let $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$. Then

(1) $v = P_{\mathbb{U},\mathbb{W}}v + P_{\mathbb{W},\mathbb{U}}v$

(2) $P_{\mathbb{U},\mathbb{W}}^2 = P_{\mathbb{U},\mathbb{W}}$

(3) $\text{range}(P_{\mathbb{U},\mathbb{W}}) = \mathbb{U}$

(4) $\text{null}(P_{\mathbb{U},\mathbb{W}}) = \mathbb{W}$

Theorem 14.4 Let $\mathbb{V}$ be an odd-dimensional real-vector space. Then $\mathbb{V}$ has at least one eigenvalue.

Proof. Prove by induction on n, the dimension of the vector space.

For $n = 1$. Let $v \in \mathbb{V}$ be nonzero. Since $\dim \mathbb{V} = 1$, the list (v) is a basis of $\mathbb{V}$. Since T is an operator, $Tv \in \mathbb{V}$. Then $Tv = cv \in \mathbb{V}$ for some $c \in \mathbb{R}$. Hence c is an eigenvalue.

As the inductive step, assume that $n = \dim V$ is odd and assume the result holds for all vector spaces with dimesions $1, 3, 5, \dots, n-2$.

Either T has an eigenvalue or it does not. If it does the theorem is proven.

Assume that T does not have an eighevalue. Then by theorem 14.1, T has an invariant subspace $\mathbb{U}$ of dimension 2.

Define $\mathbb{W}$ by $\mathbb{V} = \mathbb{U} \oplus \mathbb{W}$ ($\mathbb{W}$ exists by theorem 7.6).

Define the operator $S \in \mathcal{L}(\mathbb{W})$ by

$$Sw = P_{\mathbb{W},\mathbb{U}}Tw \tag{14.14}$$

This definition makes sense because $Tw \in \mathbb{V}$ and $P_{\mathbb{W},\mathbb{U}} : \mathbb{V} \mapsto \mathbb{W}$.

Hence S is an operator on $\mathbb{W}$.

Since $\dim \mathbb{U} = 2$, $\dim \mathbb{W} = \dim \mathbb{V} - 2 = n - 2$, so the inductive hypothesis applies to S.

By the inductive hypothesis, S has an eigenvalue λ with eigenvector w. Hence

$$(S - \lambda I)w = 0 \tag{14.15}$$

Consider vectors of the form

$$v = u + aw, \quad u \in \mathbb{U}, \;\; a \in \mathbb{R} \tag{14.16}$$

where w is still the eigenvector of S with eigenvalue λ. Then

$$\begin{aligned}
(T - \lambda I)v &= (T - \lambda I)(u + aw) && (14.17)\\
&= Tu - \lambda u + aTw - a\lambda w && (14.18)\\
&= Tu - \lambda u + a(Tw - \lambda w) && (14.19)\\
&= Tu - \lambda u + a(P_{\mathbb{U},\mathbb{W}}Tw + P_{\mathbb{W},\mathbb{U}}Tw - \lambda w) && (14.20)\\
&= Tu - \lambda u + a(P_{\mathbb{U},\mathbb{W}}Tw + Sw - \lambda w) && (14.21)\\
&= Tu - \lambda u + aP_{\mathbb{U},\mathbb{W}}Tw \in \mathbb{U} && (14.22)
\end{aligned}$$

The last statement, that the right hand side is in $\mathbb{U}$ follows from the following: the first term is in $\mathbb{U}$ because $\mathbb{U}$ is invariant under T; the second term is in $\mathbb{U}$ because $u \in \mathbb{U}$; and the third term is in $\mathbb{U}$ because, by definition, it is a projection onto $\mathbb{U}$. Hence

$$(T - \lambda I) : (\mathbb{U} + \operatorname{span}(w)) \mapsto \mathbb{U} \tag{14.23}$$

This is a mapping of higher dimensional subspace to a lower dimensional subspace because

$$\dim \mathbb{U} < \dim(\mathbb{U} + \operatorname{span}(w)) \tag{14.24}$$

By corollary 8.20

$$(T - \lambda I)|_{\mathbb{U}+\operatorname{span}(w)} \tag{14.25}$$

is not injective, hence its null space is non-zero (theorem 8.17).

Hence there exists a nonzero vector $v \in \mathbb{U} + \operatorname{span}(w)$ such that

$$(T - \lambda I)v = 0 \tag{14.26}$$

Thus T has an eigenvalue.

□

Topic 15

Inner Products and Norms

Definition 15.1 Let $\mathbb{V}$ be a finite dimensional non-zero vector space over $\mathbb{F}$. An **inner product** on $\mathbb{V}$ is a function

$$\langle u, v \rangle : \mathbb{V} \times \mathbb{V} \mapsto \mathbb{F} \tag{15.1}$$

with the following properties:

1. **Positivity**: For all $v \in \mathbb{V}$,
$$\langle v, v \rangle \geq 0 \tag{15.2}$$

2. **Definiteness**:
$$\langle v, v \rangle = 0 \iff v = 0 \tag{15.3}$$

3. **Additivity in first variable**: for all $u, v, w \in \mathbb{V}$,
$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle \tag{15.4}$$

4. **Homogeneity in first variable**: for all $v, w \in \mathbb{V}$ and all $a \in \mathbb{F}$,

$$\langle av, w \rangle = a\langle v, w \rangle \tag{15.5}$$

5. **Conjugate Symmetry**: for all $v, w \in \mathbb{V}$,

$$\langle v, w \rangle = \overline{\langle w, v \rangle} \tag{15.6}$$

Definition 15.2 An **inner product space** is a vector field with an inner product.

Example 15.1 Let $\mathbb{F}$ be the real numbers and $\mathbb{V} = \mathbb{R}^2$. Then the dot product is an inner product, and $\mathbb{R}^2$ is an inner product space.

Example 15.2 Let $\mathbb{P}_m(\mathbb{F})$ be the set of all polynomials with coefficients in $\mathbb{F}$. Then for $p, q \in \mathbb{P}_m(\mathbb{F})$define

$$\langle p, q \rangle = \int_0^1 p(x)\overline{q(x)}dx \tag{15.7}$$

Theorem 15.3 Properties of Inner Products

1. Inner Product with zero:

$$\langle 0, w \rangle = \langle w, 0 \rangle = 0 \tag{15.8}$$

2. Additivity in the second slot:

$$\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle \tag{15.9}$$

3. Congugate Homogeneity in the second slot:

$$\langle u, av \rangle = a^*\langle u, v \rangle \tag{15.10}$$

Proof. Proof of (1). By additivity in the first slot,

$$\langle 0, w\rangle = \langle 0+0, w\rangle = \langle 0, w\rangle + \langle 0, w\rangle \tag{15.11}$$

Hence $\langle 0, w\rangle = 0$. By conjugate symmetry,

$$\langle w, 0\rangle = \overline{\langle 0, w\rangle} = \overline{0} = 0 \tag{15.12}$$

Proof of (2)

$$\begin{aligned}
\langle u, v+w\rangle &= \overline{\langle v+w, u\rangle} \quad \text{(conjugate symmetry)} && (15.13)\\
&= \overline{\langle v, u\rangle + \langle w, u\rangle} \quad \text{(additivity in 1st slot)} && (15.14)\\
&= \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle} \quad \text{(additivity of comp. conjugate)} && (15.15)\\
&= \langle u, v\rangle + \langle u, w\rangle \quad \text{(conjugate symmetry)} && (15.16)
\end{aligned}$$

Proof of (3)

$$\begin{aligned}
\langle u, av\rangle &= \overline{\langle av, u\rangle} \quad \text{(conjugate symmetry)} && (15.17)\\
&= \overline{a\langle v, u\rangle} \quad \text{(homgeneity in first slot)} && (15.18)\\
&= \overline{a}\,\overline{\langle v, u\rangle} \quad \text{(property of complex numbers)} && (15.19)\\
&= \overline{a}\langle u, v\rangle \quad \text{(conjugate symmetry)} && \square
\end{aligned}$$

Definition 15.4 Let $v \in \mathbb{V}$ be a vector. Then the **norm of the vector** is

$$\|v\| = \sqrt{\langle v, v\rangle} \tag{15.20}$$

Example 15.3 The Euclidean norm is

$$\|(z_1, \dots, z_n)\| = \sqrt{z_1 z_1^* + \cdots + z_n z_n^*} \tag{15.21}$$

Example 15.4 Let $p \in \mathbb{P}_m(\mathbb{F})$; then

$$\|p\| = \sqrt{\int_0^1 |p(x)|^2 dx} \tag{15.22}$$

Theorem 15.5 $\|v\| = 0 \iff v = 0$

Proof. . This follows from the definiteness of the inner product. □

Theorem 15.6 $\|av\| = |a|\|v\|$ where $a \in \mathbb{F}$ and $v \in \mathbb{V}(\mathbb{F})$.

Proof.

$$\|av\|^2 = \langle av, av\rangle = a\langle v, av\rangle = aa^*\langle v, v\rangle = |a|^2\|v\|^2 \tag{15.23}$$

□

Definition 15.7 Two vectors are called **orthogonal** if $\langle u, v\rangle = 0$.

Theorem 15.8 Pythagorean Theorem If u, v are orthogonal vectors then

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 \tag{15.24}$$

Proof.

$$\begin{aligned} \|u+v\|^2 &= \langle u+v, u+v\rangle && (15.25)\\ &= \langle u, u+v\rangle + \langle v, u+v\rangle && (15.26)\\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle && (15.27)\\ &= \|u\|^2 + \|v\|^2 && (15.28)\end{aligned}$$

where the last step follows becaue $\langle v, u\rangle = \langle u, v\rangle = 0$, by orthogonality. □

Definition 15.9 An **orthogonal decomposition** of a vector is a decomposition of the vector as a sum of orthogonal vectors:

$$v = v_1 + v_2 + \cdots + v_n \tag{15.29}$$

where $\langle v_i, v_j\rangle = 0$ unless $i = j$.

Example 15.5 In Euclidean space $\mathbb{R}^3$ we can decompose a vector into components parallel to the three Cartesian axes:

$$(x, y, z) = (x, 0, 0) + (0, y, 0) + (0, 0, z) \tag{15.30}$$

Example 15.6 In Euclidian space we commonly decompose a vector u into a part that is parallel to a second vector b and one that is orthogonal to it. These are given by the dot product (the parallel component) and the original vector minus the parallel component:

$$u = (v \cdot u)v + (u - (v \cdot u)v) \tag{15.31}$$

We would like to generalize example 15.6 to general vector spaces:

$$u = av + (u - av) \tag{15.32}$$

where av is the "part" of u that is "parallel" to v and $(u - av)$ is the part that is orthogonal to v. (By "parallel" we mean a scalar multiple of v.) To force orthogonality of the second part of equation 15.32 to v means

$$0 = \langle u - av, v\rangle = \langle u, v\rangle + \langle -av, v\rangle = \langle u, v\rangle - a\|v\|^2 \tag{15.33}$$

hence

$$a = \frac{\langle u, v\rangle}{\|v\|^2} \tag{15.34}$$

Hence from equation 15.32 we have the following.

Theorem 15.10 Orthogonal Decomposition. Let $u, v \neq 0$ be vectors. Then u can be decomposed into the sum of a scalar multiple of v and an a vector that is orthogonal to v as follows:

$$u = \frac{\langle u, v\rangle}{\|v\|^2}v + \left(u - \frac{\langle u, v\rangle}{\|v\|^2}v\right) \tag{15.35}$$

The first term in equation 15.35 is a scalar multiple of v and the second term is orthogonal to v.

Theorem 15.11 Cauchy-Schwarz Inequality. Let $u, v \in \mathbb{V}$ be vectors. Then

$$|\langle u, v\rangle| \leq \|u\|\|v\| \tag{15.36}$$

Proof. If $v = 0$ then both sides are zero and the result holds. So assume that $v \neq 0$.

By the orthogonal decomposition theorem (theorem 15.10) we can write

$$u = \frac{\langle u, v\rangle}{\|v\|^2}v + \left(u - \frac{\langle u, v\rangle}{\|v\|^2}v\right) = \frac{\langle u, v\rangle}{\|v\|^2}v + w \tag{15.37}$$

where w is orthogonal to v. By the Pythagorean Theorem (theorem 15.8),

$$\|u\|^2 = \left\|\frac{\langle u, v\rangle}{\|v\|^2}v + w\right\|^2 \tag{15.38}$$

$$= \left\|\frac{\langle u, v\rangle}{\|v\|^2}v\right\|^2 + \|w\|^2 \tag{15.39}$$

$$= \frac{|\langle u, v\rangle|^2}{\|v\|^2} + \|w\|^2 \tag{15.40}$$

$$\geq \frac{|\langle u, v\rangle|^2}{\|v\|^2} \tag{15.41}$$

$$\|u\|^2\|v\|^2 \geq |\langle u, v\rangle|^2 \tag{15.42}$$

Taking square roots gives equaton 15.36. □

Example 15.7 Using the inner product of example 15.2, the Cauchy-Schwarz ineqality tells us that

$$\left|\int_0^1 p(x)\overline{q(x)}dx\right|^2 = |\langle p, q\rangle|^2 \tag{15.43}$$

$$\leq \|p\|^2\|q\|^2 \tag{15.44}$$

$$= \left(\int_0^1 |p(x)|^2 dx\right)\left(\int_0^1 |q(x)|^2 dx\right) \tag{15.45}$$

Theorem 15.12 Triangle Inequality. Let $u, v \in \mathbb{V}$. Then $\|u+v\| \leq \|u\| + \|v\|$

Proof.

$$\|u+v\|^2 = \langle u+v, u+v\rangle \tag{15.46}$$
$$= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \tag{15.47}$$
$$= \|u\|^2 + \|v\|^2 + \langle u,v\rangle + \overline{\langle u,v\rangle} \qquad \text{(Conjugate Symmetry)} \tag{15.48}$$
$$= \|u\|^2 + \|v\|^2 + 2\,\mathrm{Re}(\langle u,v\rangle) \qquad (\text{because } \mathrm{Re}(z) = (z+\overline{z})/2) \tag{15.49}$$
$$\leq \|u\|^2 + \|v\|^2 + 2|\langle u,v\rangle| \qquad (\text{because } \mathrm{Re}(z) \leq |z|) \tag{15.50}$$
$$\leq \|u\|^2 + \|v\|^2 + 2\|u\|\|v\| \qquad \text{(Cauchy-Schwarz Inequality)} \tag{15.51}$$
$$= (\|u\| + \|v\|)^2 \tag{15.52}$$

Taking square roots of both sides gives the triangle inequality. □

Theorem 15.13 Parallelogram Inequality. If $u, v \in \mathbb{V}$, then

$$\|u+v\|^2 + \|u-v\|^2 = 2\left(\|u\|^2 + \|v\|^2\right) \tag{15.53}$$

Proof.

$$\|u+v\|^2 + \|u-v\|^2 = \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \tag{15.54}$$
$$= \|u\|^2 + \|v\|^2 + \langle u,v\rangle + \langle v,u\rangle$$
$$+ \|u\|^2 + \|v\|^2 - \langle u,v\rangle - \langle v,u\rangle \tag{15.55}$$
$$= 2\left(\|u\|^2 + \|v\|^2\right) \tag{15.56}$$

Taking square roots gives the desired result. □

Topic 16

Fixed Points of Operators

Before we look at fixed points of operators, we review the analogous concept of fixed points for functions on $\mathbb{R}$. Then we will generalize from functions on $\mathbb{R}$ to operators on a vector space $\mathbb{V}$.

Definition 16.1 Let $f : \mathbb{R} \mapsto \mathbb{R}$. A number $a \in \mathbb{R}$ is called a fixed point of f if $f(a) = a$.

Example 16.1 Find the fixed points of the function $f(x) = x^4 + 2x^2 + x - 3$.

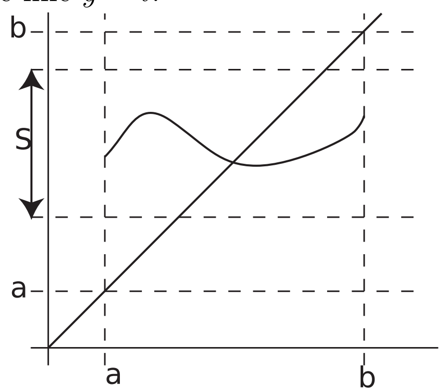
$$\begin{aligned} x &= x^4 + 2x^2 + x - 3 \\ 0 &= x^4 + 2x^2 - 3 \\ &= (x-1)(x+1)(x^2+3) \end{aligned}$$

Hence the real fixed points are $x = 1$ and $x = -1$. □

A function $f : \mathbb{R} \mapsto \mathbb{R}$ has a fixed point if and only if its graph intersects with the line $y = x$. If there are multiple intersections, then there are multiple fixed points.

Consequently a sufficient condition is that the range of f is contained in its domain (see figure 16.1).

Figure 16.1: A sufficient condition for a bounded continuous function to have a fixed point is that the range be a subset of the domain. A fixed point occurs whenever the curve of $f(t)$ intersects the line $y = t$.



Theorem 16.2 (Sufficient condition for fixed point) Suppose that $f(t)$ is a continuous function that maps its domain into a subset of itself, i.e.,

$$f(t) : [a, b] \mapsto S \subset [a, b] \tag{16.1}$$

Then $f(t)$ has a fixed point in $[a, b]$.

Proof. If $f(a) = a$ or $f(b) = b$ then there is a fixed point at either a or b. So assume that both $f(a) \neq a$ and $f(b) \neq b$. By assumption, $f(t) : [a, b] \mapsto S \subset [a, b]$, so that

$$f(a) \geq a \quad \text{and} \quad f(b) \leq b \tag{16.2}$$

Since both $f(a) \neq a$ and $f(b) \neq b$, this means that

$$f(a) > a \quad \text{and} \quad f(b) < b \tag{16.3}$$

Let $g(t) = f(t) - t$. Then g is continuous because f is continuous, and furthermore,

$$g(a) = f(a) - a > 0 \tag{16.4}$$

$$g(b) = f(b) - b < 0 \tag{16.5}$$

Hence by the intermediate value theorem, g has a root $r \in (a, b)$, where $g(r) = 0$. Then

$$0 = g(r) = f(r) - r \implies f(r) = r \tag{16.6}$$

i.e., r is a fixed point of f. □

In the case just proven, there may be multiple fixed points. If the derivative is sufficiently bounded then there will be a unique fixed point.

Theorem 16.3 (Condition for a unique fixed point) Let f be a continuous function on $[a, b]$ such that $f : [a, b] \mapsto S \subset (a, b)$, and suppose further that there exists some postive constant $K < 1$ such that

$$|f'(t)| \leq K, \quad \forall t \in [a, b] \tag{16.7}$$

Then f has a unique fixed point in $[a, b]$.

Proof. By theorem 16.2 a fixed point exists. Call it p,

$$p = f(p) \tag{16.8}$$

Suppose that a second fixed point $q \in [a, b]$, $q \neq p$ also exists, so that

$$q = f(q) \tag{16.9}$$

Hence

$$|f(p) - f(q)| = |p - q| \tag{16.10}$$

By the mean value theorem there is some number c between p and q such that

$$f'(c) = \frac{f(p) - f(q)}{p - q} \tag{16.11}$$

Taking absolute values,

$$\left|\frac{f(p) - f(q)}{p - q}\right| = |f'(c)| \leq K < 1 \tag{16.12}$$

and thence

$$|f(p) - f(q)| < |p - q| \tag{16.13}$$

This contradicts equation 16.10. Hence our assumption that a second, different fixed point exists must be incorrect. Hence the fixed point is unique. □

Theorem 16.4 Let f be as defined in theorem 16.3, and $p_0 \in (a, b)$. Then the sequence of numbers

$$\left.\begin{aligned} p_1 &= f(p_0) \\ p_2 &= f(p_1) \\ &\vdots \\ p_n &= f(p_{n-1}) \\ &\vdots \end{aligned}\right\} \tag{16.14}$$

converges to the unique fixed point of f in (a, b).

Proof. We know from theorem 16.3 that a unique fixed point p exists. We need to show that $p_i \to p$ as $i \to \infty$.

Since f maps onto a subset of itself, every point $p_i \in [a, b]$.

Further, since p itself is a fixed point, $p = f(p)$ and for each i, since $p_i = f(p_{i-1})$, we have

$$|p_i - p| = |p_i - f(p)| = |f(p_{i-1}) - f(p)| \tag{16.15}$$

If for any value of i we have $p_i = p$ then we have reached the fixed point and the theorem is proved.

So we assume that $p_i \neq p$ for all i.

Then by the mean value theorem, for each value of i there exists a number c_i between p_{i-1} and p such that

$$|f(p_{i-1}) - f(p)| = |f'(c_i)||p_{i-1} - p| \leq K|p_{i-1} - p| \tag{16.16}$$

where the last inequality follows because f' is bounded by $K < 1$ (see equation 16.7).

Substituting equation 16.15 into equation 16.16,

$$|p_i - p| = |f(p_{i-1}) - f(p)| \leq K|p_{i-1} - p| \tag{16.17}$$

Restating the same result with i replaced by $i - 1, i - 2, \dots,$

$$\left.\begin{aligned}
|p_{i-1} - p| = |f(p_{i-2}) - f(p)| \leq K|p_{i-2} - p| \\
|p_{i-2} - p| = |f(p_{i-3}) - f(p)| \leq K|p_{i-3} - p| \\
|p_{i-4} - p| = |f(p_{i-4}) - f(p)| \leq K|p_{i-4} - p| \\
\vdots \\
|p_2 - p| = |f(p_1) - f(p)| \leq K|p_1 - p| \\
|p_1 - p| = |f(p_0) - f(p)| \leq K|p_0 - p|
\end{aligned}\right\} \tag{16.18}$$

Putting all these together,

$$|p_i - p| \le K^2|p_{i-2} - p| \le K^3|p_{i-2} - p| \le \cdots \le K^i|p_0 - p| \tag{16.19}$$

Since $0 < K < 1$,

$$0 \le \lim_{i\to\infty} |p_i - p| \le |p_0 - p| \lim_{i\to\infty} K^i = 0 \tag{16.20}$$

Thus $p_i \to p$ as $i \to \infty$. □

A weaker condition that is sufficient for convergence is the Lipshitz condition.

Definition 16.5 A function f on $I \subset R$ is said to be **Lipshitz** (or **Lipshitz continuous**, or **satisfy a Lipshitz condition**) on y if there exists some constant $K > 0$ if for all $x \in I$ then

$$|f(x_1) - f(x_2)| \le K|x_1 - x_2| \tag{16.21}$$

The constant K is called a **Lipshitz Constant** for f.

Theorem 16.6 Under the same conditions as theorem 16.4 except that the condition of equation 16.7 is replaced with the following condition: $f(t)$ is Lipshitz with Lipshitz constant $K < 1$. Then fixed point iteration converges.

Proof. Lipshitz gives equation 16.16. The rest of the the proof follows as before. □

Now we turn to a discussion of fixed points on a general vector space. The Lipshitz condition can be generalized for a multivariate function to apply to a single variable. We will place the variable in the final slot of f. We state it for a two-variable function, but in fact, we could replace t in the following definition to $t_1, t_2, ...$

Definition 16.7 A function $f(t, y)$ on D is said to be **Lipshitz** in the variable y if there exists some constant $K > 0$ if for all $(x, y_1), (x, y_2) \in D$ then

$$|f(x, y_1) - f(x, y_2)| \leq K|y_1 - y_2| \tag{16.22}$$

The constant K is called a **Lipshitz Constant** for f. We will sometimes denote this as $f \in \mathcal{L}(y; K)(D)$.

Theorem 16.8 Suppose that $|\partial f/\partial y|$ is bounded by K on a set D. Then $f(t, y) \in \mathcal{L}(y, K)(D)$.

Proof. The result follows immediately from the mean value theorem. Let $(t, y_1), (t, y_2) \in D$. Then there is some number c between y_1 and y_2 such that

$$|f(t, y_1) - f(t, y_2)| = |f_y(c)||y_1 - y_2| < K|y_1 - y_2| \tag{16.23}$$

Hence f is Lipshitz in y on D. □

Definition 16.9 Let $\mathbb{V}$ be a normed vector space, $S \subset \mathbb{V}$, and $v, w \in S$. Then a **contraction** is any mapping $T : S \mapsto \mathbb{V}$ that satisfies

$$\|Tv - Tw\| \leq K\|v - w\| \tag{16.24}$$

form some $K \in \mathbb{R}$, $0 < K < 1$, for all $v, w \in S$. We will call the number K the **contraction constant**.

Definition 16.10 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let $v_1, v_2, \ldots$ be a sequence in $\mathbb{V}$. Then we say that the sequence is **Cauchy** if $\|v_m - v_n\| \to 0$ as $n, m \to \infty$. More precisely, for every $\epsilon > 0$ there exists some $N > 0$ such that for every $m, n > N$, we have $\|v_m - v_n\| < \epsilon$.

The study of Complete spaces and Cauchy sequences is beyond the scope of this class. We will just assume that we are working in a vector space in which Cauchy sequences converge.

Definition 16.11 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$. Then we say that $\mathbb{V}$ is **complete** if every Cauchy sequence in $\mathbb{V}$ converges to some element of $\mathbb{V}$.

Lemma 16.12 Let T be a contraction on a complete normed vector space $\mathbb{V}$ with contraction constant K. Then for any $v \in \mathbb{V}$

$$\|T^n v - v\| \leq \frac{1-K^n}{1-K}\|Tv - v\| \tag{16.25}$$

Proof. Use induction. For $n = 1$, the formula gives

$$\|Tv - v\| \leq \frac{1-K}{1-K}\|Tv - v\| \tag{16.26}$$

which is trivially true.

As our inductive hypothesis choose any $n > 1$ and suppose that equation 16.25 holds.

Then by the triangle inequality

$$\|T^{n+1}v - v\| \leq \|T^{n+1}v - T^n v\| + \|T^n v - v\| \tag{16.27}$$

$$\leq \|T^{n+1}v - T^n v\| + \frac{1-K^n}{1-K}\|Tv - v\| \text{ (by the inductive hyp.)} \tag{16.28}$$

$$= \|T^n Tv - T^n v\| + \frac{1-K^n}{1-K}\|Tv - v\| \tag{16.29}$$

$$\leq K^n\|Tv - v\| + \frac{1-K^n}{1-K}\|Tv - v\| \text{ (because } T \text{ is a contraction)} \tag{16.30}$$

$$= \frac{(1-K)K^n + (1-K^n)}{1-K}\|Tv - v\| \tag{16.31}$$

$$= \frac{1-K^{n+1}}{1-K}\|Tv - v\| \tag{16.32}$$

which proves the conjecture for $n+1$. □

Definition 16.13 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let T be an operator on $\mathbb{V}$. Then we say v is a **fixed point** of T if $Tv = v$.

Theorem 16.14 Contraction Mapping Theorem[1] Let T be a contraction on a normed vector space V. Then T has a unique fixed point $u \in \mathbb{V}$ such that $Tu = u$. Furthermore, any sequence of vectors $v_1, v_2, \dots$ defined by $v_k = Tv_{k-1}$ converges to the unique fixed point $Tu = u$. We denote this by $v_k \to u$.

Proof. [2] Let $\epsilon > 0$ be given and let $v \in \mathbb{V}$.

Since $K^n/(1-K) \to 0$ as $n \to \infty$ (because T is a contraction, $K < 1$), given any $v \in \mathbb{V}$, it is possible to choose an integer N such that

$$\frac{K^n\|Tv - v\|}{1-K} < \epsilon \tag{16.33}$$

[1]The contraction mapping theorem is sometimes called the Banach Fixed Point Theorem.

[2]The proof follows "Proof of Banach Fixed Point Theorem," *Encyclopedia of Mathematics* (Volume 2, 54A20:2034), PlanetMath.org.

for all $n > N$. Pick any such integer N.

Choose any two integers $m \geq n \geq N$, and define the sequence

$$\left.\begin{aligned} v_0 &= v \\ v_1 &= Tv \\ v_2 &= Tv_1 \\ &\vdots \\ v_n &= Tv_{n-1} \\ &\vdots \end{aligned}\right\} \tag{16.34}$$

Then since T is a contraction,

$$\|v_m - v_n\| = \|T^m v - T^n v\| \tag{16.35}$$

$$= \|T^n T^{m-n} v - T^n v\| \tag{16.36}$$

$$\leq K^n \|T^{m-n} v - v\| \tag{16.37}$$

From Lemma 16.12 we have

$$\|v_m - v_n\| \leq K^n \frac{1 - K^{m-n}}{1 - K} \|Tv - v\| \tag{16.38}$$

$$= \frac{K^n - K^m}{1 - K} \|Tv - v\| \tag{16.39}$$

$$\leq \frac{K^n}{1 - K} \|Tv - v\| < \epsilon \tag{16.40}$$

Therefore v_n is a Cauchy sequence, and every Cauchy sequence on a complete normed vector space converges. Hence $v_n \to u$ for some $u \in \mathbb{V}$.

Either u is a fixed point of T or it is not a fixed point of T.

Suppose that u is not a fixed point of T. Then $Tu \neq u$ and hence there exists some $\delta > 0$ such that

$$\|Tu - u\| > \delta \tag{16.41}$$

On the other hand, because $v_n \to u$, there exists an integer N such that for all $n > N$,

$$\|v_n - u\| < \delta/2 \tag{16.42}$$

Hence

$$\begin{aligned}
\|Tu - u\| &\leq \|Tu - v_{n+1}\| + \|v_{n+1} - u\| && (16.43)\\
&= \|Tu - Tv_n\| + \|u - v_{n+1}\| && (16.44)\\
&\leq K\|u - v_n\| + \|u - v_{n+1}\| \qquad \text{(because } T \text{ is a contraction)} && (16.45)\\
&\leq \|u - v_n\| + \|u - v_{n+1}\| \qquad \text{(because } K < 1\text{)} && (16.46)\\
&= 2\|u - v_n\| && (16.47)\\
&< \delta && (16.48)
\end{aligned}$$

This is a contradiction. Hence u must be a fixed point of T.

To prove uniqueness, suppose that there is another fixed point $w \neq u$.

Then $\|w - u\| > 0$ (otherwise they are equal). But

$$\|u - w\| = \|Tu - Tw\| \leq K\|u - w\| < \|u - w\| \tag{16.49}$$

which is impossible and hence and contradiction.

Thus u is the unique fixed point of T. □

Lemma 16.15 Let $\mathbb{V}$ be the vector space consisting of integrable functions on an interval (a, b), and let $f \in \mathbb{V}$. Then the **sup-norm** defined by

$$\|f\|_\infty = \sup\{|f(x)| : x \in (a, b)\} \tag{16.50}$$

is a norm.

The proof is left as an exercise.

The notation for the sup-norm comes from the **p-norm**,

$$\|f\|_p = \left(\int_a^b |f(x)|^p dx\right)^{1/p} \tag{16.51}$$

then

$$\lim_{p\to\infty} \|f\|_p = \|f\|_\infty \tag{16.52}$$

See any text on analysis for a discussion of this.

Theorem 16.16 Existence of Solutions to the Initial Value Problem. Let $D \in \mathbb{R}^2$ be convex and suppose that f is continuously differentiable on D. Then the initial value problem

$$y' = f(t, y), \quad y(t_0) = y_0 \tag{16.53}$$

has a unique solution $\phi(t)$ in the sense that $\phi'(t) = f(t, \phi(y)), \phi(t_0) = y_0$.

Proof. We begin by observing that ϕ is a solution of equation 16.53 if and only if it is a solution of

$$\phi(t) = y_0 + \int_{t_0}^t f(x, \phi(x))dx \tag{16.54}$$

Our goal will be to prove 16.54.

Let $\mathbb{V}$ be the set of all continuous integrable functions on an interval (a, b) that contains t_0.

Define the map $T \in \mathcal{L}(\mathbb{V})$:$\phi \mapsto T\phi$ by

$$T\phi = y_0 + \int_{t_0}^{t} f(x, \phi(x))dx \tag{16.55}$$

We will assume $b \geq t > t_0 \geq a$. The proof for $t < t_0$ is completely analogous.

We will use the sup-norm on (a, b). Let $g, h \in \mathbb{V}$.

$$\begin{aligned} \|Tg - Th\|_\infty &= \sup_{a \leq t \leq b} |Tg - Th| && (16.56) \\ &= \sup_{a \leq t \leq b} \left| y_0 + \int_{t_0}^{t} f(x, g(x))dx - y_0 - \int_{t_0}^{t} f(x, h(x))dx \right| && (16.57) \\ &= \sup_{a \leq t \leq b} \left| \int_{t_0}^{t} \left[f(x, g(x)) - f(x, h(x)) \right] dx \right| && (16.58) \end{aligned}$$

Since f is continuously differentiable it is differentiable and its derivative is continuous. Thus the derivative is bounded (otherwise it could not be continuous on all of (a, b)). Therefore by theorem 16.8, it is Lipshitz in its second argument. Consequently there is some $K \in \mathbb{R}$ such that

$$\begin{aligned} \|Tg - Th\|_\infty &\leq L \sup_{a \leq t \leq b} \int_{t_0}^{t} |g(x) - h(x)|\, dx && (16.59) \\ &\leq K(t - t_0) \sup_{a \leq t \leq b} |g(x)) - h(x)| && (16.60) \\ &\leq K(b - a) \sup_{a \leq t \leq b} |g(x)) - h(x)| && (16.61) \\ &\leq K(b - a) \|g - h\| && (16.62) \end{aligned}$$

Since K is fixed, so long as the interval (a, b) is larger than $1/K$ we have

$$\|Tg - Th\|_\infty \leq K'\|g - h\|_\infty \tag{16.63}$$

where

$$K' = K(b - a) < 1 \tag{16.64}$$

Thus T is a contraction. By the contraction mapping theorem it has a fixed point; call this point ϕ. Equation 16.54 follows immediately. □

Topic 17

Orthogonal Bases

Definition 17.1 A list of vectors $B = (v_1, \dots, v_n)$ is called **orthonormal** if (a) $\|v_i\| = 1$; and $\langle v_i, v_j \rangle = 0$ for $i \neq j$. If B is also a basis of $\mathbb{V}$ then it is called an **orthonormal basis** of $\mathbb{V}$.

Theorem 17.2 Let $B = (e_1, \dots, e_m)$ be an orthonormal list of vectors in $\mathbb{V}$. Then

$$\|a_1e_1 + a_2e_2 + \cdots + a_me_m\|^2 = |a_1|^2 + |a_2|^2 + \cdots + |a_m|^2 \tag{17.1}$$

for all $a_1, a_2, \dots \in \mathbb{F}$.

Proof. This follows immediately from the Pythagorean theorem. □

Theorem 17.3 Let $B = (e_1, \dots, e_m)$ be and orthonormal list of vectors in $\mathbb{V}$. Then B is linearly independent.

Proof. Suppose there exist $a_1, \dots, a_m \in \mathbb{F}$ such that

$$0 = a_1e_1 + \cdots + a_me_m \tag{17.2}$$

Then by theorem 17.2

$$0 = \|a_1e_1 + \cdots + a_m e_m\| = |a_1|^2 + \cdots + |a_m|^2 \tag{17.3}$$

Hence $a_1 = a_2 = \cdots = a_m = 0$, and therefore B is linearly independent. □

Definition 17.4 Kronecker Delta Function.[1]

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \tag{17.4}$$

Theorem 17.5 Let $B = (e_1, \dots, e_n)$ be an orthonormal basis of $\mathbb{V}$. Then both of the following are true for every $v \in \mathbb{V}$:

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n \tag{17.5}$$

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 \tag{17.6}$$

Proof. Equation 17.6 follows from equation 17.5 and theorem 17.2.

To prove equation 17.5 pick any $v \in \mathbb{V}$. Since B is a basis of $\mathbb{V}$ there exist scalars $a_1, \dots, a_n$ such that

$$v = a_1e_1 + \cdots + a_ne_n \tag{17.7}$$

Hence

$$\langle v, e_j\rangle = \sum_{i=1}^{n} a_i\langle e_i, e_j\rangle = \sum_{i=1}^{n} a_i\delta_{ij} = a_j \tag{17.8}$$

Substituting equation 17.8 into equation 17.7 for each a_j gives equation 17.5. □

[1] Named for Leopold Kronecker (1823-1891).

Theorem 17.6 Gram-Schmidt Orthonormalization Procedure. Let $A = (v_1, \ldots, v_n)$ be a linearly independent list of vectors in $\mathbb{V}$. Then there exists an orthonormal list of vectors $B = (e_1, \ldots, e_n)$ in $\mathbb{V}$ such that

$$\text{span}(v_1, \ldots, v_j) = \text{span}(e_1, \ldots, e_j) \quad \forall j = 1, \ldots, n \tag{17.9}$$

Proof. The proof is constructive. We start by defining

$$e_1 = \frac{v_1}{\|v_1\|} \tag{17.10}$$

and then define e_j inductively for $j > 1$ from the $e_1, \ldots, e_{j-1}$. Cleary $\|e_1\| = 1$.

To illustrate this we construct the first few. We define e_2 as the part of v_2 that is orthogonal to e_1:

$$e_2 = \frac{v_2 - \langle v_2, e_1\rangle e_1}{\|v_2 - \langle v_2, e_1\rangle e_1\|} \tag{17.11}$$

which by construction satisfies $\|e_2\| = 1$ and $\langle e_2, e_1\rangle = 0$. Furthermore, $\text{span}(v_1, v_2) = \text{span}(e_1, e_2)$.

Next, we define e_3 as the part of v_3 that is orthogonal to both e_1 and e_2:

$$e_3 = \frac{v_3 - \langle v_3, e_1\rangle e_1 - \langle v_3, e_2\rangle e_2}{\|v_3 - \langle v_3, e_1\rangle e_1 - \langle v_3, e_2\rangle e_2\|} \tag{17.12}$$

Again, by construction, $\|e_3\| = 1, \langle e_3, e_1\rangle = \langle e_3, e_2\rangle = 0$ and $\text{span}(e_1, e_2, e_3) = \text{span}(v_1, v_2, v_3)$.

In general we define each e_j to have length 1 and be orthogonal to the $e_1, e_2, ..., e_{j-1}$ by the following equation:

$$e_j = \frac{v_j - \sum_{i=1}^{j-1} \langle v_j, e_i \rangle e_i}{\| v_j - \sum_{i=1}^{j-1} \langle v_j, e_i \rangle e_i \|} \tag{17.13}$$

To prove that $\text{span}(v_1, \dots, v_n) \subset \text{span}(e_1, \dots, e_n)$ let

$$v = \sum a_i v_i \tag{17.14}$$

and substitute to get an expression for v in terms of the e_i. To show that $\text{span}(e_1, \dots, e_n) \subset \text{span}(v_1, \dots, v_n)$ make the reverse subsitution, $v = \sum b_i e_i$ and rearrange to get an expression in terms of the v_i. The exact calculation is left as an exercise. This proves that the two spans are equal. □

Corollary 17.7 Let $\mathbb{V}$ be a finite-dimensional inner-product space. Then $\mathbb{V}$ has an orthonormal basis.

Proof. Pick any basis B of $\mathbb{V}$. Since B is linearly independent, we can find an orthonormal list of vectors B' with the same span, by the Gram-Schmidt orthonormalization process. Since B' spans $\mathbb{V}$ and is orthonormal, it is an orthonormal basis. □

Corollary 17.8 Let $\mathbb{V}$ be a finite dimensional inner-product space and let $B = (b_1, \dots, b_n)$ be an orthornormal list of vectors in $\mathbb{V}$. Then B can be extended to an orthonormal basis

$$E = (b_1, \dots, b_n, e_1, \dots, e_n) \tag{17.15}$$

of $\mathbb{V}$. In particular, if $\mathbb{U}$ is a substace of $\mathbb{V}$ with an orthonormal basis B, then B can be extended to a basis of $\mathbb{V}$.

Corollary 17.9 Let $\mathbb{V}$ be a finite dimensional inner-product space; let $T \in \mathcal{L}(\mathbb{V})$ be an operator on $\mathbb{V}$; and suppose that there exists a basis B of $\mathbb{V}$ such that $\mathcal{M}(T, B)$ is upper-triangular. Then there exists an orthonormal basis B' of $\mathbb{V}$ such that $\mathcal{M}(T, B')$ is upper-triangular.

Corollary 17.10 Let $\mathbb{V}$ be a complex vector space and let $T \in \mathcal{L}(\mathbb{V})$. Then T has an upper triangular matrix with respect to some orthonormal basis of $\mathbb{V}$.

Proof. From theorem 12.3, there exists some basis under which $\mathcal{M}(T)$ is uper triangular.

By Corollary 17.9 $\mathcal{M}(T)$ is upper triangular with respect to some orthonormal basis.

□

Definition 17.11 Let $U \subset \mathbb{V}$. Then the **orthogonal complement of** U is the set of all vectors in $\mathbb{V}$ that are orthogonal to all vectors in U:

$$U^{\perp} = \{v \in \mathbb{V} | \langle v, u \rangle = 0, \forall u \in U\} \tag{17.16}$$

Remark 17.12 Some properties of orthogonal complements:

1. $\mathbb{V}^{\perp} = \{0\}$
2. $\{0\}^{\perp} = \mathbb{V}$
3. $U \subset W \implies W^{\perp} \subset U^{\perp}$

Theorem 17.13 Let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Then $\mathbb{V} = \mathbb{U} \oplus \mathbb{U}^{\perp}$.

Proof. Let $v \in \mathbb{V}$ and let $B = (e_1, \ldots, e_n)$ be a orthonormal basis of $\mathbb{U}$. Then

$$v = (\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n) + v - (\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n) \tag{17.17}$$

Let

$$\left.\begin{aligned} u &= \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n \\ w &= v - u \end{aligned}\right\} \tag{17.18}$$

Since B is a basis of $\mathbb{U}$, $u \in \mathbb{U}$, and

$$\langle w, e_j\rangle = \langle v, e_j\rangle - \langle v, u\rangle = \langle v, e_j\rangle - \langle v, e_j\rangle = 0 \tag{17.19}$$

Hence $w \perp u$ for all $u \in \text{span}(B)$ and thus $w \in \mathbb{U}^\perp$.

Thus $v = u + w$ where $u \in \mathbb{U}$ and $w \in \mathbb{U}^\perp$. Hence $\mathbb{V} = \mathbb{U} + \mathbb{U}^\perp$.

To show that the sum is actually a direct sum, let $v \in \mathbb{U} \cap \mathbb{U}^\perp$.

Thus $v \in \mathbb{U}$ and v is orthogonal to every vector in $\mathbb{U}$ because $v \in \mathbb{U}^\perp$.

In particular, v is orthogonal to itself, which means $\langle v, v\rangle = 0$. Hence $v = 0$.

Therefore

$$\mathbb{U} \cap \mathbb{U}^\perp = \{0\} \tag{17.20}$$

Hence by theorem 4.8

$$\mathbb{V} = \mathbb{U} \oplus \mathbb{U}^\perp \tag{17.21}$$

□

Corollary 17.14 If $\mathbb{U}$ is a subspace of $\mathbb{V}$ then $U = (\mathbb{U}^\perp)^\perp$.

Definition 17.15 Supose that $\mathbb{U}$ is a subspace of $\mathbb{V}$, and $v = u + w$, where $u \in \mathbb{U}$ and $w \in \mathbb{U}^\perp$. Then the **orthogonal projection** of $\mathbb{V}$ onto $\mathbb{U}$ is given by the operator $P_\mathbb{U}$ where $P_\mathbb{U} v = u$. (In the notation of definition 14.2, $P_\mathbb{U} = P_{\mathbb{U},\mathbb{U}^\perp}$.)

Remark 17.16 Properties of the Orthogonal Projection. Let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Then

1. range $P_{\mathbb{U}} = \mathbb{U}$
2. null $P_{\mathbb{U}} = \mathbb{U}^{\perp}$
3. $v - P_{\mathbb{U}}v \in \mathbb{U}^{\perp}$ for every $v \in \mathbb{V}$.
4. $P_{\mathbb{U}}^2 = P_{\mathbb{U}}$
5. $\|P_{\mathbb{U}}v\| \leq \|v\|$ for every $v \in \mathbb{V}$.

Theorem 17.17 Let $\mathbb{U}$ be a subspace of $\mathbb{V}$ and let $v \in \mathbb{V}$. Then

$$\|v - P_{\mathbb{U}}v\| \leq \|v - u\| \tag{17.22}$$

for every $u \in \mathbb{U}$ (The shortest path from a point to a line is a normal).

Proof.

$$\begin{aligned} \|v - P_{\mathbb{U}}v\|^2 &\leq \|v - P_{\mathbb{U}}v\|^2 + \|P_{\mathbb{U}}v - u\|^2 && (17.23)\\ &= \|v - P_{\mathbb{U}}v + P_{\mathbb{U}}v - u\|^2 && \text{Pythagorean Theorem} \quad (17.24)\\ &= \|v - u\|^2 && (17.25) \end{aligned}$$

Equation 17.24 follows the previous line via the Pythagorean theorem, because the first vector $v - P_{\mathbb{U}}v \in \mathbb{U}^{\perp}$ and the second vector, $P_{\mathbb{U}}v - u \in \mathbb{U}$, hence they are perpendicular. □

Topic 18

Fourier Series

Theorem 18.1 Let $\mathbb{V}$ be the set of all integrable functions $f : [a, b] \to \mathbb{C}$ and let f be any positive real-valued function on $[a, b]$. Then $\mathbb{V}$ is a normed inner product space with inner product

$$\langle f, g\rangle = \int_a^b k(x)f(x)\overline{g(x)}dx \tag{18.1}$$

We observe that $\mathbb{V}$ is an infinite-dimensional vector space (consider the set of functions $\{1, x, x^2, x^3, \dots\}$, which is linearly independent). We can extend our definition of a basis to infinite dimensions as follows.

Definition 18.2 Let $\mathbb{V}$ be an infinite dimensional vector space. Then a sequence $f_0, f_1, f_2, \dots \in \mathbb{V}$ is called a **complete basis for** $\mathbb{V}$ if, for every $f \in \mathbb{V}$, there exists a sequence of constants $c_0, c_1, \dots \in \mathbb{C}$ such that

$$f = \sum_{k=0}^{\infty} c_i f_i = c_0 f_0 + c_1 f_1 + c_2 f_2 + \cdots \tag{18.2}$$

and is called a **complete orthonormal basis if** $\langle f_i, f_j \rangle = \delta_{ij}$.[1]

Example 18.1 Let $\mathbb{V}$ be the set of integrable functions on $[-1, 1]$ and consider the sequence of functions $1, x, x^2, x^3, \ldots$. Then by Taylor's theorem, for any $f \in \mathbb{V}$, there exist a sequence of constants $a_0, a_1, \ldots$ given by

$$a_k = \frac{f^{(k)}(0)}{k!} \tag{18.3}$$

such that

$$f(x) = \sum_{k=0}^{\infty} a_k f_k \tag{18.4}$$

Hence the sequence $1, x, x^2, \ldots$ is a complete basis of $\mathbb{V}$. □

Given a complete basis for a vector space along with an inner product and its associated norm, we can use the Gram-Schmidt process to create an orthognal basis for the space.

Example 18.2 Using the inner product

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\overline{g(x)}dx \tag{18.5}$$

on the real vector space defined in the previous example, use the Gram-Schmidt process to find an orthogonal basis from the complete basis $1, x, x^2, \ldots$

Denote the original basis by $f_j = x^j, j = 0, 1, 2, \ldots$ the orthogonal basis by p_j, and the normalized basis by q_j. Then since

$$\|f_0\|^2 = \langle f_0, f_0 \rangle = \int_{-1}^{1} dx = 2 \tag{18.6}$$

[1]Here $\delta_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$ is the Kroneker delta function.

we can define $p_0 = f_0 = 1$ and

$$q_0 = \frac{p_0}{\|p_0\|} = \frac{1}{\sqrt{2}} \tag{18.7}$$

Next we calculate

$$\langle f_1, q_0 \rangle = \frac{1}{\sqrt{2}} \int_{-1}^{1} x^2 dx = 0 \tag{18.8}$$

and thus

$$p_1 = f_1 - \langle f_1, q_0 \rangle q_0 = f_1 = x \tag{18.9}$$

$$\|p_1\|^2 = \int_{-1}^{1} x^2 dx = \frac{2}{3} \tag{18.10}$$

$$q_1 = \frac{p_1}{\|p_1\|} = \sqrt{\frac{3}{2}} x \tag{18.11}$$

Next,

$$p_2 = f_2 - \langle f_2, q_0 \rangle q_0 - \langle f_2, q_1 \rangle q_1 \tag{18.12}$$

$$\langle f_2, q_0 \rangle = \int_{-1}^{1} (x^2) \left(\frac{1}{\sqrt{2}} \right) dx = \frac{1}{\sqrt{2}} \times \frac{2}{3} = \frac{\sqrt{2}}{3} \tag{18.13}$$

$$\langle f_2, q_1 \rangle = \int_{-1}^{1} (x^2) \left(\sqrt{\frac{3}{2}} \right) x dx = 0 \text{ (odd function)} \tag{18.14}$$

$$p_2 = x^2 - \frac{\sqrt{2}}{3} \times \frac{1}{\sqrt{2}} - 0 \times \sqrt{\frac{3}{2}} x = x^2 - \frac{1}{3} \tag{18.15}$$

$$\|p_2\|^2 = \int_{-1}^{1} \left(x^2 - \frac{1}{3} \right)^2 dx = \frac{8}{45} \tag{18.16}$$

$$\|p_2\| = \frac{2}{3} \sqrt{\frac{2}{5}} \tag{18.17}$$

$$q_2 = \frac{p_2}{\|p_2\|} = \frac{3}{2} \sqrt{\frac{5}{2}} \left(x^2 - \frac{1}{3} \right) \tag{18.18}$$

and so forth. □

Remark 18.3 The sequence of orthonormal functions generated in the previous example are related to the Legendre polynomials, which are solutions of the initial value problem

$$\frac{d}{dx}\left[(1-x^2)\frac{d}{dx}P_n(x)\right]+n(n+1)P_n(x)=0, \quad P_n(1)=1 \tag{18.19}$$

in other words, they are eigenfunctions of the operator $T \in \mathcal{L}(V)$, where V is the vector space of functions on $[-1,1]$, given by

$$Tf=[(1-x)^2f']' \tag{18.20}$$

with eigenvalues $n(n+1)$. It turns out the eigenfunctions of certain differential operators will always produce orthogonal bases. See any book on boundary value problems or the Sturm-Liouville operator for more details.

Table of first several Legendre Polynomals, solutions to Legendre's equation (equation 18.19). The Legendre Polynomials can be generated recursively via the relation $(n+1)P_{n+1}(x)=(2n+1)xP_n(x)-nP_{n-1}x$.

n	$P_n(x)$
0	1
1	x
2	$\frac{1}{2}(3x^2-1)$
3	$\frac{1}{2}(5x^3-3x)$
4	$\frac{1}{8}(35x^4-30x^2+3)$
5	$\frac{1}{8}(63x^5-70x^3+15x)$

Theorem 18.4 Let $\mathbb{V}$ be an infinite dimensional real-valued function space on an interval $I \subset \mathbb{R}$ and let $\phi_0, \phi_1, \dots$ be a complete orthonormal basis on $\mathbb{V}$. Then for any $f \in \mathbb{V}$, the **Generalized Fourier Series** of f is given by

$$f \sim \sum_{i=0}^{\infty} \langle f, \phi_i \rangle \phi_i \tag{18.21}$$

Proof. Since $\phi_0, \phi_1, \dots$ is a complete basis, there are some numbers $c_i \in \mathbb{R}$ such that

$$f = \sum_{i=0}^{\infty} c_i \phi_i \tag{18.22}$$

$$\langle f, \phi_j \rangle = \sum_{i=0}^{\infty} c_i \langle \phi_i, \phi_j \rangle = \sum_{i=0}^{\infty} c_i \delta_{ij} = c_j \tag{18.23}$$

Plugging the second equation into the first gives the desired result. □

Remark 18.5 In the previous theorem we overlooked what we mean by convergence of the series $\sum_{k=0}^{\infty} c_k \phi_k$. This is a subtle point that we will not concern ourselves with in this class. In particular, the convergence of the series only satisfies the concept of **convergence in the mean**, namely, that

$$\left\| \sum_{k=0}^{\infty} c_k \phi_k - s_n \right\| \to 0 \text{ as } n \to \infty \tag{18.24}$$

The consequence is that the equality may not hold at a countable number of points, in the sense that at any point x_0, equation 18.21 really means

$$\frac{1}{2}\left[f(x_0^+) + f(x_0^-) \right] = \sum_{i=0}^{\infty} \langle f, \phi_i \rangle \phi_i(x_0) \tag{18.25}$$

Example 18.3 Consider the space of integrable functions on $[-\pi, \pi]$ with an inner product given by

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{g(x)} dx \tag{18.26}$$

Then the set of functions $\phi_k = e^{ikx}, k = 0, \pm 1, \pm 2 \dots$ is a orthonormal because

$$\langle \phi_n, \phi_n \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} dx = 1 \tag{18.27}$$

and for $n \neq m$

$$\begin{aligned}
\langle \phi_n, \phi_m \rangle &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{inx} e^{-imx} dx && (18.28)\\
&= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(n-m)x} dx && (18.29)\\
&= \frac{1}{2\pi} \frac{1}{i(n-m)} e^{i(n-m)x} \Big|_{-\pi}^{\pi} && (18.30)\\
&= \frac{1}{2i\pi(n-m)} \left[e^{i(n-m)\pi} - e^{-i(n-m)\pi} \right] && (18.31)\\
&= \frac{1}{\pi(n-m)} \sin(n-m)\pi = 0 && (18.32)
\end{aligned}$$

The Generalized Fourier Series with respect to this basis[2] is

$$f \sim \sum_{k=-\infty}^{\infty} c_k e^{ikx} \tag{18.33}$$

where

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx \quad \square \tag{18.34}$$

[2] We haven't actually shown that ϕ_j form a basis, only that they are orthonormal. To show that it is a basis we have to show that it spans the space.

Example 18.4 Repeat the previous example with the set of functions

$$\phi_k = \frac{1}{\sqrt{2\pi}}, \frac{\sin kx}{\sqrt{\pi}}, \frac{\cos jx}{\sqrt{\pi}}, k, j = 0, 1, 2, \dots \tag{18.35}$$

and the inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(x)g(x)dx \tag{18.36}$$

The standard notation is to define

$$a_k = \frac{1}{\sqrt{\pi}}\int_{-\pi}^{\pi} f(x)\cos kx dx \tag{18.37}$$

$$b_k = \frac{1}{\sqrt{\pi}}\int_{-\pi}^{\pi} f(x)\sin kx dx \tag{18.38}$$

$$\tag{18.39}$$

Then the Fourier series is

$$f \sim \frac{a_0}{2} + \sum_{k=1}^{\infty}[a_k \cos kx + b_k \sin kx] \quad \Box \tag{18.40}$$

The coefficents in the Fourier series are called the **Fourier Coefficients**

$$c_j = \langle f, \phi_j\rangle \tag{18.41}$$

The following result tells us that the sum of the Fourier coefficients is bounded by the square of the norm, $\|f\|^2$, in the sense that

$$\sum |c_j|^2 \le \|f\|^2 \tag{18.42}$$

Theorem 18.6 Bessel's Inequality. Let $\mathbb{V}$ be an infinite dimensional real-valued function space on an interval $I \subset \mathbb{R}$ and let $\phi_0, \phi_1, \dots$ be a complete orthonormal basis on $\mathbb{V}$. Then for any smooth function f,

$$\sum_{k=0}^{\infty} |\langle f, \phi_k \rangle|^2 \leq \|f\|^2 \tag{18.43}$$

Proof. Let

$$s_n = \sum_{k=0}^{n} \langle f, \phi_k \rangle \phi_k \tag{18.44}$$

Since the ϕ_k are orthonormal,

$$\langle s_n, \phi_i \rangle = \left\langle \sum_{k=0}^{n} \langle f, \phi_k \rangle \phi_k, \phi_i \right\rangle = \sum_{k=0}^{n} \langle f, \phi_k \rangle \langle \phi_k, \phi_i \rangle = \langle f, \phi_i \rangle \tag{18.45}$$

Therefore,

$$\langle f - s_n, \phi_i \rangle = \langle f, \phi_i \rangle - \langle s_n, \phi_i \rangle = \langle f, \phi_i \rangle - \langle f, \phi_i \rangle = 0 \tag{18.46}$$

i.e., $f - s_n$ and ϕ_i are orthogonal. Hence we can apply the Pythagorean theorem,

$$\|f\|^2 = \|f - s_n + s_n\|^2 = \|f - s_n\|^2 + \|s_n\|^2 \tag{18.47}$$

Solving for $\|s_n\|^2$ gives

$$\|s_n\|^2 = \|f\|^2 - \|f - s_n\|^2 \leq \|f\|^2 \tag{18.48}$$

But

$$\begin{aligned}
\|s_n\|^2 &= \langle s_n, s_n\rangle && (18.49)\\
&= \left\langle \sum_{j=0}^{n} \langle f, \phi_j\rangle \phi_j, \sum_{k=0}^{n} \langle f, \phi_k\rangle \phi_k \right\rangle && (18.50)\\
&= \sum_{j=0}^{n} \langle f, \phi_j\rangle \overline{\sum_{k=0}^{n} \langle f, \phi_k\rangle \langle \phi_k, \phi_j\rangle} && (18.51)\\
&= \sum_{j=0}^{n} \langle f, \phi_j\rangle \overline{\langle f, \phi_j\rangle} && (18.52)\\
&= \sum_{j=0}^{n} |\langle f, \phi_j\rangle|^2 && (18.53)
\end{aligned}$$

Substituting the right hand side of 18.53 into the the left hand side of equation 18.48 gives

$$\sum_{j=0}^{n} |\langle f, \phi_j\rangle|^2 = \|s_n\|^2 \le \|f\|^2 \tag{18.54}$$

The sequence of partial sums s_n is a bounded increasing sequence and hence it converges. Taking the limit as $n \to \infty$ gives the desired result. □

The following gives an even stronger result.

Theorem 18.7 Parseval's Theorem Under the same conditions as the previous theorem,

$$\sum_{k=0}^{\infty} |\langle f, \phi_k \rangle|^2 = \|f\|^2 \tag{18.55}$$

Proof. From equation 18.47

$$\|f\|^2 = \|f - s_n + s_n\|^2 = \|f - s_n\|^2 + \|s_n\|^2 \tag{18.56}$$

By the previous theorem $s_n \to f$ hence $\|f - s_n\| \to 0$. The desired result follows by taking the limit as $n \to \infty$. □

The following Best Mean Approximation Theorems tells us that the truncated fourier series is the best approximation to a function among all the possible linear combinations of basis functions.

Theorem 18.8 Best Mean Approximation Theorem. Under the same conditions as the previous theorems, the $n-$term truncation of the Fourier series is the best approximation (in the mean) of all possible expansions of the form $\sum_{k=0}^{n} a_k \phi_k$, in the sense that for any sequence of numbers $a_0, a_1, \dots,$

$$\left\| f - \sum_{k=1}^{n} a_k \phi_k \right\| \geq \left\| f - \sum_{k=1}^{n} \langle f, \phi_k \rangle \phi_k \right\| \tag{18.57}$$

with equality holding only when

$$a_k = \langle f, \phi_k \rangle \text{ for all } k = 0, 1, \dots, n \tag{18.58}$$

Proof. Let f be any function in our vector space and define

$$c_k = \langle f, \phi_k \rangle \tag{18.59}$$

$$s_n = \sum_{k=0}^{n} c_k \phi_k \tag{18.60}$$

$$t_n = \sum_{k=0}^{n} a_k \phi_k \tag{18.61}$$

To verify equation 18.57, we need to show that

$$\|f - t_n\| \geq \|f - s_n\| \tag{18.62}$$

and that equality only holds when $a_k = c_k$ for all k.

By the linearity of the inner product,

$$\|f - t_n\|^2 = \langle f - t_n, f - t_n \rangle \tag{18.63}$$

$$= \langle f, f \rangle + \langle t_n, t_n \rangle - \langle f, t_n \rangle - \langle t_n, f \rangle \tag{18.64}$$

and that

$$\langle t_n, t_n \rangle = \left\langle \sum_{j=0}^{n} a_j \phi_j, \sum_{k=0}^{n} a_k \phi_k \right\rangle \tag{18.65}$$

$$= \sum_{j=0}^{n} a_j \sum_{k=0}^{n} \overline{a_k} \langle \phi_i, \phi_j \rangle \tag{18.66}$$

$$= \sum_{j=0}^{n} a_j \sum_{k=0}^{n} \overline{a_k} \delta_{ij} \tag{18.67}$$

$$= \sum_{j=0}^{n} a_j \overline{a_j} \tag{18.68}$$

$$= \sum_{j=0}^{n} |a_j|^2 \tag{18.69}$$

Furthermore,

$$\langle f, t_n \rangle = \left\langle f, \sum_{k=0}^{n} a_k \phi_k \right\rangle = \sum_{k=0}^{n} \overline{a_k} \langle f, \phi_k \rangle = \sum_{k=0}^{n} \overline{a_k} c_k \tag{18.70}$$

$$\langle t_n, f \rangle = \left\langle \sum_{k=0}^{n} a_k \phi_k, f \right\rangle = \sum_{k=0}^{n} a_k \langle \phi_k, f \rangle = \sum_{k=0}^{n} a_k \overline{c_k} \tag{18.71}$$

Therefore

$$\|f - t_n\|^2 = \|f\|^2 + \sum_{k=0}^{n} |a_k|^2 - \sum_{k=0}^{n} \overline{a_k} c_k - \sum_{k=0}^{n} a_k \overline{c_k} \tag{18.72}$$

Similarly,

$$\|f - s_n\|^2 = \|f\|^2 + \sum_{k=0}^{n} |c_k|^2 - \sum_{k=0}^{n} \overline{c_k} c_k - \sum_{k=0}^{n} c_k \overline{c_k} = \|f\|^2 - \sum_{k=0}^{n} |c_k|^2 \tag{18.73}$$

Hence

$$\sum_{k=0}^{n} |c_k|^2 = \|f\|^2 - \|f - s_n\|^2 \tag{18.74}$$

Rearranging equation 18.72,

$$\begin{aligned}
\|f - t_n\|^2 &= \|f\|^2 + \sum_{k=0}^{n} |a_k|^2 - \sum_{k=0}^{n} \overline{a_k} c_k - \sum_{k=0}^{n} a_k \overline{c_k} && (18.75)\\
&= \|f\|^2 + \sum_{k=0}^{n} |a_k|^2 - \sum_{k=0}^{n} 2\,\mathrm{Re}(a_k c_k) && (18.76)
\end{aligned}$$

But for any complex numbers x and y,

$$\begin{aligned}
|x - y|^2 &= (x - y)\bar{(}x - y) = x\overline{x} - x\overline{y} - y\overline{x} + y\overline{y} && (18.77)\\
&= |x|^2 + |y|^2 - 2\,\mathrm{Re}(xy) && (18.78)
\end{aligned}$$

Hence

$$\begin{aligned}
\|f - t_n\|^2 &= \|f\|^2 + \sum_{k=0}^{n} |a_k|^2 + \sum_{k=0}^{n} \left(|a_k - c_k|^2 - |a_k|^2 - |c_k|^2\right) && (18.79)\\
&= \|f\|^2 + \sum_{k=0}^{n} |a_k - c_k|^2 - \sum_{k=0}^{n} |c_k|^2 && (18.80)\\
&= \|f\|^2 + \sum_{k=0}^{n} |a_k - c_k|^2 - \|f\|^2 + \|f - s_n\|^2 \text{ (from eq. 18.74)} && (18.81)\\
&= \sum_{k=0}^{n} |a_k - c_k|^2 + \|f - s_n\|^2 && (18.82)\\
&\geq \|f - s_n\|^2 && (18.83)
\end{aligned}$$

and equality only holds if each of the terms in the first sum is zero, namely, whence $a_k = c_k$ for all k. □

Topic 19

Triangular Decomposition

Corollary 19.1 Schur's Theorem. Let $\mathbb{V}$ be a complex inner-product space; and let $T \in \mathcal{L}(\mathbb{V})$ be an operator on $\mathbb{V}$. Then there exists an orthonormal basis B of $\mathbb{V}$ such that $\mathcal{M}(T, B)$ is upper-triangular.

Proof. This follows from corollary 17.9. □

Definition 19.2 Let U be a complex valued matrix. Then the **Conjugate Transpose of** U, or **Adjoint matrix**[1] denoted by U^*, is the complex conjugate of the matrix transpose.

$$U^* = \overline{(U^T)} = \overline{U}^T \tag{19.1}$$

There could be some confusion here since we use the * to indicate complex conjugate when it is applied to a vector or scalar, but a conjugate transpose when it is applied to a matrix. We will always me the conjugate transpose when * is applied to a matrix. Some authors use the † to designate a conjugate transpose, e.g. as $U^\dagger$.

Definition 19.3 A matrix U is said to be **unitary** if $U^{-1} = U^*$. If U is real and

[1]Not to be confused with the adjoint map defined in the next section.

unitary then $U^{-1} = U^T$.

Theorem 19.4 Properties of unitary matrices.

1. The rows are orthogonal unit vectors.
2. The columns are orthogonal unit vectors.
3. $\det U = 1$
4. U is invertible.
5. The rows form an orthonormal basis of $\mathbb{C}^n$.
6. The columns form an orthonormal basis of C^n.

Proof. (1) and (2) follow from the fact that $\langle \text{row}_i, \text{column}_j \rangle = \delta_{ij}$ (because $U^{-1} = U^*$) and the fact that $\text{row}_i = \overline{(\text{column}_i)^T}$.

(3) follows because if λ is an eigenvalue with eigenvector v we have hence

$$\begin{aligned} Uv = \lambda v &\implies \overline{Uv} = \overline{\lambda v} && (19.2)\\ &\implies \overline{v^T} U^* = \overline{\lambda}\overline{v}^T && (19.3)\\ &\implies (\overline{v^T} U^*)(Uv) = (\lambda v)(\overline{\lambda}\overline{v}^T) && (19.4)\\ &\implies \overline{v^T} I v = |\lambda|^2 \|v\|^2 && (19.5)\\ &\implies \|v\|^2 = |\lambda|^2 \|v\|^2 && (19.6)\\ &\implies |\lambda|^2 = 1 && (19.7) \end{aligned}$$

since the determinant is the product of the eigenvalues, $\det U = 1$.

(4) follows because $\det U \neq 0$.

(5) and (6) follow from (1) and (2) because the rows (columns) are linearly independent and spanning. □

Theorem 19.5 The Schur Decomposition Let A be any $n \times n$ square matrix over $\mathbb{C}$. Then there exists a unitary matrix U such that

$$A = U^{-1}TU \tag{19.8}$$

where T is an upper triangular matrix.

Proof. For $n = 1$ the result automatically holds since a 1×1 matrix is upper triangular.

Let u_1 be any unit-length eigenvector of A with eigenvalue λ_1. We know that at least one eigenvector exists because $\det(A - \lambda I) = 0$ has at least one complex root.

Define each vector $e_i \in \mathbb{F}^n$ to be all zero except in the ith component, e.g.,

$$\left.\begin{aligned} e_1 &= (1, 0, 0, 0, \dots, 0) \\ e_2 &= (0, 1, 0, 0, \dots, 0) \\ e_3 &= (0, 0, 1, 0, \dots, 0) \\ &\vdots \\ e_n &= (0, 0, \dots, 0, 0, 1) \end{aligned}\right\} \tag{19.9}$$

Then $E = (e_1, \dots, e_n)$ is a basis of $\mathbb{F}^n$.

Since our first eigenvector $u_1 \neq 0$ it has at least one non-zero component. Pick one of these components and designate the r'th component.

Then the list of vectors

$$E' = (u_1, e_1, e_2, \dots, e_{r-1}, e_{r+1}, e_{r+2}, \dots, e_n) \quad \text{(with } e_r \text{ missing)} \tag{19.10}$$

is also a basis of $\mathbb{F}^n$.

Use the Gram-Schmidt process to obtain an orthonormal basis $B = (v_1, \dots, v_n)$ of $\mathbb{F}^n$ from E', starting with $v_1 = u_1$.

Let M be the matrix whose columns are v_i:

$$M = \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \tag{19.11}$$

Then

$$AM = A\begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} \lambda_1 v_1 & Av_2 & \cdots & Av_n \end{pmatrix} \tag{19.12}$$

because $Av_1 = Au_1 = \lambda_1 u_1$. Similary, M^*, the conjugate transpose of M, has v_i^{*T} (the transpose of the complex conjuage of v_i) as row i.

$$M^*AM = \begin{pmatrix} v_1^{*T} \\ v_2^{*T} \\ \vdots \\ v_n^{*T} \end{pmatrix} \begin{pmatrix} \lambda_1 v_1 & Av_2 & \cdots & Av_n \end{pmatrix} \tag{19.13}$$

$$= \begin{pmatrix} v_1^{*T}\lambda_1 v_1 & v_1^{*T}Av_2 & v_1^{*T}Av_3 & \cdots & v_1^{*T}Av_n \\ v_2^{*T}\lambda_1 v_1 & v_2^{*T}Av_2 & v_2^{*T}Av_3 & \cdots & v_2^{*T}Av_n \\ \vdots & \vdots & & & \vdots \\ v_n^{*T}\lambda_1 v_1 & v_n^{*T}Av_2 & v_n^{*T}Av_3 & \cdots & v_n^{*T}Av_n \end{pmatrix} \tag{19.14}$$

$$= \begin{pmatrix} \lambda_1 & v_1^{*T}Av_2 & v_1^{*T}Av_3 & \cdots & v_1^{*T}Av_n \\ 0 & v_2^{*T}Av_2 & v_2^{*T}Av_3 & \cdots & v_2^{*T}Av_n \\ \vdots & \vdots & & & \vdots \\ 0 & v_n^{*T}Av_2 & v_n^{*T}Av_3 & \cdots & v_n^{*T}Av_n \end{pmatrix} \tag{19.15}$$

where the last step follows because the $(v_1, \dots, v_n)$ are orthonormal. Hence M^*AM

has the form

$$M^*AM = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & \boxed{B} & \\ 0 & & & \end{pmatrix} \tag{19.16}$$

where we are again using the "*" notation to refer to values we don't care about; and B is an $(n-1)\times(n-1)$ matrix.

Here is the inductive step: Let W be an $(n-1)\times(n-1)$ unitary matrix for which

$$W^*BW = T_1 \tag{19.17}$$

where T_1 is upper-triangular.

Then define the matrix Y by

$$Y = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boxed{W} & \\ 0 & & & \end{pmatrix} \tag{19.18}$$

Since W is unitary, so is Y. From equation 19.16

$$Y^*(M^*AM)Y = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boxed{W^*} & \\ 0 & & & \end{pmatrix} \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & \boxed{B} & \\ 0 & & & \end{pmatrix} \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boxed{W} & \\ 0 & & & \end{pmatrix} \tag{19.19}$$

Hence from equation 19.17,

$$Y^*(M^*AM)Y = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & W*BW & \\ 0 & & & \end{pmatrix} = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & T_1 & \\ 0 & & & \end{pmatrix} \tag{19.20}$$

where T_1 is upper triangular. Hence

$$Y^*M^*AMY = T \tag{19.21}$$

where T is upper triangular.

Define $U = MY$, which is unitary because the product of unitary matrices is unitary. Since $U^* = (MY)^* = Y^*M^*$, we have

$$U^*MU = T \tag{19.22}$$

where T is upper triangular. □

Remark 19.6 If A is unitary then the Schur decomposition of A gives

$$A = U^{-1}DU \tag{19.23}$$

where D is diagonal.

Topic 20

The Adjoint Map

Definition 20.1 Let $\mathbb{V}$ be vector space over $\mathbb{F}$. Then a **linear functional**[1] on $\mathbb{V}$ is a linear map $\varphi : \mathbb{V} \mapsto \mathbb{F}$.

Theorem 20.2 Let φ be a linear functional on $\mathbb{V}$. Then there is a unique vector $v \in \mathbb{V}$ such that

$$\varphi(u) = \langle u, v \rangle \tag{20.1}$$

for every $u \in \mathbb{V}$.

Proof. Let $B = (e_1, \dots, e_n)$ be an orthonormal basis of $\mathbb{V}$ and pick any $u \in \mathbb{V}$. Then by theorem 17.5

$$u = \langle u, e_1 \rangle e_1 + \cdots + \langle u, e_n \rangle e_n \tag{20.2}$$

hence by linearity of φ

$$\varphi(u) = \langle u, e_1 \rangle \varphi(e_1) + \cdots + \langle u, e_n \rangle \varphi(e_n) \tag{20.3}$$

[1] Recall from definition 8.1 that a linear map is a map that the properties of additivity ($\varphi(u+v) = \varphi(u) + \varphi(v), u, v \in \mathbb{V}$) and homogeneity ($\varphi(av) = a\varphi(v), v \in \mathbb{V}, a \in \mathbb{F}$)

By conjugate homogeneity in the second argument of the inner product (theorem 15.3),

$$\varphi(u) = \langle u, \varphi(e_1)^* e_1\rangle + \cdots + \langle u, \varphi(e_n)^* e_n\rangle \tag{20.4}$$

If we define v by

$$v = \varphi(e_1)^* e_1 + \cdots + \varphi(e_n)^* e_n \tag{20.5}$$

then

$$\langle u, v\rangle = \langle u, \varphi(e_1)^* e_1 + \cdots + \varphi(e_n)^* e_n\rangle \tag{20.6}$$

$$= \langle u, \varphi(e_1)^* e_1\rangle + \cdots + \langle u, \varphi(e_n)^* e_n\rangle \tag{20.7}$$

$$= \varphi(u) \tag{20.8}$$

as desired. Note that v does not depend on u, only on ϕ and the basis.

To prove uniqueness, suppose that there exist v, w such that

$$\varphi(u) = \langle u, v\rangle = \langle u, w\rangle \tag{20.9}$$

for all $u \in \mathbb{V}$. Then

$$0 = \langle u, v\rangle - \langle u, w\rangle = \langle u, v - w\rangle \tag{20.10}$$

This must hold for all u, so if we pick $u = v - w$ then

$$0 = \langle v - w, v - w\rangle \implies v - w = 0 \tag{20.11}$$

hence $v = w$, proving uniqueness. □

Definition 20.3 Let $\mathbb{V}, \mathbb{W}$ be finite dimensional inner product spaces over $\mathbb{V}$ and let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then the **adjoint of** T is the function $T^* : \mathbb{W} \mapsto \mathbb{V}$ defined as the unique vector $v \in \mathbb{V}$ such that, given $w \in \mathbb{W}$,

$$\langle Tv, w\rangle = \langle v, T^* w\rangle \tag{20.12}$$

Note: The use of the "$*$" superscript for both adjoint and complex conjugate can be a bit confusing; however, we should be able to distinguish them from the context. Whenever "*" is applied to a map or operator, in means adjoint; when it is applied to a vector or scalar it means complex conjugate. Thus T^* is the adjoint of T, while $(Tv)*$ is the complex conjugate of Tv.

Theorem 20.4 If $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ then $T^* \in \mathcal{L}(\mathbb{W}, \mathbb{V})$

Proof. Additivity:

$$\langle v, T^*(u+w)\rangle = \langle Tv, u+w\rangle \tag{20.13}$$
$$= \langle Tv, u\rangle + \langle Tv, w\rangle \tag{20.14}$$
$$= \langle v, T^* u\rangle + \langle v, T^*\rangle w \tag{20.15}$$
$$= \langle v, T^* u + T^* w\rangle \tag{20.16}$$
$$\implies T^*(u+w) = T^* u + T^* w \tag{20.17}$$

Homogeneity:

$$\langle v, T^*(aw)\rangle = \langle Tv, aw\rangle \tag{20.18}$$
$$= a^*\langle Tv, w\rangle \tag{20.19}$$
$$= a^*\langle v, T^* w\rangle \tag{20.20}$$
$$= \langle v, aT^* w\rangle \tag{20.21}$$

$$\implies T^*aw = aT^*w \tag{20.22}$$

□

Theorem 20.5 Properties of the Adjoint Map.

1. Additivity: $(S+T)^* = S^* + T^*$
2. Conjugate Homogeneity: $(aT)^* = a^*T^*$
3. Adjoint of Adjoint: $(T^*)^* = T$
4. Identity: $I^* = I$
5. Adjoint Product: $(ST)^* = T^*S^*$

Proof. (Exercise.) □

Theorem 20.6 Let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$. Then

1. $\text{null}(T^*) = (\text{range}(T))^\perp$
2. $\text{range}(T^*) = (\text{null}(T))^\perp$
3. $\text{null}(T) = (\text{range}(T^*))^\perp$
4. $\text{range}(T) = (\text{null}(T^*))^\perp$

Proof. (1) Let $w \in \mathbb{W}$, and suppose $w \in \text{null}(T^*)$

$$T^*w = 0 \iff \langle v, T^*w\rangle = 0 \quad (\forall v \in \mathbb{V}) \tag{20.23}$$

$$\iff \langle Tv, w\rangle = 0 \quad (\forall v \in \mathbb{V}) \tag{20.24}$$

$$\iff w \in (\text{range } T)^\perp \tag{20.25}$$

(2) is the orthogonal complement of (3), so we prove (3) first.

(3) Let $w \in \mathbb{W}$, and suppose that $w \in \text{null } T$. The replace T^* with T everywhere in

equations 20.23 through 20.25:

$$Tw = 0 \iff \langle v, Tw \rangle = 0 \quad (\forall v \in \mathbb{V}) \tag{20.26}$$

$$\iff \langle T^* v, w \rangle = 0 \quad (\forall v \in \mathbb{V}) \tag{20.27}$$

$$\iff w \in (\text{range } T^*)^{\perp} \tag{20.28}$$

Returning to (2), we take the orthognoal complement of (3)

$$\text{null}(T) = (\text{range}(T^*))^{\perp} \iff (\text{null } T)^{\perp} = ((\text{range}(T^*))^{\perp})^{\perp} = (\text{range}(T^*) \tag{20.29}$$

To prove (4), take the orthogonal complement of (1):

$$\text{null}(T^*) = (\text{range } T)^{\perp} \iff (\text{null}(T^*))^{\perp} = ((\text{range } T)^{\perp})^{\perp} = \text{range } T \tag{20.30}$$

□

Next we recall the definition of the adjoint matrix as the conjugate transpose (see definition 19.2). Then the adjoint operator and the adjoint matrix are related in the following way:

Theorem 20.7 Let $T \in \mathcal{L}(\mathbb{V}, \mathbb{W})$, and let $E = (e_1, \dots, e_n)$ and $F = (f_1, \dots, f_m)$ be orthonromal bases of $\mathbb{V}$ and $\mathbb{W}$ respectively. Then the matrix of the adjoint of T is the adjoint (conjugate transpose) of the matrix of T:

$$\mathcal{M}(T^*, F, E) = \mathcal{M}(T, E, F)^* \tag{20.31}$$

Proof. (Exercise.) □

Definition 20.8 Let $T \in \mathcal{L}V$ be an operator. Then we say T is **self-adjoint** or **Hermitian** if $T = T^*$.

Theorem 20.9 Let $\mathbb{V}$ be a finite-dimensional, nonzero, inner-product space over $\mathbb{F}$ (either $\mathbb{R}$ or $\mathbb{C}$) and let $T \in \mathcal{L}(V)$ be a self-adjoint operator over $\mathbb{V}$. The the eigenvalues of T are all real.

Proof. Let λ be an eigenvalue of $\mathbb{V}$ with nonzero eigenvector v. Then

$$\begin{aligned}
\lambda\|v\|^2 &= \lambda\langle v, v\rangle &(20.32)\\
&= \langle \lambda v, v\rangle = \langle Tv, v\rangle \text{ (def. of eigenvalue)} &(20.33)\\
&= \langle v, T^*v\rangle = \langle v, Tv\rangle \text{ (T is self-adjoint)} &(20.34)\\
&= \langle v, \lambda v\rangle = \overline{\lambda}\langle v, v\rangle = \overline{\lambda}\|v\|^2 &(20.35)
\end{aligned}$$

Since $v \neq 0$ then $\|v\|^2 \neq 0$ and can be cancelled to give $\lambda = \overline{\lambda}$, which is only possible of $\lambda \in \mathbb{R}$. □

Theorem 20.10 Let $\mathbb{V}$ be a complex inner product space and let $T \in \mathcal{L}(\mathbb{V})$ such that

$$\langle Tv, v\rangle = 0 \tag{20.36}$$

for all $v \in \mathbb{V}$. Then $T = 0$.

Proof. Since $\langle Tv, v\rangle = 0$ for all v, it is true for $v = u + w$ and $v = u - w$:

$$0 = \langle T(u + w), u + w\rangle = \langle Tu, u\rangle + \langle Tu, w\rangle + \langle Tw, u\rangle + \langle Tw, w\rangle \tag{20.37}$$

$$0 = \langle T(u - w), u - w\rangle = \langle Tu, u\rangle - \langle Tu, w\rangle - \langle Tw, u\rangle + \langle Tw, w\rangle \tag{20.38}$$

Subtracting gives

$$0 = 2\langle Tu, w\rangle + 2\langle Tw, u\rangle \implies \langle Tu, w\rangle = -\langle Tw, u\rangle \tag{20.39}$$

Next we apply the same idea to $v = u + iw$ and $v = u - iw$:

$$0 = \langle T(u + iw), u + iw\rangle = \langle Tu, u\rangle + \langle Tu, iw\rangle + \langle iTw, u\rangle + \langle iTw, iw\rangle \tag{20.40}$$

$$0 = \langle T(u - iw), u - iw\rangle = \langle Tu, u\rangle - \langle Tu, iw\rangle - \langle iTw, u\rangle + \langle iTw, iw\rangle \tag{20.41}$$

Subtracting,

$$0 = -2i\langle Tu, w\rangle + 2i\langle Tw, u\rangle \tag{20.42}$$

$$\implies 0 = 2\langle Tu, w\rangle - 2\langle Tw, u\rangle \text{ (multiply by } i) \tag{20.43}$$

$$\implies \langle Tu, w\rangle = \langle Tw, u\rangle \tag{20.44}$$

Adding equations 20.39 and 20.44 gives

$$\langle Tu, w\rangle = 0 \tag{20.45}$$

Since this must hold for all u, w, it certainly holds for $w = Tu$. Then for any $u \in \mathbb{V}$,

$$0 = \langle Tu, w\rangle = \langle w, w\rangle \tag{20.46}$$

which is true if and only if $w = 0$. Hence for any $u \in \mathbb{V}$, $Tu = 0$. Hence $T = 0$. □

Remark 20.11 The last result only holds if $\mathbb{V}$ is complex; if $\mathbb{V}$ is real equations 20.40 and following do not hold. Consequently equation 20.39 does not imply that $T = 0$

Example 20.1 (Example of Remark 20.11). Let $\mathbb{V}$ be a real vector space and let T be the 90 rotation about the origin operator,e.g, if $v = (x, y) \in \mathbb{V}$,

$$Tv = T(x, y) = (-y, x) \tag{20.47}$$

Let $w = Tv$. Then

$$\langle Tv, w\rangle = \langle w, w\rangle = \|w\|^2 \tag{20.48}$$

Furthermore, if $v = (x, y)$,

$$Tw = TTv = T(-y, x) = (-x, -y) \tag{20.49}$$

Hence

$$\langle Tw, v\rangle = \langle(-x, -y), (x, y)\rangle = -\langle(x, y), (x, y)\rangle = -\langle v, v\rangle = -\|v\|^2 \tag{20.50}$$

Equation 20.39 tells us that $\langle Tu, w\rangle = -\langle Tw, u\rangle$; from equations 20.48 and 20.50 we get

$$\|v\|^2 = \|w\|^2 \implies \|Tv\| = \|v\| \tag{20.51}$$

In other words, the rotation operator is length-preserving, but not the zero operator.

The following theorem is also only true on complex vectors, but is false on real vector spaces.

Theorem 20.12 Let $\mathbb{V}$ be a complex inner product space and let $T \in \mathcal{L}(\mathbb{V})$ be an operator on $\mathbb{V}$. Then T is self-adjoint if and only if $\langle Tv, v\rangle \in \mathbb{R}$ for every $v \in \mathbb{V}$.

Proof. Let $v \in \mathbb{V}$. Then since any real number is its own complex conjugate, if $\langle Tv, v\rangle = 0$ then

$$0 = \langle Tv, v\rangle - \overline{\langle Tv, v\rangle} = \langle Tv, v\rangle - \langle v, Tv\rangle \tag{20.52}$$

$$= \langle Tv, v\rangle - \langle T^*v, v\rangle \tag{20.53}$$

$$= \langle Tv - T^*v, v\rangle \tag{20.54}$$

$$= \langle (T - T^*)v, v\rangle \tag{20.55}$$

Hence $T - T^* = 0$ by theorem 20.10 and thus $T = T^*$.

To prove the converse, suppose that T is self adjoint. Then equation 20.55 follows immediately, which (following the steps backwards), tells us that $\langle Tv, v\rangle$ is its own complex conjugate, hence it must be real. □

Theorem 20.13 Let T be a self adjoint operator on a nonzero inner product space $\mathbb{V}$ such that

$$\langle Tv, v\rangle = 0 \ \forall v \in \mathbb{V} \tag{20.56}$$

Then $T = 0$.

Proof. If $\mathbb{V}$ is complex this follows from theorem 20.10. So let us assume that $\mathbb{V}$ is real.

Since $\langle Tv, v\rangle = 0$, we have

$$0 = \langle T(u+w), u+w\rangle \tag{20.57}$$

$$= \langle Tu, u\rangle + \langle Tu, w\rangle + \langle Tw, u\rangle + \langle Tw, w\rangle \tag{20.58}$$

$$= \langle Tu, w\rangle + \langle Tw, u\rangle \tag{20.59}$$

Hence

$$-\langle Tu, w\rangle = \langle Tw, u\rangle = \langle w, T^*u\rangle \tag{20.60}$$

$$= \langle w, Tu\rangle \text{ (because } T \text{ is self adjoint)} \tag{20.61}$$

$$= \langle Tu, w\rangle \text{ (because } \mathbb{V} \text{ is real)} \tag{20.62}$$

Thus $\langle Tu, w\rangle = 0$ for all $u, w \in \mathbb{V}$. Let $w = Tu$. Then $\langle Tu, Tu\rangle = 0$ for all $u \in \mathbb{V}$. Hence $\|Tu\|^2 = 0 \implies Tu = 0$ for all u. Hence $T = 0$. □

Definition 20.14 Let $T \in \mathcal{L}\mathbb{V}$ be an operator on $\mathbb{V}$. Then we say T is **normal** or a **normal operator** if it commutes with its adjoint, i.e.,

$$T^*T = TT^* \tag{20.63}$$

Corollary 20.15 If T is self-adjoint, then it is normal.

Theorem 20.16 Let $T \in \mathcal{L}(\mathbb{V})$ be an operator over $\mathbb{V}$. Then T is normal if and only if

$$\|Tv\| = \|T^*v\| \quad \forall v \in \mathbb{V} \tag{20.64}$$

Proof.

$$T \text{ is normal} \iff TT^* - T^*T = 0 \tag{20.65}$$

$$\iff \langle (TT^* - T^*T)v, v\rangle = 0 \ \forall v \in \mathbb{V} \tag{20.66}$$

$$\iff \langle TT^*v, v\rangle = \langle T^*Tv, v\rangle \tag{20.67}$$

$$\iff \langle T^*v, T^*v\rangle = \langle Tv, Tv\rangle \tag{20.68}$$

$$\iff \|T^*v\|^2 = \|Tv\|^2$$

□

Corollary 20.17 Let $T \in \mathcal{L}(\mathbb{V})$ be normal. Then if $v \in \mathbb{V}$ is an eigenvector of T with eigenvalue λ then v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$.

Proof. Let v be an eigenvector of T with eigenvalue λ. Then

$$(T - \lambda I)v = 0 \tag{20.69}$$

Let $H = T - \lambda I$. Since T is normal,

$$HH^* = (T - \lambda I)(T - \lambda I)^* \tag{20.70}$$

$$= (T - \lambda I)(T^* - \overline{\lambda} I) \tag{20.71}$$

$$= TT^* - \overline{\lambda}T - \lambda T^* + |\lambda|^2 I \tag{20.72}$$

$$= T^*T - \lambda T^* - \overline{\lambda}T + \overline{\lambda}\lambda I \tag{20.73}$$

$$= T^*(T - \lambda I) - \overline{\lambda}(T - \lambda I) \tag{20.74}$$

$$= (T^* - \lambda I)(T - \lambda I) = H^*H \tag{20.75}$$

Hence H is normal. By theorem 20.16

$$\|Hv\| = \|H^*v\| \tag{20.76}$$

Since $Hv = 0$ (because $Tv = \lambda v$),

$$0 = \|H^*v\| = \|(T - \lambda I)^* v\| = \|(T^* - \overline{\lambda} I)v\| \tag{20.77}$$

Hence $T^*v = \overline{\lambda}v$, i.e., v is an eigenvector of T^* with eigenvalue $\overline{\lambda}$. □

Theorem 20.18 Let $T \in \mathcal{L}(\mathbb{V})$ be normal. Then the eigenvectors of T with distinct eigenvalues are orthogonal.

Proof. Let $\alpha \neq \beta$ be eigenvalues of T with eigenvectors u and v. By theorem 20.17

$$Tv = \beta v \implies T^*v = \overline{\beta} v \tag{20.78}$$

Hence

$$(\alpha - \beta)\langle u, v\rangle = \alpha\langle u, v\rangle - \beta\langle u, v\rangle \tag{20.79}$$
$$= \langle \alpha u, v\rangle - \langle u, \overline{\beta} v\rangle \tag{20.80}$$
$$= \langle Tu, v\rangle - \langle u, T^*v\rangle \tag{20.81}$$
$$= \langle Tu, v\rangle - \langle Tu, v\rangle \tag{20.82}$$
$$= 0 \tag{20.83}$$

Since $\alpha \neq \beta$, this means $\langle u, v\rangle = 0$. □

Topic 21

The Spectral Theorem

Theorem 21.1 Spectral Theorem, Complex Vector Spaces. Let $\mathbb{V}$ be a complex inner-product space and $T \in \mathcal{L}(\mathbb{V})$. Then $\mathbb{V}$ has an orthonormal basis consisting of eigenvectors of T if and only if T is normal.

Proof. ($\Longrightarrow$) Suppose that $\mathbb{V}$ has an orthonormal basis consisting of eigenvectors of T. With respect to this basis, T has a diagonal matrix (this is theorem 12.5). Since the matrix of the adjoint operator is the adjoint matrix of the matrix of the operator, T^* also has a diagonal matrix. Since any two diagonal matrices commuute, $TT^* = T^*T$. Hence T is normal.

($\Longleftarrow$) Suppose T is normal. Then by corollary 17.10, there is some orthonormal basis $E = (e_1, \dots, e_n)$ under which $\mathcal{M}(T)$ is upper triangular. Let

$$\mathcal{M}(T, E) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ 0 & & a_{nn} \end{pmatrix} \tag{21.1}$$

where by definition of the matrix of an operator

$$Te_i = a_{1i}e_1 + \cdots + a_{ni}e_n \tag{21.2}$$

Hence

$$Te_1 = a_{11}e_1 + \cdots + a_{n1}e_n = a_{11}e_1 \tag{21.3}$$

so that

$$\|Te_1\|^2 = |a_{11}|^2 \tag{21.4}$$

Similarly,

$$\mathcal{M}(T^*, E) = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \ddots & \vdots \\ 0 & & a_{nn} \end{pmatrix}^* = \begin{pmatrix} \overline{a_{11}} & & 0 \\ \vdots & \ddots & \\ \overline{a_{1n}} & \cdots & \overline{a_{nn}} \end{pmatrix} \tag{21.5}$$

Again looking at the first column,

$$T^*e_1 = \overline{a_{11}}e_1 + \cdots + \overline{a_{1n}}e_n = \sum_{i=1}^n \overline{a_{1i}}e_i \tag{21.6}$$

$$\|T^*e_1\|^2 = \langle \sum_{i=1}^n \overline{a_{1i}}e_i, \sum_{j=1}^n \overline{a_{1j}}e_j \rangle = \sum_{i=1}^n \sum_{j=1}^n \langle \overline{a_{1i}}e_i, \overline{a_{1j}}e_j \rangle \tag{21.7}$$

$$= \sum_{i=1}^n \sum_{j=1}^n \overline{a_{1i}}\overline{\overline{a_{1j}}}\langle e_i, e_j \rangle = \sum_{i=1}^n \sum_{j=1}^n \overline{a_{1i}}\overline{\overline{a_{1j}}}\delta_{ij} = |a_{11}|^2 + \cdots + |a_{1n}|^2 \tag{21.8}$$

Since T is normal, by theorem 20.16, we require

$$\|Te_1\| = \|T^*e_1\| \tag{21.9}$$

$$|a_{11}|^2 = |a_{11}|^2 + \cdots + |a_{1n}|^2 \tag{21.10}$$

Hence the only nonzero element in the first row is the one in the first column:

$$a_{12} = a_{13} = \cdots = a_{1n} = 0 \tag{21.11}$$

Therefore $\mathcal{M}(T)$ has the form

$$\mathcal{M}(T) = \begin{pmatrix} a_1 1 & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{pmatrix} \tag{21.12}$$

Repeating the same argument with e_2 we find that the only nonzero element in the second row is a_{22}; and then repeat for the remaining rows, to show that the matrix is diagonal.

Since the matrix is diagonal, E is an orthonormal basis of eigenvalues (theorem 12.5).

□

Theorem 21.2 Real Spectral Theorem. Let $\mathbb{V}$ be a real inner-product space and let $T \in \mathcal{L}(\mathbb{V})$. Then $\mathbb{V}$ has an orthonormal basis consisting of eigenvectors of T if and only if T is self-adjoint.

Corollary 21.3 Let $T \in \mathcal{L}(\mathbb{V})$ be self-adjoint with distinct eigenvalues $\lambda_1, \dots, \lambda_m$. Then

$$\mathbb{V} = \text{null}(\mathbb{V} - \lambda_1 I) \oplus \cdots \oplus \text{null}(\mathbb{V} - \lambda_m I) \tag{21.13}$$

where each subspace $\mathbb{U}_i = \text{null}(\mathbb{V} - \lambda_i I)$ is orthogonal to each of the other $\mathbb{U}_i$.

Lemma 21.4 Let $T \in \mathcal{L}(V)$ be self-adjoint. If $\alpha, \beta \in \mathbb{R}$ satisfy

$$\alpha^2 < 4\beta \tag{21.14}$$

then

$$T^2 + \alpha T + \beta I \tag{21.15}$$

is invertible.

Proof. Let $v \in \mathbb{V}$ be nonzero. Then

$$\begin{aligned}\langle (T^2 + \alpha T + \beta I)v, v\rangle &= \langle T^2 v, v\rangle + \alpha\langle Tv, v\rangle + \beta\langle v, v\rangle &(21.16)\\ &= \langle Tv, T^* v\rangle + \alpha\langle Tv, v\rangle + \beta\|v\|^2 &(21.17)\\ &= \langle Tv, Tv\rangle + \alpha\langle Tv, v\rangle + \beta\|v\|^2 &(21.18)\\ &= \|Tv\|^2 + \alpha\langle Tv, v\rangle + \beta\|v\|^2 &(21.19)\end{aligned}$$

From the Cauchy Schwarz inequality (theorem 15.11)

$$|\langle Tv, v\rangle| \leq \|Tv\|\|v\| \tag{21.20}$$

$$\implies \alpha\langle Tv, v\rangle \geq -|\alpha||\langle Tv, v\rangle| \geq -|\alpha|\|Tv\|\|v\| \tag{21.21}$$

hence

$$\begin{aligned}\langle (T^2 + \alpha T + \beta I)v, v\rangle &= \|Tv\|^2 - |\alpha|\|Tv\|\|v\| + \beta\|v\|^2 &(21.22)\\ &= \|Tv\|^2 - |\alpha|\|Tv\|\|v\| + \frac{|\alpha|^2}{4}\|v\|^2\\ &\quad - \frac{|\alpha|^2}{4}\|v\|^2 + \beta\|v\|^2 &(21.23)\\ &= \left(\|Tv\| - \frac{|\alpha|\|v\||}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right)\|v\|^2 > 0 &(21.24)\end{aligned}$$

where the last inequality follows because $v \neq 0$ and $\alpha^2 < 4\beta$. Hence the inner product on the left is non-zero. Hence

$$(T^2 + \alpha T + \beta I)v \neq 0 \tag{21.25}$$

so that $T^2 + \alpha T + \beta I$ is injective (theorem 8.17). Hence it is invertible. □

Lemma 21.5 If $T \in \mathcal{L}(\mathbb{V})$ is self-adjoint in a real vector space $\mathbb{V}$, then T has an eigenvalue.

Proof. Let $n = \dim(\mathbb{V})$ and pick any $v \neq 0, v \in \mathbb{V}$. Let

$$S = (v, Tv, T^2v, \ldots, T^nv) \tag{21.26}$$

Since length$(S) > n$, S cannot be linearly independent. Thus there exists real numbers, not all zero, such that

$$0 = a_0v + a_1Tv + a_2T^2v + \cdots + a_nT^nv \tag{21.27}$$

Define the polynomial $p(x)$ by

$$p(x) = a_0 + a_1x + \cdots + a_nx^n \tag{21.28}$$

Then $p(x)$ can be factored

$$p(x) = c(x^2 + \alpha_1x + \beta_1)\cdots(x^2 + \alpha_px + \beta_p)(x - \lambda_1)\cdots(x - \lambda_m) \tag{21.29}$$

where $c \neq 0$, $\alpha_j, \beta_j, \lambda_j \in \mathbb{R}$, each $\alpha_j^2 < 4\beta_j$, and $m + p \geq 1$ (There are m real roots and $p/2$ complex conjugate pairs of roots). Hence

$$0 = c(T^2 + \alpha_1 T + \beta_1 I) \cdots (T^2 + \alpha_p T + \beta_p I)(T - \lambda_1 I) \cdots (T - \lambda_m I)v \qquad (21.30)$$

Since each T is self-adjoint and $\alpha_j^2 < 4\beta_j$, the previous lemma tells us that each $T^2 + \alpha_j T + \beta_j I$ is invertible; hence (also $c \neq 0$),

$$0 = (T - \lambda_1 I) \cdots (T - \lambda_m I)v \qquad (21.31)$$

Thus for at least one j, $T - \lambda_j I$ is not injective, and λ_j is an eigenvalue of T. □

Proof. (Proof of theorem 21.2.) ($\Longrightarrow$) Suppose $\mathbb{V}$ has an orthonormal basis consisting of eigenvectors of T.

Then with respect to this basis, T has a diagonal matrix (theorem 12.5).

The eigenvalues are all real. This is because the vector space is real. To see this, suppose there was a complex eigenvalue λ with nonzero imaginary part, with eigenvector v. Then since $Tv = \lambda v \in \mathbb{V}$ we have $\lambda v \in \mathbb{V}$. Since $v \in \mathbb{V}$, then v must be real. Hence λv is complex, which means it cannot be a part of a real vector space; since this would violate closure we can't have a non-real eigenvalue.

Thus $\mathcal{M}T$ is both diagonal and real. Hence $T = T^*$, making T self-adjoint.

($\Longleftarrow$) Suppose that T is self-adjoint. Prove by induction on the dimension of $\mathbb{V}$.

If $\dim \mathbb{V} = 1$ then $\mathbb{V}$ has an orthonormal basis consisting of eigenvectors (all vectors are parallel so any vector is an eigenvector).

Inductive Hypothesis: suppose $\dim \mathbb{V} = n > 1$, and assume that any vector space with dimension $< n$ has an orthonormal basis consisting solely of eigenvectors of T.

Let λ, u be an eigenvalue,eigenvector pair with $\|u\| = 1$.

Let $\mathbb{U}$ be the subspace of $\mathbb{V}$ spanned by u.

Let $v \in \mathbb{U}^\perp$; then $\langle u, v\rangle = 0$. Since T is self adjoint,

$$\langle u, Tv\rangle = \langle Tu, v\rangle = \langle \lambda u, v\rangle = \lambda\langle u, v\rangle = 0 \tag{21.32}$$

Hence whenver $v \in \mathbb{U}^\perp$ we have $Tv \in \mathbb{U}^\perp$, i.e., $U^\perp$ is invariant under T.

Let $S \in \mathcal{L}(\mathbb{U}^\perp)$ be defined as the restriction of T to $U^\perp$, i.e., $S = T|_{\mathbb{U}^\perp}$.

Let $v, w \in \mathbb{U}^\perp$. Then

$$\langle Sv, w\rangle = \langle Tv, w\rangle = \langle v, Tw\rangle = \langle v, Sw\rangle \tag{21.33}$$

Hence S is self-adjoint.

By the inductive hypothesis there is an orthonromal basis of $\mathbb{U}^\perp$ consisting solely of eigenvectors of S. Joining u to this list of eigenvectors of S (each of which is also an eigenvector of T) gives a basis of $\mathbb{V}$ that consists solely of eigenvectors of T. □

Theorem 21.6 Spectral Theorem for Matrices Let A be matrix over $\mathbb{F}$. Then A is normal if and only if it is diagonalizable, and there exists unitary matrix U such that

$$A = UDU^* \tag{21.34}$$

where D is a diagonal matrix consisting of the eigenvalues of A.

Lemma 21.7 Let T be a normal triangular matrix. Then T is a diagonal matrix.

Proof. Let T be a triangular matrix with $T_{ij} = 0$ for $i > j$. Since T is normal,

$TT^* = T^*T$, and in particular, the diagonal elements are equal:

$$(TT^*)_{ii} = (T^*T)_{ii} \tag{21.35}$$

Then

$$(T^*T)_{ii} = \sum_{k=1}^{n} (T^*)_{ik} T_{ki} = \sum_{k=1}^{i} \overline{T_{ki}} T_{ki} \tag{21.36}$$

$$(TT^*)_{ii} = \sum_{k=1}^{n} T_{ik} (T^*)_{ki} = \sum_{k=i}^{n} T_{ik} \overline{T_{ik}} \tag{21.37}$$

Then equating the diagonal components,

$$\overline{T_{1i}} T_{1i} + \cdots + \overline{T_{ii}} T_{ii} = T_{ii} \overline{T_{ii}} + \cdots + T_{in} \overline{T_{in}} \tag{21.38}$$

Hence $T_{ij} = 0$ if $i \neq j$. □

Proof. (Proof of theorem 21.6) Suppose T is normal. By the Schur Decomposition (theorem 19.5),

$$A = UTU^* \tag{21.39}$$

where T is upper triangular. Hence

$$A^*A = (UTU^*)^* UTU^* = UT^*U^*UTU^* = UT^*TU^* \tag{21.40}$$

$$AA^* = UTU^*UT^*U^* = UTT^*U^* \tag{21.41}$$

Since A is normal, $AA^* = A^*A$ and therefore

$$UT^*TU^* = UTT^*U^* \tag{21.42}$$

$$U^*(UT^*TU^*)U = U^*(UTT^*U^*)U \tag{21.43}$$

$$T^*T = TT^* \tag{21.44}$$

Hence T is also normal. By the lemma, Since T is triangular and normal it is diagonal.

□

Topic 22

Normal Operators

Recall that an operator is self-adjoint if $T = T^*$ and is normal if $TT^* = T^*T$. In this section we will be concerned with real inner product spaces only, hence

$$\mathcal{M}(T^*) = (\mathcal{M}(T))^T \tag{22.1}$$

Theorem 22.1 Let $\mathbb{V}$ be a two-dimensional real inner product space and suppose that $T \in \mathcal{L}(\mathbb{V})$ is an operator over $\mathbb{V}$. Then the following are equivalent:

(1) T is normal but not self-adjoint.

(2) For every othonormal basis B of $\mathbb{V}$

$$\mathcal{M}(T, B) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{22.2}$$

where $b \neq 0$.

(3) For any orthonormal basis B of $\mathbb{V}$, equation 22.2 holds with $b > 0$.

Proof. We will show that (1) $\implies$ (2) $\implies$ (3) $\implies$ (1) which is equivalent to proving that any one of the statements implies any one of the others.

((1) $\implies$ (2)) Here we begin by assuming (1), i.e., $T^*T = TT^*$ but $T^* \neq T$.

Since $\mathbb{V}$ is two-dimensional than $\mathcal{M}(T)$ is a 2×2 matrix.

Let $B = (e_1, e_2)$ be any orthonormal basis of $\mathbb{V}$, and let

$$\mathcal{M}(T) = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \tag{22.3}$$

To prove (2) we must show that $a = d$, $b = -c \neq 0$. By the definition of $\mathcal{M}(T, B)$, (equation 12.1)

$$Te_1 = ae_1 + be_2 \tag{22.4}$$

$$Te_2 = ce_1 + de_2 \tag{22.5}$$

Since B is orthonormal, $\langle e_i, e_j\rangle = \delta_{ij}$; therefore

$$\|Te_1\|^2 = \langle Te_1, Te_1\rangle = \langle ae_1 + be_2, ae_1 + be_2\rangle = a^2 + b^2 \tag{22.6}$$

$$\|Te_2\|^2 = \langle T_e2, Te_2\rangle = \langle ce_1 + de_2, ce_1 + de_2\rangle = c^2 + d^2 \tag{22.7}$$

Since the matrix of the adjoint is the adjoint of the matrix,

$$\mathcal{M}(T^*) = \mathcal{M}(T)^* = \begin{pmatrix} a^* & b^* \\ c^* & d^* \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{22.8}$$

where the last step follows because the vector space is real. we have (again by the definition of $\mathcal{M}(T^*)$, equation 12.1),

$$T^*e_1 = ae_1 + ce_2 \tag{22.9}$$

$$T^*e_2 = be_1 + de_2 \tag{22.10}$$

so that

$$\|T^*e_1\| = \langle ae_1 + ce_2, ae_1 + ce_2\rangle = a^2 + c^2 \tag{22.11}$$

$$\|T^*e_2\| = \langle be_1 + de_2, be_1 + de_2\rangle = b^2 + d^2 \tag{22.12}$$

By theorem 20.16, since T is normal,

$$\|Te_i\| = \|T^*e_i\| \implies \begin{cases} a^2 + b^2 = a^2 + c^2 \\ c^2 + d^2 = b^2 + d^2 \end{cases} \tag{22.13}$$

This tells us that $b^2 = c^2$ or $b = \pm c$.

We are also given that T is not self-adjoint, from which we conclude that

$$\mathcal{M}(T) \neq \mathcal{M}(T)^* \implies \begin{pmatrix} a & c \\ b & d \end{pmatrix} \neq \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{22.14}$$

so that $b \neq c$. Therefore $b = -c$, and

$$\mathcal{M}(T) = \begin{pmatrix} a & -b \\ b & d \end{pmatrix} \tag{22.15}$$

Furthermore, $b \neq 0$ because otherwise the matrix would be diagonal and hence self-adjoint, and we are given that the operator is not self adjoint, hence neither is the matrix. Since the operator is normal, $TT^* = T^*T$ hence

$$\mathcal{M}(T)\mathcal{M}(T^*) = \mathcal{M}(T^*)\mathcal{M}(T) \tag{22.16}$$

But

$$\mathcal{M}(T)\mathcal{M}(T^*) = \begin{pmatrix} a & -b \\ b & d \end{pmatrix} \begin{pmatrix} a & b \\ -b & d \end{pmatrix} = \begin{pmatrix} a^2+b^2 & ab-bd \\ ab-bd & b^2+d^2 \end{pmatrix} \tag{22.17}$$

$$\mathcal{M}(T^*)\mathcal{M}(T) = \begin{pmatrix} a & b \\ -b & d \end{pmatrix} \begin{pmatrix} a & -b \\ b & d \end{pmatrix} = \begin{pmatrix} a^2+b^2 & -ab+bd \\ -ab+bd & b^2+d^2 \end{pmatrix} \tag{22.18}$$

Substituting into equation 22.16 and equating like components of the matrix,

$$-ab + bd = ab - bd \implies ab = bd \tag{22.19}$$

Since $b \neq 0$, we have $a = d$, and thus

$$\mathcal{M}(T, B) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{22.20}$$

((2) $\implies$ (3)) Here was assume that (2) is true, hence for every orthonormal basis $B = (e_1, e_2)$, equation 22.2 holds with $b \neq 0$:

$$\mathcal{M}(T, B) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{22.21}$$

Pick any orthonormal basis B. Then if $b > 0$, then statement (3) follows immediately. If $b < 0$ then consider the basis $E = (-e_1, -e_2)$. By definition of $\mathcal{M}(T, B)$ (equation

12.1), we have

$$\begin{cases} Te_1 = ae_1 + be_2 \\ Te_2 = -be_1 + ae_2 \end{cases} \tag{22.22}$$

Thus

$$T(-e_1) = -Te_1 = -ae_1 - be_2 \tag{22.23}$$

$$T(-e_2) = -Te_2 = be_1 - ae_2 \tag{22.24}$$

so that by the definition of $\mathcal{M}(T, E)$,

$$\mathcal{M}(T, E) = \begin{pmatrix} -a & -b \\ b & -a \end{pmatrix} \tag{22.25}$$

Since $b < 0$, $-b > 0$ and therefore (3) holds for basis E.

((3) $\implies$ (1)) We begin by assuming that (3) is true thus equation 22.2 holds for some $b > 0$.

$$\mathcal{M}(T, B) = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{22.26}$$

Since $b > 0$, $b \neq -b$ and therefore $\mathcal{M}(T) \neq (\mathcal{M}(T))^*$. So T is not self-adjoint. To show that T is normal we must verify $TT^* = T^*T$, which we do with matrix multiplication.

$$\mathcal{M}(T)(\mathcal{M}(T))^* = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} a & b \\ -b & a \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & 0 \\ 0 & a^2 + b^2 \end{pmatrix} \tag{22.27}$$

$$(\mathcal{M}(T))^*\mathcal{M}(T) = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} a & -b \\ b & a \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & 0 \\ 0 & a^2 + b^2 \end{pmatrix} \tag{22.28}$$

$$\implies \mathcal{M}(T)(\mathcal{M}(T))^* = (\mathcal{M}(T))^*\mathcal{M}(T) \tag{22.29}$$

Therefore $TT^* = T^*T$ □

Notation: We will sometimes divide a matrix into blocks and refer to each of the blocks as a matrix in its own right. For example, we might write

$$\begin{pmatrix} a & b & c & d & e \\ f & g & h & i & j \\ k & l & m & n & o \\ p & q & r & s & t \\ u & v & w & x & y \end{pmatrix} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{22.30}$$

where

$$A = \begin{pmatrix} a & b \\ f & g \end{pmatrix}, \quad B = \begin{pmatrix} c & d & e \\ h & i & j \end{pmatrix}, \quad C = \begin{pmatrix} k & l \\ p & q \\ u & v \end{pmatrix}, \quad D = \begin{pmatrix} m & n & o \\ r & s & t \\ w & x & y \end{pmatrix} \tag{22.31}$$

The matrices A, B, C, D are called **sub-matrices** of the original matrix.

Multiplication of compatibly blocked matrices proceeds like normal matrix multplication. Thus

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} P & Q \\ R & S \end{pmatrix} = \begin{pmatrix} AP + BR & AQ + BS \\ CP + DR & CQ + DS \end{pmatrix} \tag{22.32}$$

so long as all the multiplicatons are well-defined (i.e., it is possible to multiply A and P, etc.).

A matrix square M is called **Block (Upper) Triangular** if it can be written in the form

$$M = \begin{pmatrix} * & * & \dots & * \\ 0 & * & & * \\ \vdots & \ddots & \ddots & \dots \\ 0 & \dots & 0 & * \end{pmatrix} \tag{22.33}$$

where each of the *'s represent a sub-matrix of A and each 0 represents a matrix of all-zeros. A similar definition holds for Block (Lower) Triangular matrices.

A square matrix A is called **block-diagonal** if it can be represented as

$$A = \begin{pmatrix} A_1 & 0 & \dots & 0* \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & A_m \end{pmatrix} \tag{22.34}$$

for some square matrices $A_1, \dots, A_m$. If B is also block diagonal with

$$B = \begin{pmatrix} B_1 & 0 & \dots & 0* \\ 0 & B_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & B_m \end{pmatrix} \tag{22.35}$$

with each B_j the same size as each A_j, then

$$AB = \begin{pmatrix} A_1B_1 & 0 & \dots & 0* \\ 0 & A_2B_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & A_mB_m \end{pmatrix} \tag{22.36}$$

Theorem 22.2 Let $T \in \mathcal{L}(\mathbb{V})$ be normal, with $\mathbb{U}$ a subspace of $\mathbb{V}$ that is invariant under T. Then:

(1) $\mathbb{U}^\perp$ is invariant under T.

(2) $\mathbb{U}$ is invariant under T^*.

(3) $(T|_\mathbb{U})^* = (T^*)|_\mathbb{U}$.

(4) $T|_\mathbb{U}$ is normal on $\mathbb{U}$.

(5) $T|_{\mathbb{U}^\perp}$ is normal on $\mathbb{U}^\perp$.

Proof. (proof of theorem 22.2 (1))

Let $E = (e_1, \dots, e_m)$ be an orthonormal basis of $\mathbb{U}$.

Extend E to an orthonormal basis $B = (e_1, \dots, e_m, f_1, \dots, f_n)$ of $\mathbb{V}$ (corollary 17.8).

Then $F = (f_1, \dots, f_n)$ is an orthonormal basis of $\mathbb{U}^\perp$.

Since $\mathbb{U}$ is invariant under T there exist a_{ij} such that

$$Te_j = a_{j1}e_1 + \cdots + a_{jm}e_m \tag{22.37}$$

$$= a_{j1}e_1 + \cdots + a_{jm}e_m + 0 \cdot f_1 + \cdots + 0 \cdot f_n \tag{22.38}$$

and since B is an orthonormal basis of $\mathbb{V}$ there exist b_{ij} and c_{ij} such that

$$Tf_j = b_{j1}e_1 + \cdots + b_{jm}e_m + c_{j1}f_1 + \cdots + c_{jn}f_n \tag{22.39}$$

Hence by definition of the matrix of an operator (equation 12.1),

$$\mathcal{M}(T, B) = \begin{pmatrix} a_{11} & \dots & a_{m1} & b_{11} & \dots & b_{n1} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{1m} & \dots & a_{mm} & b_{1m} & \dots & b_{nm} \\ 0 & \dots & 0 & c_{11} & \dots & c_{n1} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & c_{1n} & \dots & c_{nn} \end{pmatrix} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix} \tag{22.40}$$

where A is an $m \times m$ matrix; B is $m \times n$; C is $n \times n$; and 0 is the $n \times m$ zero matrix.

By theorem 17.5, for any vector $v \in \mathbb{V}$,

$$\|Tv\|^2 = \sum_{k=1}^{m} |\langle Tv, e_k\rangle|^2 + \sum_{k=1}^{n} |\langle Tv, f_k\rangle|^2 \tag{22.41}$$

hence

$$\|Te_j\|^2 = \sum_{k=1}^{m} |\langle Te_j, e_k\rangle|^2 = \sum_{k=1}^{m} |\langle a_{j1}e_1 + \cdots + a_{jm}e_m, e_k\rangle|^2 \tag{22.42}$$

$$= \sum_{k=1}^{m} \left| \sum_{p=1}^{m} a_{jp}\langle e_p, e_k\rangle \right|^2 = \sum_{k=1}^{m} \left| \sum_{p=1}^{m} a_{jp}\delta_{pk} \right|^2 = \sum_{k=1}^{m} |a_{jk}|^2 \tag{22.43}$$

In other words, $\|Te_j\|^2$ is the sum of the squares of the jth *column* of A. Then

$$\sum_{j=1}^{m} \|Te_j\|^2 = \sum_{j=1}^{m}\sum_{k=1}^{m} |a_{jk}|^2 \tag{22.44}$$

which is the sum of the squares of all the elements of A.

Similarly,$\|T^*e_j\|^2$ is the sum of the squares of the jth *row* of $(A\ B)$, i.e.,

$$\|T^*e_j\|^2 = \sum_{k=1}^{m} |a_{kj}|^2 + \sum_{k=1}^{n} |b_{kj}|^2 \tag{22.45}$$

so that

$$\sum_{j=1}^{m} \|T^*e_j\|^2 = \sum_{j=1}^{m}\sum_{k=1}^{m} |a_{kj}|^2 + \sum_{j=1}^{m}\sum_{k=1}^{n} |b_{kj}|^2 \tag{22.46}$$

Since T is normal, by theorem 20.16 $\|Tv\| = \|T^*v\|$ for all $v \in \mathbb{V}$. Hence, in particular, $\|Te_j\| = \|T^*e_j\|$, which means that

$$\sum_{j=1}^{m} \|Te_j\|^2 = \sum_{j=1}^{m} \|T^*e_j\|^2 \tag{22.47}$$

Substituting equation 22.44 and 22.46 into equation 22.47 gives

$$\sum_{j=1}^{m}\sum_{k=1}^{m} |a_{jk}|^2 = \sum_{j=1}^{m}\sum_{k=1}^{m} |a_{kj}|^2 + \sum_{j=1}^{m}\sum_{k=1}^{n} |b_{kj}|^2 \tag{22.48}$$

Cancelling the first terms,

$$\sum_{j=1}^{m}\sum_{k=1}^{n} |b_{kj}|^2 = 0 \tag{22.49}$$

which means that a

$$b_{jk} = 0 \tag{22.50}$$

for all j, k, i.e., $B = 0$ and

$$\mathcal{M}(T) = \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} \tag{22.51}$$

Thus $Tf_k \in \text{span}(f_1, \dots, f_n)$ and, since $F = (f_1, \dots, f_n)$ is a basis of $\mathbb{U}^\perp$, $Tf_k \in \mathbb{U}^\perp$.

Since any $v \in \mathbb{U}^\perp$ can be written as a linear combination of the vectors in F, this means that $v \in \mathbb{U}^\perp \implies Tv \in \mathbb{U}^\perp$.

Hence $U^\perp$ is invariant under T.

(proof of theorem 22.2 (2)) By the same type of argument we made in the proof of

(1), we obtain

$$\mathcal{M}(T^*) = \begin{pmatrix} A^T & 0 \\ 0 & C^T \end{pmatrix} \tag{22.52}$$

Hence T^*e_j is a linear combination of the $E = (e_1, \dots, e_m)$.

Since any vector $u \in \mathbb{U}$ is a linear combination of E, we have $u \in \mathbb{U} \implies Tu \in \mathbb{U}$. Thus U is invariant under T^*.

<u>(proof of theorem 22.2 (3))</u> Define $S = T|_{\mathbb{U}}$ and pick any $v \in \mathbb{U}$. Then

$$\langle u, S^*v \rangle = \langle Su, v \rangle = \langle Tu, v \rangle = \langle u, T^*v \rangle \tag{22.53}$$

for all $u \in \mathbb{U}$.

By (2), U is invariant under T^*, hence

$$v \in \mathbb{U} \implies T^*v \in \mathbb{U}, \text{ i.e.,} \tag{22.54}$$

$$v \in U \implies (T^*)|_{\mathbb{U}} v \in \mathbb{U} \tag{22.55}$$

Hence

$$S^*v = T^*v \implies (T|_{\mathbb{U}})^* = T^*|_{\mathbb{U}} \tag{22.56}$$

<u>(proof of theorem 22.2 (4))</u> To prove that $T|_{\mathbb{U}}$ is normal on $\mathbb{U}$ we need to show that it commutes with its adjoint. But

$$(T|_{\mathbb{U}})^*(T|_{\mathbb{U}}) = (T^*|_{\mathbb{U}})(T|_{\mathbb{U}}) \qquad \text{(by (3))} \tag{22.57}$$

$$= (T|_{\mathbb{U}})(T^*|_{\mathbb{U}}) \qquad \text{(T is self-adjoint so)} T^*T = TT^* \tag{22.58}$$

$$= (T|_{\mathbb{U}})(T|_{\mathbb{U}})^* \qquad \text{(by (3))} \tag{22.59}$$

Thus $(T|_{\mathbb{U}})$ is normal.

(proof of theorem 22.2 (5)) The proof is analogous to the proof of (4). $\square$

Theorem 22.3 Let $\mathbb{V}$ be a real inner product space and $T \in \mathcal{L}(\mathbb{V})$ be an operator over $\mathbb{V}$. Then T is normal if and only if there is some orthonormal basis of $\mathbb{V}$ under which $\mathcal{M}(T)$ is block diagonal and each block is either 1×1 or 2×2 with the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{22.60}$$

with $b > 0$.

Proof. ($\Longrightarrow$) Suppose that T is normal, and prove by induction on $n = \dim(\mathbb{V})$.

For $n = 1$, the result follows immediately.

For $n = 2$, then either T is self-adjoint or is not self-adjoint. If it is self-adjoint, then by the real spectral theorem (theorem 21.2), it has a basis given entirely by eigenvectors of T, and hence by theorem 12.8, $\mathcal{M}(T)$ is diagonal with respect to some basis. If T is not self-adjoint, then by theorem 22.1 $\mathcal{M}(T)$ has the form shown.

As our inductive hypothesis, assume $n > 2$ and that the result holds for $n - 1$.

By theorem 14.1, T has an invariant subspace $\mathbb{U}$ of either dimension 1 or 2, such that the dimension 1 subspaces have real eigenvalues and in the dimension 2 subspaces $T|_{\mathbb{U}}$ have only have complex conjugate eigenvalue pairs with non-zero imaginary parts (i.e., they do not have real eigenvalues).

If $\dim \mathbb{U} = 1$, choose any $v \in \mathbb{U}$ such that $\|v\| = 1$ as a basis of $\mathbb{U}$.

If $\dim \mathbb{U} = 2$, then $T|_{\mathbb{U}}$ is normal (theorem 22.2, number (4)). But $T|_{\mathbb{U}}$ is not self-adjoint. To see this, note that if T were self-adjoint, it would have a real eigenvalue by lemma 21.5 and we have just noted that $T|_{\mathbb{U}}$ does not have real eigenvalues on the 2-dimensional subspaces.

Since $T|_{\mathbb{U}}$ is normal but not self adjoint (in dimension 2), theorem 22.1 applies. Then there is some basis in which $\mathcal{M}(T|_{\mathbb{U}})$ has the form given by equation 22.60.

Next observe that $\mathbb{U}^{\perp}$ is invariant under T and $T|_{\mathbb{U}}$ is normal on $\mathbb{U}^{\perp}$ by theorem 22.2. Since $\dim \mathbb{U}^{\perp} < n$, the inductive hypothesis holds on it.

Hence there is some basis B under which $\mathcal{M}(T|_{\mathbb{U}^{\perp}})$ has the properties predicted by the theorem. Adding to this basis to the basis of $\mathbb{U}^{\perp}$ give a basis B' under which $\mathcal{M}(T, B')$ has the properties of the theorem.

($\Longleftarrow$) Suppose there is some basis E under which $\mathcal{M}(T)$ is block diagonal

$$\mathcal{M}(T) = \text{diagonal}(A_1, \dots, A_n) \tag{22.61}$$

with the blocks A_i either 1×1 or 2×2 with the form given in equation 22.60. Then

$$\mathcal{M}(T^*) = \text{diagonal}(A_1^T, \dots, A_n^T) \tag{22.62}$$

For each A_i, if A_i is 1×1, then $A_i = A_i^T$, and hence commutes with it's transpose. If A_i is 2×2 then it has the form of equation 22.60, and we have already shown in equation 22.27 that this matrix commutues with its transpose. Hence

$$A_i A_i^T = A_i^T A_i \tag{22.63}$$

and therefore

$$\mathcal{M}(T)\mathcal{M}(T*) = \text{diagonal}(A_1, \dots, A_n)\text{diagonal}(A_1^T, \dots, A_n^T) \tag{22.64}$$

$$= \text{diagonal}(A_1 A_1^T, \dots, A_n A_n^T) \tag{22.65}$$

$$= \text{diagonal}(A_1^T A_1, \dots, A_n^T A_n) \tag{22.66}$$

$$= \text{diagonal}(A_1^T, \dots, A_n^T)\text{diagonal}(A_1, \dots, A_n) \tag{22.67}$$

$$= \mathcal{M}(T*)\mathcal{M}(T) \tag{22.68}$$

Thus T is normal. □

Topic 23

Normal Matrices

Definition 23.1 An $n \times n$ square matrix is a **normal matrix** if all of its eigenvectors are orthogonal.

Theorem 23.2 A normal matrix is sef-adjoint if and only if all of its eigenvalues are real.

Proof. Let A be a normal matrix. Then it has n orthogonal eigenvectors $v_1, \dots, v_n$ with eigevalues $\lambda_1, \dots, \lambda_n$. Define U as the matrix whose column vectors are the orthonormal eigenvectors v_i and $\Lambda = \text{diagonal}(\lambda_1, \dots, \lambda_n)$. Then $U^{-1} = U^*$ and

$$NU = N \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \tag{23.1}$$

$$= \begin{pmatrix} Nv_1 & \cdots & Nv_n \end{pmatrix} \tag{23.2}$$

$$= \begin{pmatrix} \lambda_1 v_1 & \cdots & \lambda_n v_n \end{pmatrix} \tag{23.3}$$

$$= \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \text{diagonal}(\lambda_1, \dots, \lambda_n) \tag{23.4}$$

$$= U\Lambda \tag{23.5}$$

Hence

$$U^*NU = \Lambda \tag{23.6}$$

or equivalently

$$N = U\Lambda U^* \tag{23.7}$$

So N is similar to the diagonal matrix Λ and has the same eigenvalues.

N is self-adjoint if and only if

$$N^* = N \tag{23.8}$$

$$(U\Lambda U^*)^* = U^*\Lambda U \tag{23.9}$$

$$U\Lambda^* U = U^*\Lambda U \tag{23.10}$$

$$\Lambda^* = \Lambda \tag{23.11}$$

which is true if and only if $\overline{\lambda_i} = \lambda_i$, i.e., the eigenvalues are all real. □

Theorem 23.3 If N is a normal matrix then there exists square, commuting matrices A and B such that

$$N = A + iB \text{ and } AB = BA \tag{23.12}$$

Proof. Let $\Lambda = U^*NU$ be the canonical diagonal form of N. Then we can write

$$\Lambda = \Lambda_r + i\Lambda_i \tag{23.13}$$

where

$$\Lambda_r = \text{diagonal}(\,\text{Re}(\lambda_1)\,, \ldots, \,\text{Re}(\lambda_n)\,) \tag{23.14}$$

$$\Lambda_i = \text{diagonal}(\,\text{Im}(\lambda_1)\,, \ldots, \,\text{Im}(\lambda_n)\,) \tag{23.15}$$

Hence

$$N = U(\Lambda_r + i\Lambda_i)U^* = U\Lambda_r U^* + iU\Lambda_i U^* \tag{23.16}$$

Therefore

$$A = U\Lambda_r U^*, \qquad B = U\Lambda_i U^* \tag{23.17}$$

and

$$\begin{aligned} AB &= U\Lambda_r U^* U\Lambda_i U^* && (23.18)\\ &= U\Lambda_r\Lambda_i U^* && (23.19)\\ &= U\Lambda_i\Lambda_r U^* \quad \text{(because diagonal matrices commute)} && (23.20)\\ &= U\Lambda_i U^* U\Lambda_r U^* && (23.21)\\ &= BA && (23.22) \end{aligned}$$

Hence $N = A + iB$ where A and B commute. □

The following theorem shows that our previous definition of normality is consistent with definition 23.1.

Theorem 23.4 N is normal if and only if it commutes with its adjoint, i.e,

$$NN^* = N^*N \tag{23.23}$$

Proof. ($\Longrightarrow$) Suppose N is normal. Then by equation 23.6

$$\begin{aligned} NN^* &= U\Lambda U^* U\Lambda^* U^* && (23.24)\\ &= U\Lambda\Lambda^* U^* = U\Lambda^*\Lambda U^* && (23.25)\\ &= U\Lambda^* U^* U\Lambda U^* = N^*N && (23.26) \end{aligned}$$

($\Longleftarrow$) Suppose that $N^*N = NN^*$. We must show that N has a complete set of orthogonal eigenvectors.

Define the matrices A and B by

$$A = \frac{1}{2}(N + N^*), \qquad B = \frac{1}{2i}(N - N^*) \tag{23.27}$$

Then

$$N = A + iB \tag{23.28}$$

Furthermore, since $NN^* = N^*N$,

$$\begin{aligned}
4iAB &= (N + N^*)(N - N^*) && (23.29)\\
&= N^2 + N^*N - NN^* - (N^*)^2 && (23.30)\\
&= N^2 + NN^* - N^*N - (N^*)^2 && (23.31)\\
&= N(N + N^*) - N^*(N + N^*) && (23.32)\\
&= (N - N^*)(N + N^*) && (23.33)\\
&= 4iBA && (23.34)
\end{aligned}$$

Hence $AB = BA$.

Let $a_1 \geq a_2 \geq \cdots \geq a_n$ be the eigenvalues of A and let $\alpha = \text{diagonal}(a_1, \dots, a_n)$. We know the eigenvalues are real because $A = A^*$ is self adjoint (see theorem 20.9).

Since A is real and self-adjoint it has n mutually orthogonal unit vectors (real spectral theorem). Let V be the matrix whose columns are eigenvectors of A. Then

$$V^*AV = \alpha \tag{23.35}$$

To see this

$$V^*AV = V^*\begin{pmatrix} Av_1 & \cdots & Av_n \end{pmatrix} \tag{23.36}$$

$$= V^*\begin{pmatrix} \alpha_1 v_1 & \cdots & \alpha_n v_n \end{pmatrix} \tag{23.37}$$

$$= V^*V\alpha = \alpha \tag{23.38}$$

Let

$$K = V^*BV \tag{23.39}$$

Then since B is self adjoint,

$$K^* = (V^*BV)^* = V^*B^*V = V^*BV = K \tag{23.40}$$

hence K is self-adjoint; but K is not necessarily diagonal.

Since A and B commute we have

$$\alpha K = (V^*AV)(V^*BV) = V^*ABV = V^*BAV = (V^*BV)(V^*AV) = K\alpha \tag{23.41}$$

and therefore K and α commute. Hence $(\alpha K)_{ij} = (K\alpha)_{ij}$,

$$\sum_{p=1}^{n} \alpha_{ip} K_{pj} = \sum_{p=1}^{n} K_{ip}\alpha_{pj} \tag{23.42}$$

$$\sum_{p=1}^{n} \alpha_i \delta_{ip} K_{pj} = \sum_{p=1}^{n} K_{ip}\alpha_j \delta_{pj} \tag{23.43}$$

$$\alpha_i K_{ij} = K_{ij}\alpha_j \tag{23.44}$$

Thuse $K_{ij} = 0$ unless $\alpha_i = \alpha_j$.

If the eigenvalues are distinct, then K is diagonal.

If the eigenvalues are not distinct then we can write α as a block diagonal matrix

$$\alpha = \text{diagonal}(\Lambda_1, \dots, \Lambda_m) \tag{23.45}$$

where each Λ_i is a diagonal matrix with all of its diagonal elements equal, and correspondingly

$$K = \text{diagonal}(K_1, \dots, K_m) \tag{23.46}$$

where each of the K_i is self-adjoint and has the same dimensions as the corresponding Λ_i. (This follows from equation 23.44.)

Since K_i is self adjoint, there is some self-adjoint matrix W_i such that $W_i^* K_i W_i$ is diagonal. Let

$$W = \text{diagonal}(W_1, \dots, W_m) \tag{23.47}$$

Then W is also self-adjoint and

$$P = W^*KW = \begin{pmatrix} W_1^* & & 0 \\ & \ddots & \\ 0 & & W_m^* \end{pmatrix} \begin{pmatrix} K_1 & & 0 \\ & \ddots & \\ 0 & & K_m \end{pmatrix} \begin{pmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_m \end{pmatrix} \tag{23.48}$$

$$= \begin{pmatrix} W_1^* K_1 W & & 0 \\ & \ddots & \\ 0 & & W_m^* K_m W_m \end{pmatrix} \tag{23.49}$$

is diagonal. Now compute

$$W^*\alpha W = \begin{pmatrix} W_1^* & & 0 \\ & \ddots & \\ 0 & & W_n^* \end{pmatrix} \begin{pmatrix} \Lambda_1 & & 0 \\ & \ddots & \\ 0 & & \Lambda_m \end{pmatrix} \begin{pmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_m \end{pmatrix} \tag{23.50}$$

$$= \begin{pmatrix} W_1^*\Lambda_1 W_1 & & 0 \\ & \ddots & \\ 0 & & W_m^*\Lambda_m W_m \end{pmatrix} \tag{23.51}$$

$$= \begin{pmatrix} W_1^*\lambda_1 I W_1 & & 0 \\ & \ddots & \\ 0 & & W_m^*\lambda_m I W_m \end{pmatrix} \tag{23.52}$$

$$= \alpha \tag{23.53}$$

Let $U = VW$. Then

$$U^*AU = (W^*V^*)A(VW) = W^*(V^*AV)W = W^*\alpha W = \alpha \tag{23.54}$$

is a real diagonal matrix, as is

$$U^*BU = (W^*V^*)B(VW) = W^*(V^*BV)W = W^*KW = P \tag{23.55}$$

Therefore

$$U^*NU = U^*(A+iB)U = U^*AU + iU^*BU = \alpha + iP \tag{23.56}$$

is diagonal. Hence N is is similar to a diagonal matrix, and the columns of U are the orthogonal eigenvectors of N.

□

Topic 24

Positive Operators

Definition 24.1 Let $\mathbb{V}$ be a finite-dimensioned non-zero inner-product space over $\mathbb{V}$ and let $T \in \mathcal{L}(\mathbb{V})$ be a self-adjoint operator over $\mathbb{V}$. Then we say T is a **positive operator** operator, or a **positive semi-definite operator**, or just T is **positive**, if T is self-adjoint and

$$\langle Tv, v\rangle \geq 0 \tag{24.1}$$

for all $v \in \mathbb{V}$.

The term positive is a bit misleading because equality is allowed in equation 24.1; a better term might be non-negative rather than positive. A strictly positive operator that satisfies

$$\langle Tv, v\rangle > 0 \tag{24.2}$$

unless $v = 0$ is call **positive definite**. When equality is allowed, the "definite" becomes "semi-definite" but in our shorthand term, we call it positive.

Example 24.1 Every orthogonal projection is positive.

Definition 24.2 A self-adjoint matrix A is called positive definite if

$$\overline{v}^T Av > 0 \tag{24.3}$$

for all $v \neq 0$, and is called positive-semi-definite if

$$\overline{v}^T Av \geq 0 \tag{24.4}$$

Definition 24.3 An operator S is called the **square root** of an operator T if $S^2 = T$.

Example 24.2 Let

$$T(x, y, z) = (z, 0, 0) \tag{24.5}$$

and

$$S(x, y, z) = (y, z, 0) \tag{24.6}$$

Then

$$S^2(x, y, z) = S(y, z, 0) = (z, 0, 0) = T(x, y, z) \tag{24.7}$$

Hence S is a square root of T.

Example 24.3 Square Root of a Matrix Find the square root of

$$T = \begin{pmatrix} 125 & -75 \\ -75 & 125 \end{pmatrix} \tag{24.8}$$

Let $M = \sqrt{T}$ be denoted by

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \tag{24.9}$$

Then

$$T = M^2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 + bc & ab + bd \\ ac + cd & bc + d^2 \end{pmatrix} \tag{24.10}$$

This gives four equations in four unknowns:

$$a^2 + bc = 125 \tag{24.11}$$

$$ab + bd = -75 \tag{24.12}$$

$$ac + cd = -75 \tag{24.13}$$

$$bc + d^2 = 125 \tag{24.14}$$

The equations are non-linear and there are four solutions for M: Substracting equation (24.11) from (24.14) we find

$$a^2 = d^2$$

This gives two choices, $a = d$ or $a = -d$.

We can rule out $a = -d$ since that gives a contradiction in (24.12) and (24.13).

Start with $a = d$. Substituting this into equations (24.12) and (24.13) gives

$$2ab = -75 \implies b = -75/2a \tag{24.15}$$

$$2ac = -75 \implies c = -75/2a \tag{24.16}$$

From equation (24.11)

$$125 = a^2 + bc = a^2 + \frac{75^2}{4a} \tag{24.17}$$

Rearranging,

$$a^4 - 125a^2 + \frac{75^2}{4} = 0 \tag{24.18}$$

This is a quadratic in a^2

$$a^2 = \frac{125 \pm \sqrt{125^2 - 75^2}}{2} = \frac{125 \pm 100}{2} \tag{24.19}$$

Hence $a^2 = 225/2$ or $a^2 = 125/2$.

Thus $a = \pm 5/\sqrt{2}$ or $a = \pm 15/\sqrt{2}$.

For $a = 5/\sqrt{2}$, then $d = 5/\sqrt{2}$. Hence $b = c = -75/2a = -15/\sqrt{2}$.

For $a = d = -5/\sqrt{2}$, this gives $b = c = 15/\sqrt{2}$.

The remaining two solutions are negatives of the first two. Finally we get the four possible matrices:

$$M = \frac{1}{\sqrt{2}}\begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} -5 & 15 \\ 15 & -5 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 5 & -15 \\ -15 & 5 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 15 & -5 \\ -5 & 15 \end{pmatrix} \tag{24.20}$$

Theorem 24.4 Let $T \in \mathcal{L}(\mathbb{V})$. Then the following are equivalent.

(1) T is positive.

(2) T is self-adjoint and all of its eigenvalues are non-negative.

(3) T has a positive square root.

(4) T has a self-adjoint square root.

(5) There exists some $S \in \mathcal{L}(\mathbb{V})$ such that $T = S^*S$.

Proof. $\underline{((1) \implies (2))}$ Let T be positive. The by definition of a positive operator, it is self-adjoint.

Let λ be an eigenvalue of T with non-zero eigenvector. Then

$$0 \leq \langle Tv, v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle = \lambda\|v\|^2 \tag{24.21}$$

Since $\|v\|^2 > 0$ this gives $\lambda \geq 0$.

$\underline{((2) \implies (3))}$ By (2) T is self adjoint and has non-negative eigenvalues.

If $\mathbb{V}$ is real, by the real spectral theorem (theorem 21.2), T has an orthonormal basis of eigenvectors.

If $\mathbb{V}$ is complex, since every self-adjoint map is normal, T is normal, and the complex spectral theorem (theorem 21.1) gives the same result.

Let $E = (e_1, \dots, e_n)$ be a basis of $\mathbb{V}$ consisting of eigenvectors of T with corresponding eigenvalues $(\lambda_1, \dots, \lambda_n)$.

We are given that each $\lambda_i \geq 0$ so they each have a real square root.

Define $S \in \mathcal{L}(\mathbb{V})$ by

$$Se_j = \sqrt{\lambda_i}e_j \tag{24.22}$$

Any vector $v \in \mathbb{V}$ has an expansion $v = a_1e_1 + \cdots + a_ne_n$ for some $a_1, \dots, a_n$. Hence

$$\begin{aligned} \langle Sv, v\rangle &= \left\langle S\sum_{i=1}^n a_ie_i, \sum_{j=1}^n a_je_j\right\rangle && (24.23)\\ &= \left\langle \sum_{i=1}^n a_iSe_i, \sum_{j=1}^n a_je_j\right\rangle && \text{(linearity)} \quad (24.24)\\ &= \left\langle \sum_{i=1}^n a_i\lambda_ie_i, \sum_{j=1}^n a_je_j\right\rangle && \text{(eigenvalues)} \quad (24.25) \end{aligned}$$

Hence by linearity, homogenity, and conjugate homogeneity in the second argument,

$$\langle Sv, v\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\lambda_i\overline{a_j}\langle e_i, e_j\rangle \tag{24.26}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\lambda_i\overline{a_j}\delta_{ij} \tag{24.27}$$

$$= \sum_{i=1}^{n} \lambda_i|a_i|^2 \geq 0 \tag{24.28}$$

becasue it is the sum of non-negative numbers. Hence S is positive.

Furthermore, since any $v = a_1e_1 + \cdots + a_ne_n$,

$$S^2v = S^2(a_1e_1 + \cdots + a_ne_n) \tag{24.29}$$

$$= S(a_1Se_1 + \cdots + a_nSe_n) \tag{24.30}$$

$$= S(a_1\sqrt{\lambda_1}e_1 + \cdots + a_n\sqrt{\lambda_n}e_n) \tag{24.31}$$

$$= a_1\sqrt{\lambda_1}Se_1 + \cdots + a_n\sqrt{\lambda_n}Se_n \tag{24.32}$$

$$= a_1\lambda_1e_1 + \cdots + a_n\lambda_ne_n \tag{24.33}$$

$$= a_1Te_1 + \cdots + a_nTe_n \quad (\text{because } Te_j = \lambda_je_j) \tag{24.34}$$

$$= Tv \tag{24.35}$$

Hence S is a postive square root of T, which proves statement (3).

<u>((3) $\implies$ (4))</u> Since T has a positive square root S, the square root is self adjoint, by definition of a positive operator.

<u>((4) $\implies$ (5))</u> If (4) is true then T has a self-adjoint square root S. Since $S = S^*$ then $T = S^2 = SS^* = S^*S$, which is (5).

((5) $\implies$ (1)) Suppose that there exists some operator S such that $T = S^*S$. Then

$$T^* = (S^*S)^* = S^*(S^*)^* = S^*S = T \tag{24.36}$$

Therefore T is self-adjoint. To show that T is positive, for any $v \in \mathbb{V}$,

$$\langle Tv, v\rangle = \langle S^*Sv, v\rangle = \langle Sv, Sv\rangle = \|Sv\| \geq 0 \tag{24.37}$$

which proves that T is positive. Hence (1) follows from (5). □

Lemma 24.5 Let T be self-adjoint, and let $\lambda_1, \dots, \lambda_n$ be the distinct

$$\mathbb{V} = \text{null}(T - \lambda_1 I) \oplus \cdots \oplus \text{null}(T - \lambda_n I) \tag{24.38}$$

and each vector in $\text{null}(T - \lambda I)$ is orthogonal to all of the subspaces in the decomposition.

Proof. By the spectral theorems (theorems 21.1 and 21.2) $\mathbb{V}$ has an basis consisting of eigenvectors of T.

Equation 24.38 follows from theorem 12.8, statement (5). □

Theorem 24.6 Let T be a positive operator on $\mathbb{V}$. Then T has a unique positive square root.

Proof. (Omitted; see Axler if you are interested. The proof uses lemma 24.5) □

Theorem 24.7 A self-adjoint matrix is positive definite if and only if all of its eigenvalues are positive.

Proof. ($\implies$) Let M be a matrix with only positive eigenvalues. We have already observed (remark 19.6) that if M is self-adjoint then the Schur decomposition gives

$$D = U^{-1}MU \tag{24.39}$$

where D is a diagonal matrix consisting of the eigenvalues of M and $U^{-1} = U^*$.

Let u, v be vectors such that $u = Uv$, i.e, for any u, $v = U^{-1}u = U^*u$ Then

$$\begin{aligned}
\langle Mu, u\rangle &= \langle MUv, Uv\rangle && (24.40)\\
&= \langle U^*MUv, v\rangle && (24.41)\\
&= \langle Dv, v\rangle && (24.42)\\
&= \langle(\lambda_1 v_1, \dots, \lambda_n v_n), v\rangle && (24.43)\\
&= \lambda_1|v_1|^2 + \cdots + \lambda_n|v_n|^2 > 0 && (24.44)
\end{aligned}$$

unless $v = 0$, which only occurs if $u = 0$.

($\Longleftarrow$) Suppose that $\langle Mu, u\rangle > 0$ unless $u = 0$.

Let $(u_1, \dots, u_n)$ be unit length eigenvectors of M with eigenvalues λ_i. Then

$$\langle Mu_i, u_i\rangle = \lambda_i\|u_i\|^2 = \lambda_i \tag{24.45}$$

Since any vector v can be written as a linear combindation of the u_i,

$$v = a_1u_1 + \cdots + a_nu_n \tag{24.46}$$

$$\langle Mv, v\rangle = \langle a_1u_1 + \cdots + a_nu_n, a_1u_1 + \cdots + a_nu_n\rangle \tag{24.47}$$

$$= \sum_{i=1}^{n} a_i\langle u_i, a_1u_1 + \cdots + a_nu_n\rangle \tag{24.48}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\overline{a_j}\langle u_i, u_j\rangle = \sum |a_i|^2\lambda_i \tag{24.49}$$

which is greater than zero if and only if all eigenvalues are positive. □

Topic 25

Isometries

Definition 25.1 A operator $S \in \mathcal{L}(\mathbb{V})$ is called an **isometry** if it preserves length in the sense

$$\|Sv\| = \|v\| \tag{25.1}$$

for all $v \in \mathbb{V}$.

Example 25.1 Suppose that $S \in \mathcal{L}(\mathbb{V})$ such that

$$S(e_j) = \lambda_j e_j \tag{25.2}$$

where $E = (e_1, \dots, e_n)$ is an orthonormal basis and $|\lambda_j| = 1$ for all $j = 1, \dots, n$. Then S is an isometry. To see this, let $v \in \mathbb{V}$; then since E is a basis (see theorem 17.5),

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n \tag{25.3}$$

$$\|v\|^2 = |\langle v, e_1\rangle|^2 + \cdots + |\langle v, e_n\rangle|^2 \tag{25.4}$$

Since the λ_j are eigenvalues of S corresponding to the eigenvectors e_j,

$$Sv = \langle v, e_1 \rangle Se_1 + \cdots + \langle v, e_n \rangle Se_n \tag{25.5}$$

$$= \lambda_1 \langle v, e_1 \rangle e_1 + \cdots + \lambda_n \langle v, e_n \rangle e_n \tag{25.6}$$

Using the fact that $|\lambda_i| = 1$ and that $\langle e_i, e_j \rangle = \delta_{ij}$ gives

$$\|Sv\| = \langle \sum_i \lambda_i \langle v, e_i \rangle e_i, \sum_j \lambda_j \langle v, e_j \rangle e_j \rangle \tag{25.7}$$

$$= \sum_{i,j} \lambda_i \overline{\lambda_j} \langle v, e_i \rangle \overline{\langle v, e_j \rangle} \langle e_i, e_j \rangle \tag{25.8}$$

$$= \sum_{i,j} \lambda_i \overline{\lambda_j} \langle v, e_i \rangle \overline{\langle v, e_j \rangle} \delta_{ij} \tag{25.9}$$

$$= |\lambda_1|^2 |\langle v, e_1 \rangle|^2 + \cdots + |\lambda_n|^2 |\langle v, e_n \rangle|^2 \tag{25.10}$$

$$= |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2 \tag{25.11}$$

$$= \|v\|^2 \tag{25.12}$$

where the last line follows from equation 25.4. Therefore S is an isometry. □

Theorem 25.2 Properties of Isometries. Let $S \in \mathcal{L}(\mathbb{V})$. Then the following are equivalent:

1. S is an isometry.
2. $\langle Su, Sv \rangle = \langle u, v \rangle$ for all $u, v \in \mathbb{V}$.
3. $S^*S = I$
4. $(Se_1, \ldots, Se_n)$ is orthonormal whenever $(e_1, \ldots, e_n)$ is orthonormal.
5. There exists an orthonormal basis $(e_1, \ldots, e_n)$ such that $(Se_1, \ldots, Se_n)$ is orthonormal.

6. S^* is an isometry.

7. $\langle S^*u, S^*v\rangle = \langle u, v\rangle$ for all $u, v \in \mathbb{V}$.

8. $SS^* = I$

9. $(S^*e_1, \dots, S^*e_n)$ is orthonormal whenever $(e_1, \dots, e_n)$ is orthonormal.

10. There exists an orthonormal basis $(e_1, \dots, e_n)$ such that $(S^*e_1, \dots, S^*e_n)$ is orthonormal.

Proof. We will show that (1) through (5) are equivalent; conclusions (6) through (10) are analogous to (1) through (5) and hence the proof that (6) through (10) are equivalent to each other is completely analogous. To show that (1) through (5) are equivalent to (6) through (10) we compare items (3) and (8), and observe that $SS^* = I$ if and only if $S^*S = I$.

<u>((1) $\implies$ (2))</u> Assume that S is an isometry; then $\|Sx\| = \|x\|$ for all $x \in \mathbb{V}$.

Then for any $u, v \in \mathbb{V}$,

$$\begin{aligned}
\langle Su, Sv\rangle &= \frac{1}{4}\left(\|Su + Sv\|^2 - \|Su - Sv\|^2\right) && \text{(homework)} \quad (25.13)\\
&= \frac{1}{4}\left(\|S(u + v)\|^2 - \|S(u - v)\|^2\right) && (25.14)\\
&= \frac{1}{4}\left(\|u + v\|^2 - \|u - v\|^2\right) && \text{(isometry)} \quad (25.15)\\
&= \langle u, v\rangle && \text{(homework)} \quad (25.16)
\end{aligned}$$

<u>((2) $\implies$ (3))</u> Suppose that $\langle Su, Sv\rangle = \langle u, v\rangle$ for all $u, v \in \mathbb{V}$. Then

$$\langle (S^*S - I)u, v\rangle = \langle S^*Su, v\rangle - \langle u, v\rangle = \langle Su, Sv\rangle - \langle u, v\rangle = 0 \quad (25.17)$$

If we let $v = (S^*S - I)u$ this means $\langle v, v\rangle = 0$ which is true if and only if $v = 0$; hence

$$(S^*S - I)u = 0 \tag{25.18}$$

for all $u \in \mathbb{U}$. Thus

$$S^*S - I = 0 \implies S^*S = I \tag{25.19}$$

<u>((3) $\implies$ (4))</u> Assume that $S^*S = I$. Let $E = (e_1, \dots, e_n)$ be orthonormal.

$$\langle Se_j, Se_k \rangle = \langle S^*Se_j, e_k \rangle = \langle e_j, e_k \rangle = \delta_{jk} \tag{25.20}$$

Hence $(Se_1, \dots, Se_n)$ is an orthonormal list of vectors. Thus (3) $\implies$ (4).

<u>((4) $\implies$ (5))</u> Assume that $(Se_1, \dots, Se_n)$ is an orthonormal list of vectors whenever $(e_1, \dots, e_n)$ is orthonormal. Pick any such orthonormal list of vectors $(e_1, \dots, e_n)$. Then $(Se_1, \dots, Se_n)$ is orthonormal. Thus (4) $\implies$ (5).

<u>((5) $\implies$ (1))</u> Assume that there exists an orthonormal list $(e_1, \dots, e_n)$ such that $(Se_1, \dots, Se_n)$ is orthonormal.

Let $v \in \mathbb{V}$. Then since $\langle Se_i, Se_j \rangle = \delta_{ij}$,

$$\begin{aligned}
\|Sv\|^2 &= \left\| S \sum_{i=1}^n \langle v, e_i \rangle e_i \right\|^2 && (25.21)\\
&= \left\| \sum_{i=1}^n \langle v, e_i \rangle Se_i \right\|^2 && (25.22)\\
&= \left\langle \sum_{i=1}^n \langle v, e_i \rangle Se_i, \sum_{j=1}^n \langle v, e_j \rangle Se_j \right\rangle && (25.23)\\
&= \sum_{i=1}^n \sum_{j=1}^n \langle v, e_j \rangle \overline{\langle v, e_i \rangle} \langle Se_i, Se_j \rangle && (25.24)\\
&= \sum_{i=1}^n |\langle v, e_i \rangle|^2 = \|v\|^2 && (25.25)
\end{aligned}$$

where the last step comes from theorem 17.5. Thus S is an isometry and (5) $\Longrightarrow$ (1). □

Remark 25.3 Every isometry is normal, since $S^*S = I = SS^*$.

Theorem 25.4 Let $\mathbb{V}$ be a complex non-zero inner-product space and let $S \in \mathcal{L}(\mathbb{V})$. Then S is an isometry if and only if there is an orthonormal basis of $\mathbb{V}$ consisting of eigenvalues of S with $|\lambda_i| = 1$ for all $i = 1, \dots, n$.

Proof. ($\Longrightarrow$) Suppose that S is an isometry.

Hence there exists an orthonormal basis $(e_1, \dots, e_n)$ consisting solely of eigenvectors of S (Spectral Theorem, theorem 21.1). Let $(\lambda_1, \dots, \lambda_n)$ be the corresponding eigenvalues. Then

$$|\lambda_j| = \|\lambda_j e_j\| = \|Se_j\| = \|e_j\| = 1 \tag{25.26}$$

($\Longleftarrow$) This was proven already as example 25.1. □

Theorem 25.5 Let $\mathbb{V}$ be a real inner-product space with $S \in \mathcal{L}(\mathbb{V})$. S is an isometry if and only f there is an orthonormal basis of $\mathbb{V}$ in which S has a block-diagonal matrix where each block is either

(a) a 1×1 matrix with a 1 or -1; or

(b) a 2×2 matrix of the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ with $a^2 + b^2 = 1$ and $b > 0$.

Remark 25.6 The 2×2 blocks have the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \tag{25.27}$$

which is a rotation by an angle θ in $\mathbb{R}^2$; thus every rotation in $\mathbb{R}^n$ is composed of a sequence of rotations in the coordinate planes.

Corollary 25.7 Let $\mathbb{V} = \mathbb{R}^n$ where n is odd, and let S be an isometry on $\mathbb{V}$. Then S has either -1 or 1 (or possibly both) as an eigenvalue.

Proof. (Theorem 25.5.) ($\Longrightarrow$)

Let S be an isometry. Then S is normal (see remark 25.3).

By theorem 22.3 there is an orthonormal basis under which $\mathcal{M}(T)$ is block diagnoal with the blocks either 1×1 or 2×2 with the form $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ with $b > 0$.

We need to show that $a^2 + b^2 = 1$ for each 2×2 block; and that the 1×1 blocks are ± 1.

Let λ be any 1×1 diagonal block. Then there is a basis vector e_j such that λ is an eigevalue

$$Se_j = \lambda e_j \tag{25.28}$$

Since S is an isometry, $\|Se_j\| = \|e_j\| = 1$ (the last step follows because e_j is a vector in an orthonormal basis, hence has length 1). Hence $|\lambda| = 1$; since $\mathbb{V}$ is real, this means $\lambda = \pm 1$.

Let

$$\Lambda = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \tag{25.29}$$

be any 2×2 diagonal block. Then there are basis vectors e_j, e_{j+1} such that

$$Se_j = ae_j + be_{j+1} \tag{25.30}$$

To see this let $\alpha = (x, y)^T \in \mathbb{R}^2$. Then

$$\Lambda\alpha = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax - by \\ bx + ay \end{pmatrix} = a \begin{pmatrix} x \\ y \end{pmatrix} + b \begin{pmatrix} -y \\ x \end{pmatrix} \tag{25.31}$$

$$= a\alpha + b\beta \tag{25.32}$$

where $\langle \alpha, \beta \rangle = 0$ and $\|\alpha\| = \|\beta\|$. Then let construct e_j and e_k in $\mathbb{R}^n$ by replacing the remaining components with 0. The result $Se_j = ae_j + be_{j+1}$ follows.

Hence since S is an isometry,

$$1 = \|e_j\|^2 = \|Se_j\|^2 \tag{25.33}$$

$$= \langle ae_j + be_{j+1}, ae_j + be_{j+1} \rangle \tag{25.34}$$

$$= a^2\langle e_j, e_j \rangle + ab\langle e_j, e_{j+1} \rangle + ba\langle e_{j+1}, e_j \rangle + b^2\langle e_{j+1}, e_{j+1} \rangle \tag{25.35}$$

$$= a^2 + b^2 \tag{25.36}$$

($\Longleftarrow$) Suppose that for some orthonormal basis the matrix has the desired form. Then we can decomposed $\mathbb{V}$ into subspaces such that

$$\mathbb{V} = \mathbb{U}_1 \oplus \cdots \oplus \mathbb{U}_m \tag{25.37}$$

where each $\mathbb{U}_i$ is a subspace of dimension 1 or 2, each subspace is orthogonal, and because of the form of the matrix, each $S|_{\mathbb{U}_j}$ is an isometry on U_j. (To verify the last statement in dimension 2, let $v \in \mathbb{U}_j = (0, \ldots, 0, x, y, 0, \ldots)^T$. Then

$$\|Sv\| = \|\text{diagonal}(\cdots,\Lambda,\cdots)(0,\cdots,0,x,y,0,\cdots,0)\| \tag{25.38}$$

$$= \left\| \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right\| = \left\| a \begin{pmatrix} x \\ y \end{pmatrix} + b \begin{pmatrix} -y \\ x \end{pmatrix} \right\| \tag{25.39}$$

$$= \|au + bu'\| \tag{25.40}$$

where $u = (x,y)^T$, $u' = (-y,x)^T$, and hence $\langle u, u'\rangle = 0$ and $\|u\| = \|u'\|$. Hence

$$\|Sv\|^2 = \langle au + bu', au + bu'\rangle = (a^2+b^2)\|u\|^2 = \|u\|^2 \tag{25.41}$$

Therefore each $S|_{\mathbb{U}_j}$ is an isometry on U_j.)

Finally, to show that S is an isometry, let $v \in \mathbb{V}$. Then we can write

$$v = u_1 + \cdots + u_m \tag{25.42}$$

where $u_j \in \mathbb{U}_j$. Since $S_{\mathbb{U}_j}$ is an isometry on $\mathbb{U}_j$,

$$\|Sv\|^2 = \|Su_1 + \cdots + Su_m\|^2 \tag{25.43}$$

$$= \langle Su_1 + \cdots + Su_m, Su_1 + \cdots + Su_m\rangle \tag{25.44}$$

$$= \sum_{i,j}\langle Su_i, Su_j\rangle = \sum_{i,j}\langle S^*Su_i, u_j\rangle = \sum_{i,j}\langle u_i, u_j\rangle \tag{25.45}$$

$$= \sum_i \langle u_i, u_i\rangle \text{ (because the } \mathbb{U}_j \text{ are orthogonal)} \tag{25.46}$$

$$= \sum \|u_i\|^2 = \|v\|^2 \tag{25.47}$$

which proves that S is an isometry. □

Topic 26

Singular Value Decomposition

Theorem 26.1 Polar Decomposition. Let $T \in \mathcal{L}(\mathbb{V})$. Then there exists an isometry $S \in \mathcal{L}(\mathbb{V})$ such that

$$T = S\sqrt{T^*T} \tag{26.1}$$

Lemma 26.2 Let $\mathbb{U}$ be a subspace of $\mathbb{V}$. Then

$$\dim \mathbb{V} = \dim \mathbb{U} + \dim \mathbb{U}^\perp \tag{26.2}$$

Proof. This follows from $\mathbb{V} = \mathbb{U} \oplus \mathbb{U}^\perp$ □

Proof. (of theorem 26.1). Let $v \in \mathbb{V}$. Then

$$\|Tv\|^2 = \langle Tv, Tv\rangle = \langle T^*Tv, v\rangle \tag{26.3}$$

$$= \langle \sqrt{T^*T}\sqrt{T^*T}v, v\rangle \tag{26.4}$$

$$= \langle \sqrt{T^*T}v, \sqrt{T^*T}v\rangle \tag{26.5}$$

$$= \|\sqrt{T^*T}v\|^2 \tag{26.6}$$

Now define

$$S' : \text{range}\sqrt{T^*T} \mapsto \text{range}T \tag{26.7}$$

by

$$S'\left(\sqrt{T^*T}v\right) = Tv \tag{26.8}$$

To prove that we can extend S' to an isometry $S \in \mathcal{L}(\mathbb{V})$, let $v_1, v_2 \in \mathbb{V}$ such that

$$\sqrt{T^*T}v_1 = \sqrt{T^*T}v_2 \tag{26.9}$$

Then

$$\begin{aligned}
\|Tv_1 - Tv_2\| &= \|T(v_1 - v_2)\| && (26.10)\\
&= \|S'\sqrt{T^*T}(v_1 - v_2)\| \qquad \text{(by eq. 26.8)} && (26.11)\\
&= \|S'(\sqrt{T^*T}v_1 - \sqrt{T^*T}v_2)\| && (26.12)\\
&= 0 \qquad \text{(by equation 26.9)} && (26.13)
\end{aligned}$$

Hence

$$Tv_1 = Tv_2 \tag{26.14}$$

Thus S' is well defined and injective.

We must also show that S' is linear. This is left as an exercise.

Let $u \in \text{range}(\sqrt{T^*T})$. Then for some v, we have $u = \sqrt{T^*T}v$, hence

$$\begin{aligned}
\|S'u\| &= \|S'(\sqrt{T^*T}v)\| && (26.15)\\
&= \|Tv\| \text{ (equation 26.8)} && (26.16)\\
&= \|\sqrt{T^*T}v\| \text{ (equation 26.6)} && (26.17)\\
&= \|u\| && (26.18)
\end{aligned}$$

Therefore S' is an isometry.

Since S' is injective we have by theorem 8.19 that

$$\begin{aligned}
\dim \text{range}\sqrt{T^*T} &= \dim \text{null}(S') + \dim \text{range}(S') && (26.19)\\
&= \dim \text{range}(S') && (26.20)\\
&= \dim \text{range}(T) && (26.21)
\end{aligned}$$

where the last step follows from equation 26.7. From the lemma

$$\dim \text{range}(T) + \dim \text{range}T^\perp = \dim \mathbb{V} = \dim \text{range}\sqrt{T^*T} + \dim(\text{range}\sqrt{T^*T})^\perp \tag{26.22}$$

hence by equation 26.21,

$$\dim(\text{range}\sqrt{T^*T})^\perp = \dim(\text{range}T)^\perp \tag{26.23}$$

Choose orthonormal bases $E = (e_1, \dots, e_n)$ of $(\text{range}\sqrt{T^*T})^\perp$ and $F = (f_1, \dots, f_n)$ of $(\text{range}(T))^\perp$; these bases have the same length by equation 26.23.

Define

$$S'' : (\text{range}\sqrt{T^*T})^\perp \mapsto (\text{range}T)^\perp \tag{26.24}$$

$$S''(a_1e_1 + \cdots + a_ne_n) = a_1f_1 + \cdots + a_nf_n \tag{26.25}$$

Then for all $w \in (\text{range}\sqrt{T^*T})^\perp$,

$$\|S''w\|^2 = \left\|S''\sum_i a_ie_i\right\|^2 \tag{26.26}$$

$$= \left\langle \sum_i a_if_i, \sum_j a_jf_j \right\rangle \tag{26.27}$$

$$= \sum_{i,j} a_i\overline{a_j}\langle f_i, f_j\rangle \tag{26.28}$$

$$= \sum_i |a_i|^2 = \|w\|^2 \tag{26.29}$$

hence S'' is an isometry.

Define S as the operator

$$Sv = \begin{cases} S'v, & v \in \text{range}\sqrt{T^*T} \\ S''v, & v \in (\text{range}\sqrt{T^*T})^\perp \end{cases} \tag{26.30}$$

Let $v \in \mathbb{V}$ be any vector in $\mathbb{V}$. Then there exists $u \in \text{range}\sqrt{T^*T}$ and $w \in (\text{range}\sqrt{T^*T})^\perp$ such that

$$v = u + w \implies Sv = S'u + S''w \tag{26.31}$$

But by definiton of S',

$$S\sqrt{T^*T}v = S'\sqrt{T^*T}v = Tv \tag{26.32}$$

whence $T = S\sqrt{T^*T}$, which is the desired formula (equation 26.1). To prove that S

is an isometry, frrom equation 26.31,

$$\begin{aligned} \|Sv\|^2 &= \|S'u + S''w\|^2 && (26.33)\\ &= \|S'u\|^2 + \|s''w\|^2 && \text{(Pythagorean Theorem)} \quad (26.34)\\ &= \|u\|^2 + \|w\|^2 && (S' \text{ and } S'' \text{ are isometries}) \quad (26.35)\\ &= \|v\|^2 && \text{(Pythagorean Theorem)} \quad (26.36) \end{aligned}$$

Hence S is an isometry. □

Example 26.1 Find a Polar Decomposition of

$$T = \begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix} \tag{26.37}$$

The polar decomposition is $T = S\sqrt{T^*T}$ for some isometry S, which we must find.

Since T is real, then

$$T^* = T^T = \begin{pmatrix} 11 & -2 \\ -5 & 10 \end{pmatrix} \tag{26.38}$$

Hence

$$T^*T = \begin{pmatrix} 11 & -2 \\ -5 & 10 \end{pmatrix} \begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix} = \begin{pmatrix} 125 & -75 \\ -75 & 125 \end{pmatrix} \tag{26.39}$$

From example 24.3 we found four solutons for

$$M = \frac{1}{\sqrt{2}}\begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} -5 & 15 \\ 15 & -5 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 5 & -15 \\ -15 & 5 \end{pmatrix}, \frac{1}{\sqrt{2}}\begin{pmatrix} 15 & -5 \\ -5 & 15 \end{pmatrix} \tag{26.40}$$

Lets look at the first solution:

$$M = \frac{1}{\sqrt{2}} \begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix} \tag{26.41}$$

We can verify that this works:

$$M^2 = \frac{1}{\sqrt{2}} \begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix} \times \frac{1}{\sqrt{2}} \begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix} \tag{26.42}$$

$$= \frac{1}{2} \begin{pmatrix} 250 & -150 \\ -150 & 250 \end{pmatrix} \tag{26.43}$$

$$= \begin{pmatrix} 125 & -75 \\ -75 & 125 \end{pmatrix} = T^*T \tag{26.44}$$

Thus, as expected, $M = \sqrt{T^*T}$. It only remains to find S. But by the polar decomposition theorem,

$$T = S\sqrt{T^*T} = SM \implies S = TM^{-1} \tag{26.45}$$

We can calculate that

$$M^{-1} = -\frac{1}{20\sqrt{2}} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \tag{26.46}$$

Hence

$$S = TM^{-1} = -\frac{1}{20\sqrt{2}} \begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix} \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} = \frac{1}{5\sqrt{2}} \begin{pmatrix} -7 & 1 \\ -1 & -7 \end{pmatrix} \tag{26.47}$$

To verify that S is an isometry, we calculate

$$S^*S = S^TS = \frac{1}{5\sqrt{2}}\begin{pmatrix} -7 & 1 \\ -1 & -7 \end{pmatrix} \times \frac{1}{5\sqrt{2}}\begin{pmatrix} -7 & -1 \\ 1 & -7 \end{pmatrix} = I \tag{26.48}$$

Hence S is an isometry. Thus a polar decomposition of T is

$$T = S\sqrt{T^*T} = \frac{1}{5\sqrt{2}}\begin{pmatrix} -7 & 1 \\ -1 & -7 \end{pmatrix} \times \frac{1}{\sqrt{2}}\begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix} \quad \square \tag{26.49}$$

Definition 26.3 (**Singular Value**). Let $T \in \mathcal{L}(\mathbb{V})$ and define $S = \sqrt{T^*T}$. The eigenvalues of S are called the **singular values of** T.

Example 26.2 Find the singular values of

$$T = \begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix} \tag{26.50}$$

The singular values of T are the eigenvalues of M, where $M = \sqrt{T^*T}$. From the previous example,

$$M = \sqrt{T^*T} = \frac{1}{\sqrt{2}}\begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix} \tag{26.51}$$

The characteristic equation for M gives

$$0 = \det(M - \lambda I) = \begin{vmatrix} -\frac{15}{\sqrt{2}} - \lambda & \frac{5}{\sqrt{2}} \\ \frac{5}{\sqrt{2}} & -\frac{15}{\sqrt{2}} - \lambda \end{vmatrix} \tag{26.52}$$

Evaluating the determinant,

$$0 = \left(-\frac{15}{\sqrt{2}} - \lambda\right)^2 - \left(\frac{5}{\sqrt{2}}\right)^2 \tag{26.53}$$

$$= \frac{225}{2} + 15\sqrt{2}\lambda + \lambda^2 - \frac{25}{2} \tag{26.54}$$

$$= 100 + 15\sqrt{2}\lambda + \lambda^2 \tag{26.55}$$

Thus the singular values of T (the eigenvalues of $\sqrt{T^*T}$ are

$$\lambda = \frac{1}{2}\left(-15\sqrt{2} \pm \sqrt{(15\sqrt{2})^2 - 400}\right) \tag{26.56}$$

$$= \frac{1}{2}\left(-15\sqrt{2} \pm \sqrt{50}\right) \tag{26.57}$$

$$= \frac{1}{2}\left(-15\sqrt{2} \pm 5\sqrt{2}\right) \tag{26.58}$$

$$= \left\{-10\sqrt{2}, -5\sqrt{2}\right\} \quad \square \tag{26.59}$$

Lemma 26.4 Let λ be an eigenvalue of $\sqrt{M}$. Then λ^2 is an eigenvalue of M. The converse also holds, in the sense that if λ^2 is an eigenvalue of M then at least one of the square roots is also an eigenvalue of $\sqrt{M}$.

Proof. ($\Longrightarrow$) Let λ be an eigenvalue of $\sqrt{M}$ with eigenvector v. Then

$$\sqrt{M}v = \lambda v \implies \sqrt{M}\sqrt{M}v = \lambda\sqrt{M}v = \lambda^2 v \tag{26.60}$$

$$\implies Mv = \lambda^2 v \tag{26.61}$$

($\Longleftarrow$) Let λ^2 be an eigenvalue of M with eigenvector v. Then

$$0 = (M - \lambda^2 I)v = (\sqrt{M} - \lambda I)(\sqrt{M} + \lambda I)v \tag{26.62}$$

Hence either $\sqrt{\lambda}$ or $-\sqrt{\lambda}$ is an eigenvalue of $\sqrt{M}$. □

Corollary 26.5 The singular values of T are the square roots of eigenvalues of T^*T.

Theorem 26.6 (**Singular Value Decomposition.**) Let $T \in \mathcal{L}(\mathbb{V})$ with singular values $s_1, \dots, s_n$. Then there exist orthonormal bases $E = (e_1, \dots, e_n)$ and $F = (f_1, \dots, f_n)$ of $\mathbb{V}$ such that

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n \tag{26.63}$$

for every $v \in \mathbb{V}$.

Proof. Let $M = \sqrt{T^*T}$. We are given that the s_i are singular values of T; hence they are eigenvalues of M.

By the spectral theorem (theorem 21.1; you should verify that M is normal), there is an orthonormal basis $E = (e_1, \dots, e_n)$ $\mathbb{V}$ consisting solely of eigenvectors of M,

$$Me_j = s_je_j \tag{26.64}$$

Hence every $v \in \mathbb{V}$ can be expanded in this basis,

$$v = \langle v, e_1\rangle e_1 + \cdots + \langle v, e_n\rangle e_n \tag{26.65}$$

Multiplying by M,

$$Mv = M\langle v, e_1\rangle e_1 + \cdots + M\langle v, e_n\rangle e_n \tag{26.66}$$

$$= s_1\langle v, e_1\rangle e_1 + \cdots + s_n\langle v, e_n\rangle e_n \tag{26.67}$$

By the polar decomposition theorem, there is an isometry S such that

$$T = S\sqrt{T^*T} = SM \tag{26.68}$$

Multiplying equation 26.67 by S,

$$Tv = SMv = s_1\langle v, e_1\rangle Se_1 + \cdots + s_n\langle v, e_n\rangle Se_n \tag{26.69}$$

Define

$$f_j = Se_j, \qquad j = 1, \dots, n \tag{26.70}$$

Then $F = (f_1, \dots, f_n)$ is an isometry because S is an isometry, by theorem 25.2 conclusion (4), and

$$Tv = s_1\langle v, e_1\rangle f_1 + \cdots + s_n\langle v, e_n\rangle f_n \tag{26.71}$$

□

Let us consider what the singular value decomposition means in terms of matrices. We will assume the usual Euclidean inner product. Let $(e_1, \dots, e_n)$ and $(f_1, \dots, f_n)$ be orthonormal bases. Let E and F be the matrices whose columns are given by e_i and f_i. Then

$$E_{ij} = e_{ji}, \quad F_{ij} = f_{ji}, \quad E^*_{ij} = \overline{e}_{ji} \tag{26.72}$$

Define the matrix $A = FSE^*$, where S is the diagonal matrix of singular values. Then

$$A_{ij} = (FSE^*)_{ij} \tag{26.73}$$

$$= \sum_k F_{ik}(SE^*)_{kj} \tag{26.74}$$

$$= \sum_k f_{ki} \sum_\ell S_{k\ell}(E^*)_{\ell j} \tag{26.75}$$

$$= \sum_k f_{ki} \sum_\ell \delta_{k\ell} s_\ell \overline{e}_{j\ell} = \sum_k f_{ki} s_k \overline{e}_{jk} \tag{26.76}$$

Let v be any vector with components $(v_1, \dots, v_n)$ in terms of the e_k as a basis. Then

$$(Av)_\alpha = \sum_\beta A_{\alpha\beta} v_\beta \tag{26.77}$$

$$= \sum_\beta \sum_k f_{k\alpha} s_k \overline{e}_{\beta k} v_\beta \tag{26.78}$$

$$= \sum_k f_{k\alpha} s_k \sum_\beta \overline{e}_{\beta k} v_\beta \tag{26.79}$$

$$= \sum_k f_{k\alpha} s_k \langle v, e_k \rangle \tag{26.80}$$

Hence

$$Av = \sum_k f_k s_k \langle v, e_k \rangle = s_1 \langle v, e_1 \rangle f_1 + \cdots + s_n \langle v, e_n \rangle f_n \tag{26.81}$$

This proves the following theorem.

Theorem 26.7 Singular Value Decomposition for Matrices. Let A be any matrix over $\mathbb{R}$ or $\mathbb{C}$. Then there exist unitary matrices E and F such that

$$A = FSE^* \tag{26.82}$$

where S is a diagonal matrix of singular values of A, and in particular, for any orthonormal bases, $(e_1, \dots, e_n)$ and $(f_1, \dots, f_n)$, the matrices given by

$$E = \begin{pmatrix} e_1 & \cdots & e_n \end{pmatrix}, \quad F = \begin{pmatrix} f_1 & \cdots & f_n \end{pmatrix} \tag{26.83}$$

form a singular value decomposition given by equation 26.82, i.e, the columns of E and F are the basis vectors.

Example 26.3 Find a singular value decomposition of

$$T = \begin{pmatrix} 11 & -5 \\ -2 & 10 \end{pmatrix} \tag{26.84}$$

From example (26.2), the singular values of T are

$$\lambda = \left\{ -10\sqrt{2}, -5\sqrt{2} \right\} \tag{26.85}$$

and the polar decomposition of T is $T = SM$ where

$$M = \sqrt{T^*T} = \frac{1}{\sqrt{2}} \begin{pmatrix} -15 & 5 \\ 5 & -15 \end{pmatrix} \tag{26.86}$$

and

$$S = \frac{1}{5\sqrt{2}} \begin{pmatrix} -7 & 1 \\ -1 & -7 \end{pmatrix} \tag{26.87}$$

From the proof of the singular value decomposition theorem we define e_i as the orthonormal eigenvectors of $M = \sqrt{T^*T}$.

The eigenvectors of M are $(-1,1)$ and $(1,1)$, hence the orthonormal eigenvectors are

$$e_1 = \left(\frac{-1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T, \qquad e_2 = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)^T \tag{26.88}$$

We also define the orthonormal basis f_j where $f_j = Se_j$ Hence

$$f_1 = Se_1 = \frac{1}{5\sqrt{2}}\begin{pmatrix} -7 & 1 \\ -1 & -7 \end{pmatrix}\begin{pmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 4/5 \\ -3/5 \end{pmatrix} \tag{26.89}$$

$$f_2 = Se_2 = \frac{1}{5\sqrt{2}}\begin{pmatrix} -7 & 1 \\ -1 & -7 \end{pmatrix}\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \begin{pmatrix} 3/5 \\ -4/5 \end{pmatrix} \tag{26.90}$$

Hence the singular value decomposition is

$$M = \begin{pmatrix} 4/5 & -3/5 \\ 3/5 & -4/5 \end{pmatrix}\begin{pmatrix} -10\sqrt{2} & 0 \\ 0 & -5\sqrt{2} \end{pmatrix}\begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \quad \square \tag{26.91}$$

Topic 27

Generalized Eigenvectors

Motivation Let $\mathbb{V}$ be a vector space over $\mathbb{F}$ and let T be an operator on $\mathbb{V}$. Then we would like to describe T by finding subspaces of $\mathbb{V}$ in which

$$\mathbb{V} = \mathbb{U}_1 \oplus \cdots \oplus \mathbb{U}_n \tag{27.1}$$

where each U_J is invariant under T. This is possible if and only if $\mathbb{V}$ has a basis consisting only of eigenvectors (see theorem 12.8). By the same theorem, this is true if annd only if

$$\mathbb{V} = \text{null}(T - \lambda_1 I) \oplus \cdots \oplus \text{null}(T - \lambda_n I) \tag{27.2}$$

where $\lambda_1, \ldots, \lambda_n$ are distinct eigenvalues.

By the corollary to the spectral theorem (Corollary 21.3) we know that this is possible whenever T is self-adjoint. The goal is to find a way to generalize this to all operators. We do this with the concept of generalized eigenvectors.

Definition 27.1 Let $T \in \mathcal{L}(\mathbb{V})$ be an operator with eigenvalue λ. A vector $v \in \mathbb{V}$ is called a **generalized eigenvector** of T with eigenvector λ if

$$(T - \lambda I)^j v = 0 \tag{27.3}$$

for some positive integer j.

Remark 27.2 Every eigenvector is a generalized eigenvector (with $j = 1$).

Example 27.1 Find the generalized eigenvectors of

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \tag{27.4}$$

The characteristic equation is

$$p(x) = 3 - 7x + 5x^2 - x^3 = (x-1)(x-1)(x-3) \tag{27.5}$$

The eigenvector of -3 satisfies

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 3a \\ 3b \\ 3c \end{pmatrix} \tag{27.6}$$

Hence

$$a + b + c = 3a \tag{27.7}$$

$$2b + c = 3b \implies b = c \tag{27.8}$$

$$a + 2c = 3c \implies a = c \tag{27.9}$$

Hence an eigenvector of corresponding to the eigenvalue 3 is $(1, 1, 1)$. This is also a generalized eigenvector because all eigenvectors are also generalized eigenvectors.

We need to find two generalized eigenvectors corresponding to the eignevalue 1 because it has multiplicity of 2. One of them is the eigenvector corresponding to the eigenvalue 1, which satisfies

$$a + b + c = a \implies b = -c \tag{27.10}$$

$$2b + c = b \implies c = -b \tag{27.11}$$

$$a + 2c = c \implies a = -c \tag{27.12}$$

Hence an eigenvector corresponding to the eigenvalue 1 is $(-1, -1, 1)$. This gives our second generalized eigenvector.

The third generalized eigenvector satisfies

$$(M - \lambda I)^2 v = 0 \tag{27.13}$$

for $\lambda = 1$. Hence we need to find a vector in the null space of $(M - I)^2$. But

$$(M - I)^2 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 2 \\ 1 & 1 & 2 \\ 1 & 1 & 2 \end{pmatrix} \tag{27.14}$$

so a vector in the null space satisfies

$$a + b + 2c = 0 \tag{27.15}$$

If we choose $a = b = 1$ then $c = -1$ and we have $(1, 1, -1)$ which is a multiple of the second eigenvector. If we chose $a = 2$ and $b = 0$ then $c = -1$,giving $(2, 0, -1)$. There may be other possible solutions.

Summarizing, a set of generalized eigenvectors are

$$(1, 1, 1), (-1, -1, 1), (2, 0, -1) \quad \square \tag{27.16}$$

Theorem 27.3 Let $T \in \mathcal{L}(\mathbb{V})$ be an operator on $\mathbb{V}$ and let $k > 0$ be an integer. Then if $T^k v = 0$,

$$\{0\} = \text{null}(T^0) \subset \text{null}(T^1) \subset \cdots \subset \text{null}(T^k) \subset \text{null}(T^{k+1}) \subset \cdots \tag{27.17}$$

Proof. Let $v \in \text{null}(T^k)$. Then there is some v such that $T^k v = 0$, then

$$T^{k+1}v = T(T^k v) = T0 = 0 \implies \text{null}(T^k) \subset \text{null}(T^{k+1}) \tag{27.18}$$

Since this holds for all k, the result follows. $\square$

Theorem 27.4 If for some $m > 0$, $\text{null}(T^m) = \text{null}(T^{m+1})$ then

$$\text{null}(T^0) \subset \text{null}(T^1) \subset \cdots \subset \text{null}(T^m) = \text{null}(T^{m+1}) = \text{null}(T^{m+2}) = \cdots \tag{27.19}$$

In other words, once two nullspaces in equation 27.17 are equal, all successive nullspaces are equal.

Proof. Suppose that for some $m > 0$,

$$\text{null}(T^m) = \text{null}(T^{m+1}) \tag{27.20}$$

Let $k > 0$. We know by equation 27.17 that

$$\text{null}(T^{m+k}) \subset \text{null}(T^{m+k+1}) \tag{27.21}$$

Let $v \in \text{null}(T^{m+k+1})$. Then

$$0 = T^{m+k+1}v = T^{m+1}(T^k v) \tag{27.22}$$

Hence

$$T^k v \in \text{null}(T^{m+1}) = \text{null}(T^m) \tag{27.23}$$

where the second equality follows from our initial assumption (equation 27.20). Hence

$$T^{m+k}v = T^m T^k v = T^m 0 = 0 \tag{27.24}$$

hence $v \in \text{null}(T^{m+k})$; consequently,

$$\text{null}(T^{m+k+1}) \subset \text{null}(T^{m+k}) \tag{27.25}$$

comparison with equation 27.21 gives equality of the two nullspaces. □

Theorem 27.5 Let $T \in \mathcal{L}(\mathbb{V})$ be an operator over $\mathbb{V}$. Then equality holds in equation 27.17 for $k \geq \dim \mathbb{V}$:

$$\text{null}(T^{\dim \mathbb{V}}) = \text{null}(T^{\dim \mathbb{V}+1}) = \text{null}(T^{\dim \mathbb{V}+2}) = \cdots \tag{27.26}$$

Proof. Let $m = \dim \mathbb{V}$ and suppose that $\text{null}(T^{\dim \mathbb{V}}) \neq \text{null}(T^{\dim \mathbb{V}+1})$ (proof by contradiction). Then

$$\{0\} = \text{null}(T^0) \subset \text{null}(T^1) \subset \cdots \subset \text{null}(T^m) \subset \text{null}(T^{m+1}) \tag{27.27}$$

where the subsets are strict (not equality). Since the subsets are strict, the dimension of each set must increase by at least 1. Hence

$$\{0\} = \dim\text{null}(T^0) < \dim\text{null}(T^1) < \cdots < \dim\text{null}(T^m) < \dim\text{null}(T^{m+1}) \tag{27.28}$$

$$\dim\text{null}(T^{\dim\mathbb{V}+1}) = \dim\text{null}(T^{m+1}) > m+1 = \dim\mathbb{V}+1 \tag{27.29}$$

But $T^{\dim\mathbb{V}+1}$ is a subspace of $\mathbb{V}$ and cannot have dimension larger than $\mathbb{V}$. Hence this is a contradiction. Hence our assumption is false, and the theorem follows. □

Theorem 27.6 Let λ be an eigenvalue of $T \in \mathcal{L}(\mathbb{V})$. The set of generalized eigenvalues of T corresponding to eigenvalue λ equals $\text{null}((T-\lambda I)^{\dim\mathbb{V}})$.

Proof. Let $v \in \text{null}((T-\lambda I)^{\dim\mathbb{V}})$. Then by definition of generalized eigenvectors, v is a generalized eigenvector with eigenvalue λ. This proves that

$$\text{null}((T-\lambda I)^{\dim\mathbb{V}}) \subset \text{the set of generalized eigenvectors for } \lambda \tag{27.30}$$

Now assume that v is a generalized eigenvector with eigenvalue λ. Then there exists some positive integer j such that

$$(T-\lambda I)^j v = 0 \implies v \in \text{null}((T-\lambda I)^j) \tag{27.31}$$

Let $S = T - \lambda I$. Then

$$v \in \text{null}(S^j) \subset \text{null}(S^{j+1}) \subset \cdots \subset S^{\dim \mathbb{V}} \tag{27.32}$$

by the previous theorem. Hence

$$v \in \text{null}((T - \lambda I)^{\dim \mathbb{V}}) \tag{27.33}$$

Thus

$$\text{The set of generalized Eigenvectors for} \lambda \subset \text{null}((T - \lambda I)^{\dim \mathbb{V}}) \tag{27.34}$$

Hence the set of generalized eigenvectors for λ is equal to $\text{null}((T - \lambda I)^{\dim \mathbb{V}})$. □

Definition 27.7 An operator T is called **nilpotent** if for some $k > 0$, $T^k = 0$.

Corollary 27.8 Let $N \in \mathcal{L}(\mathbb{V})$ be nilpotent. Then $N^{\dim \mathbb{V}} = 0$.

Proof. Let $v \in \mathbb{V}$. Then since N is nilpotent there exists some j such that $N^j v = 0$. Hence $(N - 0I)^j v = 0$ which makes v a generalized eigenvector with eigenvalue 0, i.e., every vector in $\mathbb{V}$ is a generalized eigenvector to N with eigenvalue zero.

By theorem 27.6 the set of generalized eigenvectors of N with eigenvalue zero is $\text{null}(N^{\dim \mathbb{V}})$, i.e.,

$$\mathbb{V} = \text{null}(N^{\dim \mathbb{V}}) \tag{27.35}$$

hence for every $v \in \mathbb{V}$, we have $N^{\dim \mathbb{V}} v = 0$. Thus $N^{\dim \mathbb{V}} = 0$. □

Theorem 27.9 Let $T \in \mathcal{L}(\mathbb{V})$. Then

$$V = \text{range}(T) \supset \text{range}(T^2) \supset \cdots \supset \text{range}(T^k) \supset \text{range}(T^{k+1}) \supset \cdots \tag{27.36}$$

Proof. Let $w \in \text{range}(T^{k+1})$. Then for some $v \in \mathbb{V}$,

$$w = T^{k+1}v = T^k(Tv) \in \text{range}\left(T^k\right) \implies \text{range}(T^{k+1}) \subset \text{range}(T^k) \tag{27.37}$$

□

The sequence terminates when $k = \dim(\mathbb{V})$, as the following shows.

Theorem 27.10 Let $\mathbb{V}$ be a vector space and $T \in \mathcal{L}(\mathbb{V})$ an operator over $\mathbb{V}$. Then

$$\text{range}\left(T^{\dim \mathbb{V}}\right) = \text{range}\left(T^{\dim \mathbb{V}+1}\right) = \text{range}\left(T^{\dim \mathbb{V}+2}\right) = \cdots \tag{27.38}$$

Proof. (exercise) □

Topic 28

The Characteristic Polynomial

Definition 28.1 The **multiplicity of an eigenvalue** λ is the dimension of the subspace of generalized eigenvectors corresponding to λ, i.e.,

$$\text{multiplicity}(\lambda) = \dim \text{null}((T - \lambda I)^{\dim \mathbb{V}}) \tag{28.1}$$

The following theorem shows that if the matrix of T is upper triangular, the multiplicity is equal to the number of times that λ occurs on the main diagonal.

Theorem 28.2 Let $\mathbb{V}$ be a vector field over $\mathbb{F}$; let $T \in \mathcal{L}(\mathbb{V})$ an operator over $\mathbb{V}$; and let $\lambda \in \mathbb{F}$. If $B = (v_1, \ldots, v_n)$ is a basis of $\mathbb{V}$ such that $M = \mathcal{M}(T, B)$ is upper triangular, λ appears on the diagonal of M

$$\dim \text{null}((T - \lambda I)^{\dim \mathbb{V}}) \tag{28.2}$$

times.

Proof. Consider first the case with $\lambda = 0$. To prove the general case, replace T with

$T' = T - \lambda I$ in what follows.

Define $n = \dim \mathbb{V}$.

For $n = 1$ the result of course holds because M is 1×1.

Let $n > 1$ and assume the result holds for $n - 1$.

Suppose that (with respect to B), M is upper triangular; define λ_i as the diagonal elements, so that,

$$M = \begin{pmatrix} \lambda_1 & * & \cdots & * \\ 0 & \ddots & & \vdots \\ \vdots & & \lambda_{n-1} & * \\ 0 & \cdots & & \lambda_n \end{pmatrix} \tag{28.3}$$

Let $U = \text{span}\,(v_1, \dots, v_{n-1})$. By theorem 12.2 U is invariant under T. Furthermore,

$$M' = \mathcal{M}(T|_U, U) = \begin{pmatrix} \lambda_1 & \cdots & * \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{n-1} \end{pmatrix} \tag{28.4}$$

By the inductive hypothesis, 0 appears on the diagonal of M'

$$\text{Num. of zeros on diag. of } M' = \dim \text{null}((T|_U)^{\dim U}) = \dim \text{null}((T|_U)^{n-1}) \tag{28.5}$$

times, because $\dim U = n-1$. Furthermore, also because $\dim U = n-1$, by theorem 27.5 we have

$$\text{null}((T|_U)^{n-1}) = \text{null}((T|_U)^n) \tag{28.6}$$

Hence (combining the last two equations), the number of zeros on the diagonal of M'

is

$$\text{Number of zeroes on diagonal of } M' = \dim \text{null}((T|_U)^n) \tag{28.7}$$

We consider two cases: $\lambda_n = 0$ and $\lambda_n \neq 0$.

Case 1: $\lambda_n \neq 0$ By equation 28.3

$$\mathcal{M}(T^n) = \mathcal{M}(T)^n = M^n = \begin{pmatrix} \lambda_1^n & * & \cdots & * \\ 0 & \ddots & & \vdots \\ \vdots & & \lambda_{n-1}^n & * \\ 0 & \cdots & & \lambda_n^n \end{pmatrix} \tag{28.8}$$

Recall that v_n is the n^{th} basis vector; hence

$$T^n v_n = u + \lambda_n^n v_n \tag{28.9}$$

where $u \in U$; (to see this, let m_i be the ith row vector of M^n and v_{ni} the ith component of v_n; then

$$M^n v_n = \begin{pmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{pmatrix} \begin{pmatrix} v_{n1} \\ v_{n2} \\ \vdots \\ v_{nn} \end{pmatrix} = m_1 v_{n,1} + \cdots + m_{n-1} v_{n,n-1} + m_n v_{n,n} \tag{28.10}$$

The first $n-1$ terms give u; the last term is $m_n v_{n,n} = \lambda_n^n v_n$.)

Now let $v \in \text{null}(T^n)$. Then (since B is a basis),

$$v = u' + a v_n \tag{28.11}$$

where $u' \in U$ and $a \in \mathbb{F}$. Since $v \in \text{null}(T^n)$,

$$0 = T^n v = T^n u' + aT^n v_n = T^n u' + au + a\lambda_n^n v_n \tag{28.12}$$

where we have used equation 28.9 in the last step. Since U is invariant under T, and hence under T^n, the first two terms are in U, while the last term is in span (v_n) which is not in U. Since the sum is zero, each term must be zero. Hence

$$a\lambda_n^n = 0 \tag{28.13}$$

Since we have assumed (case 1) that $\lambda_n \neq 0$ this measn $a = 0$. Hence $v \in U$. But we chose v as any element of null(T^n). Hence

$$\text{null}(T^n) \subset U \tag{28.14}$$

and therefore

$$\text{null}(T^n) = \text{null}((T|_U)^n) \tag{28.15}$$

(otherwise there would be some element of null(T^n) that was not in U.)

Hence by equation 28.7

$$\text{Number of zeroes on diagonal of } M' = \dim \text{null}((T|_U)^n) = \dim \text{null}(T^n) \tag{28.16}$$

which is the objective we want to prove (eq. 28.2).

Case 2: $\lambda_n = 0$. By theorem 7.12

$$\dim(U + \text{null}(T^n)) = \dim(U) + \dim(\text{null}(T^n)) - \dim(U \cap \text{null}(T^n)) \tag{28.17}$$

hence, since

$$\dim U = n - 1 \tag{28.18}$$

and

$$\dim(U \cap \text{null}(T^n)) = \dim(T|_U)^n \tag{28.19}$$

equation 28.17 becomes

$$\dim(\text{null}(T^n)) = \dim(U + \text{null}(T^n)) + \dim(U \cap \text{null}(T^n)) - \dim(U) \tag{28.20}$$

$$= \dim(T|_U)^n + \dim(U + \text{null}(T^n)) - (n-1) \tag{28.21}$$

Consider any vector of the form

$$w = u - v_n \tag{28.22}$$

where $u \in U$. Then $w \notin U$. We want to choose u such that $w \in \text{null}(T^n)$. (This requires

$$T^n w = T^n(u - v_n) = T^n u - T^n v_n = 0 \tag{28.23}$$

$$\implies T^n u = T^n v_n \tag{28.24}$$

$$\implies T^n v_n \in \text{range}\left((T|_U)^n\right) \tag{28.25}$$

But because M is upper triangular (see equation 28.9, for example)

$$Tv_n = u + \lambda_n v_n = u \tag{28.26}$$

where $u \in U$; the second equality follows because we are assuming $\lambda_n = 0$. Hence

$$Tv_n \in U \tag{28.27}$$

and consequently

$$T^n v_n = T^{n-1}(Tv_n) = T^{n-1}u \in \text{range}\left((T|_U)^{n-1}\right) \in \text{range}\left((T|_U)^n\right) \tag{28.28}$$

where the last step comes from theorem 27.10. Hence it is possible to find such a u.)

By choosing u in this manner we have $w \notin U$ but $w \in \text{null}(T^n)$.

Hence

$$n = \dim(\mathbb{V}) \geq \dim(U + \text{null}(T^n)) > \dim U = n - 1 \tag{28.29}$$

$$\implies \dim(U + \text{null}(T^n)) = n \tag{28.30}$$

Substituting equation 28.30 into equation 28.21,

$$\dim(\text{null}(T^n)) = \dim(T|_U)^n + \dim(U + \text{null}(T^n)) - (n-1) \tag{28.31}$$

$$= \dim(T|_U)^n + n - (n-1) \tag{28.32}$$

$$= \dim(T|_U)^n + 1 \tag{28.33}$$

Using the last result in equation 28.7 gives, since $\lambda_n = 0$

$$\text{Number of zeroes on diagonal of } M = \tag{28.34}$$

$$1 + \text{Number of zeroes on diagonal of } M' = \tag{28.35}$$

$$1 + \dim \text{null}((T|_U)^n) = \dim(\text{null}(T^n)) \tag{28.36}$$

which is the required result. □

Theorem 28.3 Let $\mathbb{V}$ be a complex vectors space and suppose that $T \in \mathcal{L}(\mathbb{V})$. Then the sum of the multiplicities of all the eigenvalues of T equals $\dim \mathbb{V}$.

Proof. The previous theorem showed that the multiplicity of each eigenvalue is the number of times that λ appears on the diagonal. The total number of diagonal elements is $\dim \mathbb{V}$. □

Definition 28.4 Characteristic Polynomial. Let $\mathbb{V}$ be a complex vector space and let $T \in \mathcal{L}(\mathbb{V})$ have distinct eigenvalues $\lambda_1, \dots, \lambda_m$ with multiplicities $d_1, \dots, d_m$. Then the characteristic polynomila of T is given by

$$p(z) = (z - \lambda_1)^{d_1}(z - \lambda_2)^{d_2} \cdots (z - \lambda_m)^{d_m} \tag{28.37}$$

Theorem 28.5 Cayley-Hamilton Theorem. An operator satisfies its own characteristic polynomial, i.e., if $T \in \mathcal{L}((\mathbb{V})$ is an operator on a complex vector space $\mathbb{V}$ with characteristic polynomial $p(z)$ then

$$p(T) = 0 \tag{28.38}$$

Proof. Let $B = (v_1, \dots, v_n)$ be a basis such that $\mathcal{M}(T, B)$ is upper triangular.

We need to prove that $p(T) = 0$. This is equivalent to proving that

$$p(T)v = 0 \tag{28.39}$$

for all v. Since any v can be expanded as a sum of the basis vectors, this is also equivalent to proving that

$$p(T)v_j = 0 \tag{28.40}$$

for $j = 1, 2, \dots, n$. Since

$$p(T) = (T - \lambda_1)^{d_1}(T - \lambda_2)^{d_2} \cdots (T - \lambda_m)^{d_m} \tag{28.41}$$

we only need to prove that

$$0 = (T - \lambda_1)(P - \lambda_2) \cdots (z - \lambda_m)v_j \tag{28.42}$$

because equation 28.42 implies equation 28.40.

The proof is by induction. For $j = 1$, there is only one eigenvector, which satisfies

$$Tv_1 = \lambda_1 v_1 \implies (T - \lambda_1)v_1 = 0 \tag{28.43}$$

which is precisely equation 28.42 for $j = 1$.

For the inductive step, suppose that $1 < j \leq n$ and that equation 28.42 holds for $1, 2, \dots, j-1$, i.e.,

$$(T - \lambda_1 I)v_1 = 0 \tag{28.44}$$

$$(T - \lambda_1 I)(T - \lambda_2 I)v_2 = 0 \tag{28.45}$$

$$\vdots$$

$$(T - \lambda_1 I) \cdots (T - \lambda_{j-1} I)v_{j-1} = 0 \tag{28.46}$$

Let $M_j = \mathcal{M}(T|_{U_j})$ where $U_j = (v_1, \dots, v_j)$. Then the bottom right diagonal element of $M - \lambda_j I$ is zero, hence

$$(T|_{U_j} - \lambda_j I)v_j \in \text{span}\,(v_1, \dots, v_{j-1}) \tag{28.47}$$

i.e.,

$$v_j = \sum_{k=1}^{j-1} a_k v_k \tag{28.48}$$

$$\prod_{k=1}^{j}(T - \lambda_k I)v_j = \prod_{k=1}^{j}(T - \lambda_k I)\sum_{\ell=1}^{j-1} a_\ell v_\ell = 0 \tag{28.49}$$

which proves equation 28.42. □

Example 28.1 Verify the Cayley-Hamilton Theorem for

$$M = \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} \tag{28.50}$$

The characteristic equation is

$$p(x) = \begin{vmatrix} -1-x & 3 \\ 2 & -4-x \end{vmatrix} \tag{28.51}$$

$$= (1+x)(4+x) - 6 \tag{28.52}$$

$$= x^2 + 5x - 2 \tag{28.53}$$

The Cayley-Hamilton Theorem says that

$$P(M) = M^2 + 5M - 2I = 0 \tag{28.54}$$

But

$$P(M) = \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} + 5 \begin{pmatrix} -1 & 3 \\ 2 & -4 \end{pmatrix} - 2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{28.55}$$

$$= \begin{pmatrix} 7 & -15 \\ -10 & 22 \end{pmatrix} + \begin{pmatrix} -5 & 15 \\ 10 & -20 \end{pmatrix} - \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \tag{28.56}$$

$$= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \square \tag{28.57}$$

Topic 29

The Jordan Form

Theorem 29.1 Let $\mathbb{V}$ be a vector space over $\mathbb{F}$, $T \in \mathcal{L}(\mathbb{V})$ and p is a polynomial over $\mathbb{F}$. Then $\text{null}(p(T))$ is invariant under T.

Proof. Let $v \in \text{null}(p(T))$. Then

$$p(T)v = 0 \tag{29.1}$$

and since T commutes with $p(T)$ (because T commutes with itself),

$$p(T)Tv = Tp(T)v = T(0) = 0 \tag{29.2}$$

Therefore

$$Tv \in \text{null}(p(T)) \tag{29.3}$$

hence $\text{null}(p(T))$ is invariant under T. □

Theorem 29.2 Let $\mathbb{V}$ be a complex vector space, and let T be an operator over $\mathbb{V}$, with distinct eigenvalues $\lambda_1, \dots, \lambda_m$ and corresponding eigenspaces (subspaces spanned by the generalized eigenvectors) $U_1, \dots, U_m$. Then

(1) Each U_j is invariant under T_j

(2) Each $(T - \lambda_j I)|_{U_j}$ is nilpotent.

(3) $V = U_1 \oplus \cdots \oplus \mathbb{U}_m$

Proof. Proof of (1) By theorem 27.6

$$U_j = \text{null}((T - \lambda_j I)^{\dim \mathbb{V}}) \tag{29.4}$$

Let

$$p(z) = (z - \lambda_j)^{\dim \mathbb{V}} \tag{29.5}$$

By theorem 29.1 $\text{null}(p(T)) = U_j$ is invariant under T.

Proof of (2) This follows because corresponding to the λ_j there are generalized eigenvectors $v_{j1}, \dots, v_{jp}$ such that

$$(T - \lambda_j I)^q v_{jq} = 0 \tag{29.6}$$

Hence $T - \lambda_j I$ is nilpotent.

Proof of (3). The multiplicity of each λ_j is $\dim(U_j)$ (def. 28.1) and the sum of the multiplicities is $\dim(\mathbb{V})$ (theorem 28.3),

$$\dim \mathbb{V} = \dim U_1 + \cdots + \dim U_m \tag{29.7}$$

Define

$$U = U_1 + \cdots + U_m \tag{29.8}$$

By (1) each of the U_i is invariant under T; hence U is invariant under T.

Define $S = T|_U$. Every generalized eigenvector of T is a generalized eigenvector of

S, with the same eigenvalues; and the the eigenvalues have the same multiplicities. Hence

$$\dim U = \dim U_1 + \cdots + \dim U_m = \dim \mathbb{V} \tag{29.9}$$

But U is a subspace of $\mathbb{V}$ it must have a smaller dimension, or have the same dimension only if it is equal to $\mathbb{V}$. Hence

$$\mathbb{V} = U = U_1 + \cdots + U_m \tag{29.10}$$

which is the desired result. □

Theorem 29.3 Let $\mathbb{V}$ be a complex vector space and let $T \in \mathcal{L}(\mathbb{V})$ be an operator on $\mathbb{V}$. Then there is a basis consisting of generalized eigenvectors of T.

Proof. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of T.

Define U_j as the generalized eigenspace corresponding to λ_j.

Let $v_{j1}, \ldots, v_{jp}$ be the generalized eigenvectors corresponding to λ_j. Then by definition of U_j, it is the span of the these eigenvectors. Hence they form a basis of U_j.

Join together all of the bases of each U_j formed in this way. The result is a basis of $\mathbb{V}$ which consists solely of generalized eigenvectors of U_j. □

Theorem 29.4 Let $\mathbb{V}$ be a vector space and $N \in \mathcal{L}(\mathbb{V})$ be nilpotent. Then there is a basis B of $\mathbb{V}$ such that $\mathcal{M}(N, B)$ is **strictly** upper triangular (i.e., all the diagonal elements are zero).

Proof. Since the result is follows trivially if N is the zero operator (hence M is the zero matrix) we assume that N is not the zero operator.

Since N is nilpotent then for some m, $N^m = 0$. For this m, $\text{null}(N^m) = \mathbb{V}$, because $N^m v = 0$ for every $v \in \mathbb{V}$.

Choose a basis of $\mathbb{V} = \text{null}(N^m)$ as follows: choose any basis of $\text{null}(N)$. Then extend it to a basis of $\text{null}(N^2)$, then extend the result to a basis of $\text{null}(N^3)$, ... The result is a basis B of $\mathbb{V}$.

Form $\mathcal{M}(N, B)$. The columns are basis vectors of $\mathbb{V}$. The first columns are in $\text{null}(N)$, followed by vectors in $\text{null}(N^2)$, etc.

Since the first column v_1 is in $\text{null}(N)$, $Nv_1 = 0$. Hence $v_1 = 0$, i.e., the first column is entirely zeros. The same argument holds for any other vector in the first set of columns.

The next set of columns are in $\text{null}(N^2)$. Let v_j be any such column.

$$0 = N^2 v_j = N(Nv_j) \implies Nv_j \in \text{null}(N) \tag{29.11}$$

Hence Nv_j is a linear combination of the columns to the left of it; thus the diagonal element must be zero.

Repeat the argument for each succeeding column. □

Theorem 29.5 Let $\mathbb{V}$ be a complex vector space and $T \in \mathcal{L}V$. Let $\lambda_1, \dots, \lambda_m$ be the distinct eigenvalues of T. Then there is a basis B of $\mathbb{V}$ such that

$$\mathcal{M}(T, B) = \text{diagonal}(A_1, \dots, A_m) \tag{29.12}$$

where each A_j is upper triangular with λ_j on the diagonal.

Proof. For each λ_j, let U_j be the subspace spanned by the corresponding generalized eigenvectors.

By theorem 29.2, $(T - \lambda_j I)|_{U_j}$ is nilpotent.

For each j we can choose a basis B_j of U_j such that (by theorem 29.4)

$$\mathcal{M}(T|_{U_j} - \lambda_j I, B_i) = \begin{pmatrix} 0 & & * \\ & \ddots & \\ 0 & & 0 \end{pmatrix} \tag{29.13}$$

Define $A_j = \mathcal{M}(T|_{U_j})$. Then (add $\lambda_j I$ to the above equation),

$$A_j = \begin{pmatrix} \lambda_j & & * \\ & \ddots & \\ 0 & & \lambda_j \end{pmatrix} \tag{29.14}$$

Now form a basis of $\mathbb{V}$ by combining the bases of U_j; the result is a matrix

$$\mathcal{M}(T, B) = \begin{pmatrix} A_1 & & 0 \\ & \ddots & \\ 0 & & A_m \end{pmatrix} \tag{29.15}$$

□

Definition 29.6 Let $\mathbb{V}$ be a vector space and $T \in \mathcal{L}(\mathbb{V})$. A basis B of $\mathbb{V}$ is called a **Jordan Basis** for T if the matrix of T is block diagonal,

$$\mathcal{M}(T, B) = \begin{pmatrix} A_1 & & 0 \\ & \vdots & \\ 0 & & A_m \end{pmatrix} \tag{29.16}$$

with each block upper-triangular with λ_j on the main diagonal and 1's on the super-diagonal,

$$A_j = \begin{pmatrix} \lambda_j & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_j \end{pmatrix} \tag{29.17}$$

Theorem 29.7 Let $\mathbb{V}$ be a complex vector space. If $T \in \mathcal{L}(\mathbb{V})$ then there is a basis of $\mathbb{V}$ that is a Jordan Basis of $\mathbb{V}$.

Lemma 29.8 Let $\mathbb{V}$ be a vector space and let $N \in \mathcal{L}(\mathbb{V})$ be nilpotent. Then there exist vectors $v_1, \dots, v_k \in \mathbb{V}$ such that

(1) The following is a basis of $\mathbb{V}$:

$$(v_1, Nv_1, \dots, N^{m(v_1)}v_1, \dots, v_k, Nv_k, \dots, N^{m(v_k)}v_k) \tag{29.18}$$

(2) The following is a basis of null(N):

$$(N^{m(v_1)}v_1, \dots, N^{m(v_k)}v_k) \tag{29.19}$$

Proof. (See Axler.) □

Proof. (Theorem 29.7). Let $N \in \mathcal{L}(\mathbb{V})$ be nilpotent. Construct the vectors $v_1, \dots, v_k$ as in the lemma.

Observe that N sends the first vector in the list

$$B_j = (N^{m(v_j)} v_j, \dots, N v_j, v_j) \tag{29.20}$$

to zero and each subsequent vector in the list to the vector to its left.

By reversing each B_j defined in equation 29.20 and forming the list

$$B' = (\text{reverse}(B_1), \dots, \text{reverse}(B_k)) \tag{29.21}$$

we obtain the basis in part (1) of the lemma. Hence

$$B = (B_1, \dots, B_k) \tag{29.22}$$

is also a basis of $\mathbb{V}$, but with the property that $\mathcal{M}(N, B)$ is block diagonal with each block of the form

$$\begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & 0 \end{pmatrix} \tag{29.23}$$

This proves the theorem when T is nilpotent.

Now suppose T is any operator over $\mathbb{V}$, with distinct eigenvalues $\lambda_1, \dots, \lambda_m$ and corresponding generalized eigenspaces $U_1, \dots, U_m$. Then

$$V = U_1 \oplus \cdots \oplus U_m \tag{29.24}$$

where

$$S_j = (T - \lambda_j I)|_{U_j} \tag{29.25}$$

is nilpotent, by theorem 29.2 (2).

Now apply the argument in the previous paragraph: to each nilpotent operator S_j there is a basis of U_j such that $\mathcal{M}(S_j, B_j)$ has the form of equation 29.23. By equation 29.25, each block has the form desired (by adding $\lambda_j I$ to the form in equaiton 29.23). This gives a Jordan Block, proving the theorem. □

Example 29.1 Find the Jordan form of

$$M = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 1 \\ 1 & 0 & 2 \end{pmatrix} \tag{29.26}$$

From example 27.1 the eigenvalues of M are 1 (with multiplicity 2) and 3 (with multiplicity 1). Hence a Jordan form is

$$J = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \square \tag{29.27}$$

Thumbs up, you're finished with Linear Algebra!