# Abstract Linear Algebra
# Math 4373 Fall 2000

Noel Brady

August, 2000

# Contents

**1 Basic Theory** **2**
1.1 Vector Spaces . . . 3
1.2 Linear Transformations and Coordinates . . . 11
1.3 Linear functionals and duality . . . 17
1.4 Determinants . . . 20

**2 Structure of Linear Operators** **24**
2.1 Eigenvalues and Eigenvectors, Diagonalization . . . 25
2.2 Annihilating Polynomials, Hamilton-Cayley Theorem, Primary Decomposition Theorem . . . 29
2.3 Jordan and Rational Canonical Forms . . . 38

**3 Inner Product Spaces** **43**
3.1 Inner Product Spaces . . . 44
3.2 Diagonalization and Spectral Theorem . . . 49

**4 Miscellaneous Topics** **54**
4.1 Introduction to Linear Groups and Geometry . . . 55

# Chapter 1

# Basic Theory

This chapter gives a review of the basic notions of vector space and linear transformation that you have encountered in Math 3333. There is a somewhat more abstract/general perspective this time around however. We work with vector spaces over an arbitrary field rather than just the real or complex number field.

Most of the topics should seem familiar to you if you recall your Math 3333 course notes. Topics include: vector spaces, subspaces, direct sums, bases, dimensions, coordinates, linear transformations, matrices, change of bases and similar matrices, rank, nullity, linear functionals, dual spaces, transposes and adjoints, determinants and matrix inverses.

## 1.1 Vector Spaces

**Definition 1.1.1.** A *field* is a set $K$ together with two operations called *multiplication* (denoted by juxtaposition) and *addition* (denoted by $+$) which satisfy the following axioms:

1. $x + y = y + x$ for all $x, y \in K$
2. $x + (y + z) = (x + y) + z$ for all $x, y, z \in K$
3. There exists a unique *zero element* $0 \in K$ such that $0 + x = x$ for all $x \in K$
4. For every $x \in K$ there exists a *negative* $-x \in K$ such that $x + (-x) = 0$
5. $xy = yx$ for all $x, y \in K$
6. $x(yz) = (xy)z$ for all $x, y, z \in K$
7. There is a unique *unit element* $1 \in K \setminus \{0\}$ such that $1x = x$ for all $x \in K$
8. For every $x \in K \setminus \{0\}$ there is an *inverse* $x^{-1} \in K$ such that $xx^{-1} = 1$
9. $x(y + z) = xy + xz$ for all $x, y, z \in K$.

**Examples 1.1.2.** Examples of fields include $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Q}(\sqrt{2})$, $\mathbb{Z}_p$ ($p$ prime) and the field of *constructible numbers*. We have some inclusions

$$\mathbb{Q} \subset \mathbb{Q}(\sqrt{2}) \subset \text{ constructible numbers } \subset \mathbb{R} \subset \mathbb{C}$$

The classical question of *duplicating the cube* becomes the question of whether or not $\sqrt[3]{2}$ belongs to the field of constructible numbers.

Non-examples include $\mathbb{Z}$, $\mathbb{Z}_n$ ($n$ not prime), $\mathbb{N}$, $\mathbb{H}$, $\mathbb{Z}(\sqrt{2})$.

**Remark 1.1.3.** Which field has the most elements: $\mathbb{Q}$ or $\mathbb{R}$?

To answer this question we must think about the cardinality of sets.

- Motivation: early counting methods.
- Say that sets $X$ and $Y$ have the *same cardinality* if there exists a bijection $X \to Y$. Denote cardinality of $X$ by $Card(X)$.
- Say that $Card(X) \preceq Card(Y)$ if there is an injective map $X \to Y$.
- Schroeder-Bernstein Theorem: If $Card(X) \preceq Card(Y)$ and $Card(Y) \preceq Card(X)$, then $Card(X) = Card(Y)$. (Dynamical systems style proof)
- Diagonal counting techniques: $\mathbb{Z}$, $\mathbb{N} \times \mathbb{N}$ and $\mathbb{Q}$ all have same cardinality as $\mathbb{N}$.
- $\mathbb{R}$ has same cardinality as $(0, 1)$ and so has strictly larger cardinality than $\mathbb{Q}$.
- The power set, $2^X$, has strictly greater cardinality than the original set $X$. (pretty Cantor argument by contradiction).
- Show that $(0, 1)$ and $2^{\mathbb{N}}$ have the same cardinality (dyadic expansion of reals).

- $\mathbb{R}$ has the same cardinality as $\mathbb{R} \times \mathbb{R}$ (via $2^{\mathbb{N}}$ and dyadic expansion of reals) and hence as $\mathbb{C}$.

**Definition 1.1.4.** Let $K$ be a field. A *$K$-vector space* is a set $V$ together with an operation

$$V \times V \;\to\; V \;:(v,w)\;\mapsto\; v+w$$

called *vector addition* such that $(V,+)$ is an abelian group, and an operation

$$K \times V \;\to\; V \;:(k,v)\;\mapsto\; kv$$

called *scalar multiplication* satisfying

1. $k(u+v) = ku + kv$ for all $k \in K$ and all $u, v \in V$
2. $(k+l)u = ku + lu$ for all $k, l \in K$ and all $u \in V$
3. $k(lu) = (kl)u$ for all $k, l \in K$ and all $u \in V$
4. $1u = u$ for all $u \in V$ where $1 \in K$ is the unit element.

Elements of $V$ are called *vectors*, and elements of $K$ are called *scalars*.

**Examples 1.1.5.** Some examples of vector spaces. Throughout these examples $K$ is any field.

1. $\mathbb{R}^3$ from Calc III
2. More generally the space of $n$-tuples $K^n$ with addition defined by

$$(k_1, \ldots, k_n) \;+\; (l_1, \ldots, l_n) \;=\; (k_1 + l_1, \ldots, k_n + l_n)$$

and scalar multiplication defined by

$$l(k_1, \ldots, k_n) \;=\; (lk_1, \ldots, lk_n)$$

is a $K$-vector space.

3. $K^{m\times n}$ under usual addition and scalar multiplication of $(m \times n)$-matrices.
4. Let $S$ be any set. Then the set

$$K^S \;=\; \{f \mid f : S \to K \text{ is a function }\}$$

with addition defined (for $f, g \in K^S$) by

$$(f+g)(s) \;=\; f(s) + g(s) \text{ for all } s \in S$$

and scalar multiplication defined (for $f \in K^S$ and $k \in K$) by

$$(kf)(s) \;=\; kf(s) \text{ for all } s \in S$$

5. The set of continuous functions $\mathcal{C}([a,b])$ on the closed interval $[a,b] \subset \mathbb{R}$ under usual definition of addition and scalar multiplication of functions.

6. $K[x]$ the set of polynomials with coefficients in $K$, under usual addition and scalar multiplication of polynomials.

7. If $K$ and $L$ are fields and $K \subset L$ then $L$ may be thought of as $K$-vector space.

**Definition 1.1.6.** Let $v, u_1, \dots u_n$ be elements of a $K$-vector space. Say that $v$ is a *linear combination of* $u_1, \dots u_n$ if there are scalars $k_1, \dots k_n \in K$ such that

$$v \;=\; k_1 u_1 + \cdots k_n u_n \;=\; \sum_{i=1}^{n} k_i u_i$$

**Examples 1.1.7.** Linear combination coefficients are fairly easy to work out.

1. Show that $(1,0)$ is a linear combination of $(1,1)$ and $(2,1)$ in $\mathbb{R}^2$.

2. Show that every function $\mathbb{R} \to \mathbb{R}$ is a linear combination of an *odd* function and an *even* function.

**Definition 1.1.8.** Let $V$ be a $K$-vector space. A subset $U \subset V$ is called a *$K$-vector subspace of* $V$ if

1. $U$ is nonempty

2. $U$ is *closed* under addition: $u_1, u_2 \in U$ implies $u_1 + u_2 \in U$

3. $U$ is *closed* under scalar multiplication: $u \in U$ and $k \in K$ implies $ku \in U$.

Note that $U$ is itself a $K$-vector space.

**Examples 1.1.9.** There are many naturally occurring examples of subspaces.

1. Let $\vec{n} \in \mathbb{R}^3$, then
$$\langle \vec{n} \rangle^{\perp} \;=\; \{\vec{u} \in \mathbb{R}^3 \,|\, \vec{u} \cdot \vec{n} = 0\,\}$$
is a subspace of $\mathbb{R}^3$.

2. $\mathcal{C}^{\infty}(\mathbb{R}) \subset \cdots \subset \mathcal{C}^{1}(\mathbb{R}) \subset \mathcal{C}(\mathbb{R}) \subset \mathbb{R}^{\mathbb{R}}$

3. $\{\,\text{solutions to } \frac{d^2y}{dx^2} + \lambda^2 y = 0\} \subset \mathcal{C}^{\infty}(\mathbb{R})$

4. The set of all *odd* (resp. *even*) functions $\mathbb{R} \to \mathbb{R}$.

5. Let $u_1, \dots, u_n \in V$ where $V$ is a $K$-vector space. Then
$$\mathcal{S}(u_1, \dots, u_n) \;=\; \{\sum_{i=1}^{n} k_i u_i \,|\, k_i \in K\}$$
is called the subspace of $V$ *spanned* (or *generated*) by $u_1, \dots, u_n$. It is just the subspace of all linear combinations of $u_1, \dots, u_n$.

6. The collection of all symmetric $(n \times n)$-matrices (resp. skew-symmetric matrices) is a subspace of $K^{n \times n}$.

7. The collections of all self-adjoint (hermitian) matrices is a real subspace of $\mathbb{C}^{n\times n}$ but is **not** a complex subspace.

8. The solution set of the homogenous system $Ax = 0$ where $A$ is an $(m \times n)$-matrix, $x$ is an $(n \times 1)$-vector and $0$ is an $(m \times 1)$-vector.

9. The intersection of a family of subspaces of $V$ is again a subspace of $V$.

10. The space of polynomials of degree at most $n$ is a subspace of $K[x]$.

11. If $W_1, W_2$ are subspaces of the vector space $V$, then their *sum*, defined by

$$W_1 + W_2 = \{w_1 + w_2 \,|\, w_i \in W_i \quad (i = 1, 2)\}$$

is a subspace of $V$.

**Definition 1.1.10.** Let $V$ be a $K$-vector space. The elements $u_1, \dots, u_n \in V$ are said to be *linearly independent* if

$$k_1u_1 + \cdots + k_nu_n = 0$$

implies

$$k_1 = 0, \dots, k_n = 0\,.$$

We say that $u_1, \dots, u_n$ are *linearly dependent* if they are not linearly independent. Equivalently, we can say explicitly what it means for the collection of vectors $u_1, \dots, u_n$ to be linearly dependent. Namely, there exists scalars $k_1, \dots, k_n$, not all zero, such that

$$\sum_{i=1}^{n} k_iu_i = 0\,.$$

**Examples 1.1.11.** Here are some linearly (in)dependent collections. Can you say which is which?

1. $\langle 1, 1\rangle$ and $\langle 2, 1\rangle$

2. $\langle 1, 1\rangle$, $\langle 1, 0\rangle$ and $\langle 2, 1\rangle$

3. $\cos(4x)$ and $\sin(4x)$ in $\mathcal{C}^\infty(\mathbb{R})$

4. A non-zero odd function and a non-zero even function in $\mathbb{R}^{\mathbb{R}}$.

**Definition 1.1.12.** Say that the set $\{u_1, \dots, u_n\}$ *generates* the vector space $V$ if $\mathcal{S}(u_1, \dots, u_n) = V$.

**Definition 1.1.13.** We say that the set $\{u_1, \dots, u_n\}$ is *a basis* for the vector space $V$ if

1. $u_1, \dots, u_n$ are linearly independent

2. $u_1, \dots, u_n$ generate $V$

**Examples 1.1.14.** Find simple bases for the following spaces.

1. $K^n$

2. $K^{n\times n}$

3. Solutions to the equation $\frac{d^2y}{dx^2} + \lambda^2 y = 0$ where $\lambda > 0$ is real.

4. The space of polynomials of degree at most $n$ over $K$.

5. The space of symmetric $(n \times n)$-matrices.

6. The space of skew-symmetric $(n \times n)$-matrices.

**Lemma 1.1.15.** *Let $\{u_1, \ldots, u_n\}$ be a basis for the $K$-vector space $V$, and let $u \in V$. Then there exist uniquely determined scalars $\alpha_1, \ldots, \alpha_n \in K$ such that*

$$u \,=\, \alpha_1 u_1 + \cdots + \alpha_n u_n$$

*Proof.* There are two claims: existence and uniqueness. Their proofs involve different properties of a basis. Existence of the $\alpha_i$ follows form the fact that the $u_i$ generate $V$ and $u \in V$. Now for uniqueness. Suppose that there are scalars $\beta_i \in K$ such that

$$\alpha_1 u_1 + \cdots \alpha_n u_n \,=\, \beta_1 u_1 + \cdots \beta_n u_n\,.$$

Then we get

$$(\alpha_1 - \beta_1)u_1 + \cdots + (\alpha_n - \beta_n)u_n = 0$$

and so, by linear independence of the $u_i$, we conclude that $(\alpha_i - \beta_i) = 0$ for all $i$. That is $\alpha_i = \beta_i$ for all $i$, and so uniqueness is established. □

**Definition 1.1.16.** Let $\{u_1, \ldots, u_n\}$ be a basis for the vector space $V$, and let $u \in V$. The $n$-tuple of scalars $(\alpha_1, \ldots, \alpha_n)$ with the property that $\sum_{i=1}^{n} \alpha_i u_i = u$ is called the *coordinate $n$-tuple of the vector $u$ with respect to the basis $\{u_1, \ldots, u_n\}$.* We call the $\alpha_i$ *coordinates* of $u$ w.r.t. the basis $\{u_1, \ldots, u_n\}$.

**Lemma 1.1.17.** *Let $\{u_1, \ldots, u_n\}$ be a basis for the $K$-vector space $V$. Then the* coordinate map*,*

$$\psi : V \,\to\, K^n \,:\, u \,\mapsto\, (\alpha_1, \ldots, \alpha_n)$$

*where the $\alpha_i$ are coordinates of $u$ w.r.t. the basis $\{u_1, \ldots, u_n\}$, is an isomorphism of vector spaces (bijective, and respects addition and scalar multiplication).*

*Proof.* Exercise! □

**Theorem 1.1.18.** *Let $V$ be a $K$-vector space. If $\{u_1, \ldots, u_n\}$ is a linearly independent set of vectors in $V$ and if $\{v_1, \ldots, v_m\}$ generates $V$, then*

$$n \leq m\,.$$

*In other words, the cardinality of a linearly independent set is less than or equal to the cardinality of a generating set.*

*Proof.* We know that $u_1 \neq 0$, since it is part of a linearly independent set (verify this!). Now $\{v_1, \dots, v_m\}$ generates $V$ implies that

$$u_1 \;=\; \alpha_1 v_1 + \cdots + \alpha_m v_m$$

for some scalars $\alpha_i$. Since $u_1 \neq 0$ we know that not all the $\alpha_i$ are zero. Suppose (by reordering the $v_j$ if necessary) that $\alpha_1 \neq 0$. Then we can "solve for $v_1$" to get

$$v_1 \;=\; \frac{1}{\alpha_1} u_1 - \frac{\alpha_2}{\alpha_1} v_2 - \cdots - \frac{\alpha_m}{\alpha_1} v_m \,.$$

But this means that $\{u_1, v_2, \dots, v_m\}$ generates $V$.

Again, $u_2 \in V$ and $\{u_1, v_2, \dots, v_m\}$ generates $V$ implies that there are scalars $\beta_1, \dots, \beta_m \in K$ such that

$$u_2 \;=\; \beta_1 u_1 + \beta_2 v_2 + \cdots + \beta_m v_m \,.$$

Note that $\beta_2, \dots, \beta_m$ cannot all be zero since the set $\{u_1, u_2\}$ is linearly independent. We may assume (by relabeling if necessary) that $\beta_2 \neq 0$. As before we can "solve for $v_2$" and conclude that $\{u_1, u_2, v_3, \dots, v_m\}$ generates $V$. (Provide details!).

Note that if $n > m$ (the argument is by contradiction here) then we can proceed as above (prove the inductive step!) to replace all the $v_i$'s by $u_i$'s and get that $\{u_1, \dots, u_m\}$ generates $V$. But then we would have (under the assumption that $n > m$) scalars $\gamma_1, \dots, \gamma_m$ such that

$$u_{m+1} \;=\; \gamma_1 u_1 + \cdots + \gamma_m u_m \,.$$

But this contradicts the linear independence of $u_1, \dots, u_n$. □

**Corollary 1.1.19.** *Let $\{u_1, \dots, u_n\}$ and $\{v_1, \dots, v_m\}$ be two bases for a vector space $V$. Then $m = n$.*

*Proof.* By definition of a basis, we have that $\{u_1, \dots, u_n\}$ is linearly independent, and that $\{v_1, \dots, v_m\}$ generates $V$. Thus Theorem 1.1.18 implies that $n \leq m$.

However, we also know that $\{v_1, \dots, v_m\}$ is linearly independent, and that $\{u_1, \dots, u_n\}$ generates $V$. Now Theorem 1.1.18 gives us $m \leq n$.

Combining the two inequalities we get $n = m$. □

**Definition 1.1.20.** A vector space is said to be *finite dimensional* if it has a finite basis. Otherwise it is said to be *infinite dimensional.* If the $K$-vector space $V$ is finite dimensional, then the number of elements in any basis of $V$ is called the *dimension of* $V$, and is denoted by $dim_K(V)$.

**Examples 1.1.21.** Say which of the following vector spaces are finite dimensional. Compute the dimension in the cases where it is finite.

1. $\mathbb{R}$ as an $\mathbb{R}$-vector space.

2. $\mathbb{Q}(\sqrt{2})$ as a $\mathbb{Q}$-vector space.

3. $\mathbb{C}$ as an $\mathbb{R}$-vector space.

4. $\mathbb{R}$ as a $\mathbb{Q}$-vector space.

5. $\mathbb{C}$ as a $\mathbb{C}$-vector space.

6. $K[x]$ as a $K$-vector space.

7. Set of polynomials of degree at most $n$ as a $K$-vector space.

8. $K^{m\times n}$ as a $K$-vector space.

9. The space of $(n \times n)$-symmetric real matrices as an $\mathbb{R}$-vector space.

10. The space of $(n \times n)$-skew symmetric real matrices as an $\mathbb{R}$-vector space.

11. $\mathcal{C}([0,1])$ as an $\mathbb{R}$-vector space.

12. The space of solutions to the equation $\frac{d^2y}{dx^2} + \lambda^2 y = 0$ where $\lambda > 0$ is real, as an $\mathbb{R}$-vector space.

13. The space of solutions to the homogenous system (remember Math 3333!?)
$$A_{m\times n}x_{n\times 1} = 0_{m\times 1} .$$

14. We have seen in 1.1.5.7 that if $K \subset L$ are fields, then $L$ is a $K$-vector space. Suppose that $K \subset L \subset M$ are fields, and that $L$ is a finite dimensional $K$-vector space, and $M$ is a finite dimensional $L$-vector space. Prove that $M$ is a finite dimensional $K$-vector space, and that
$$dim_K(M) = dim_K(L)dim_L(M) .$$

**Remark 1.1.22.** The field of constructible numbers has infinite dimension over $\mathbb{Q}$. However, the rational vector subspace of $\mathbb{R}$ which is spanned by a finite set of constructible numbers, $\{\alpha_1, \dots, \alpha_n\}$, has finite dimension over $\mathbb{Q}$. It **can be shown** (see a standard text on Abstract Algebra) that $dim_{\mathbb{Q}}(\mathcal{S}(\alpha_1, \dots, \alpha_n))$ is always a power of 2. It can also be shown that $dim_{\mathbb{Q}}(\sqrt[3]{2}) = 3$. Now, if $\sqrt[3]{2}$ is constructible, it belongs to $(\mathcal{S}(\alpha_1, \dots, \alpha_n)$ for some set $\{\alpha_1, \dots, \alpha_n\}$ of constructible numbers. Therefore 1.1.21.(14) implies that 3 should divide a power of 2 which is impossible! Thus $\sqrt[3]{2}$ is not constructible. This is one neat application of vector spaces; in the resolution of a 2000 year old problem from geometry: the problem of duplicating the cube.

Next result says that (at least in finite dimensional case) we can always complete any linearly independent set to a basis, and we can always "trim away at" any generating set to obtain a basis.

**Proposition 1.1.23.** *Let $V$ be a finite dimensional vector space. Let $Z$ be a generating set for $V$, and let $X$ be a linearly independent subset of $Z$. Then there exists a basis $Y$ for $V$ such that*
$$X \subset Y \subset Z$$

*Proof.* Let $Y$ be a maximal linearly independent subset of $Z$ which contains $X$. Complete the proof!

Hint, to show generation, it suffices to show that $Y$ generates all the elements of $Z$. Given $z \in Z \setminus Y$ then $\{z, y_1, \dots, y_m\}$ is linearly dependent (why?), and so there exist scalars $\lambda_0, \dots, \lambda_m$ such that
$$\lambda_0 z + \lambda_1 y_1 + \cdots + \lambda_m y_m = 0 .$$
Moreover $\lambda_0 \neq 0$ by linear independence of $Y$ (why?!). Now finish the proof. □

**Proposition 1.1.24.** *Let $U$ be a finite dimensional $K$-vector space, and let $V \subset U$ be a subspace. Then $V$ is finite dimensional and*

$$dim_K(V) \le dim_K(U).$$

*Proof.* Exercise. □

**Definition 1.1.25.** Let $U$ and $V$ be $K$-vector spaces. The *external direct sum* of $U$ and $V$ is denoted by $U \oplus V$, and is defined to be the set $U \times V$ together with coordinate-wise addition and scalar multiplication.

$$(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2)$$

and

$$k(u, v) = (ku, kv).$$

Now suppose that $U$ and $V$ are subspaces of the $K$-vector space $W$. We say that $W$ is the *internal direct sum* of $U$ and $V$ if the following properties hold.

- $U \cap V = 0$
- $U + V = W$

**Examples 1.1.26.** Here are some natural examples of direct sums.

1. $\mathbb{R}^2 = \mathbb{R} \oplus \mathbb{R}$
2. $\mathbb{R}^{\mathbb{R}} = Even \oplus Odd$ [Think about the polynomial version of this]
3. $K^{n \times n} = Symm \oplus Skew$

**Proposition 1.1.27.** *If $U$ and $V$ are finite dimensional $K$-vector spaces, and if $W = U \oplus V$, then $W$ is also finite dimensional, and*

$$dim_K(W) = dim_K(U) + dim_K(V)$$

*Proof.* Exercise! □

**Proposition 1.1.28.** *Let $U$ and $V$ be subspaces of the $K$-vector space $W$. Then the following are equivalent.*

1. *$W$ is the internal direct sum of $U$ and $V$*
2. *Every element $w \in W$ can be written in a unique way as a sum $w = u + v$ where $u \in U$ and $v \in V$*

*In this case, $W$ is isomorphic to the external direct sum of $U$ and $V$. Conversely, if a vector space $W$ is isomorphic to the external direct sum $U \oplus V$ of vector spaces $U$ and $V$, then $W$ can be decomposed as the internal direct sum of subspaces $U'$ and $V'$ with $U'$ isomorphic to $U$ and $V'$ isomorphic to $V$.*

*Proof.* Exercise! □

**Remark 1.1.29.** Because of the isomorphism above, we denote any direct sum (internal or external) of $U$ and $V$ by $U \oplus V$.

## 1.2 Linear Transformations and Coordinates

**Definition 1.2.1.** Let $U$ and $V$ be vector spaces over the field $K$. A *linear transformation* (or *linear operator*) from $U$ to $V$ is a map $T : U \to V$ such that

$$T(u_1 + u_2) \,=\, T(u_1) \,+\, T(u_2) \quad \text{for all } u_1, u_2 \in U$$

and

$$T(\lambda u) \,=\, \lambda T(u) \quad \text{for all } u \in U \text{ and all scalars } \lambda \in K.$$

**Examples 1.2.2.** Some examples of linear operators.

1. $I_U : U \to U \; : u \mapsto u$ for all $u \in U$
2. $O : U \to V \; : u \mapsto 0$ for all $u \in U$
3. $kI_U : U \to U \; : u \mapsto ku$ for all $u \in U$. Here $k \in K$.
4. Differentiation of polynomials, $D : K[x] \to K[x]$.
5. $A_{m\times n} : K^n \to K^m \; : X_{n\times 1} \mapsto AX$
6. $T : K^{m\times n} \to K^{m\times n} \; : A \mapsto P_{m\times m} A Q_{n\times n}$ for given $P \in K^{m\times m}$ and $Q \in K^{n\times n}$.
7. $Int : \mathcal{C}(\mathbb{R}) \to \mathcal{C}^1(\mathbb{R}) \; : f \mapsto Int(f)$ where
$$Int(f)(x) \,=\, \int_0^x f(t)dt\,.$$
8. The coordinate map $V \to K^n$ which takes a vector to its coordinates wit respect to a basis $\{u_1, \ldots, u_n\}$ for $V$.
9. Rotations in $\mathbb{R}^2$
10. Reflections in $\mathbb{R}^n$ (starting with $n = 2$)
11. Projections in $\mathbb{R}^n$
12. The transpose map $K^{m\times n} \to K^{n\times m}$

**Properties 1.2.3.** Here are some elementary properties which are satisfied by linear transformations.

1. $T : U \to V$ satisfies $T(-u) = -T(u)$ for all $u \in U$
2. $T : U \to V$ satisfies $T(0) = 0$
3. $T$ preserves collinearity.
4. $T$ preserves parallelograms.
5. $T : \mathbb{R} \to \mathbb{R}$ is linear if and only if its graph is a straight line through the origin.

6. $T(\sum_{i=1}^{n} \alpha_i u_i) = \sum_{i=1}^{n} \alpha_i T(u_i)$

7. $S, T : U \to V$ linear, $\{u_i\}$ generates $U$, and $T(u_i) = S(u_i)$ for all $i$, implies that $S = T$

8. Let $\{u_i\}_{i=1}^{n}$ be a basis for $U$ and let $w_i \in V$, then there exists a unique linear mapping $T : U \to V$ such that $T(u_i) = w_i$ for all $1 \leq i \leq n$.

**Definition 1.2.4.** Let $T : U \to V$ be a linear mapping. Then

$$Ker(T) \;=\; \{u \in U \,|\, T(u) = 0\}$$

and

$$Im(T) \;=\; \{T(u) \,|\, u \in U\}$$

are vector subspaces of $U$ and $V$ respectively. (prove this!) We define

$$Rank(T) \;=\; dim_K(Im(T))$$

and

$$Nullity(T) \;=\; dim_K(Ker(T))\,.$$

**Properties 1.2.5.** Note that (prove it!) the linear map $T$ is injective if and only if $Ker(T) = 0$.

**Examples 1.2.6.** We've seen many examples in Math 3333.

1. $Ker(A_{m\times n})$ is the solution set to the homogenous system

$$A_{m\times n}X_{n\times 1} = 0_{m\times 1}\,.$$

2. General solution to linear system $A_{m\times n}X_{n\times 1} = B_{m\times 1}$ consists of a particular solution to $A_{m\times n}X_{n\times 1} = B_{m\times 1}$ plus the general solution to the homogenous system $A_{m\times n}X_{n\times 1} = 0_{m\times 1}$.

3. Projection map of $\mathbb{R}^3$ onto $\mathbb{R}^2$.

**Theorem 1.2.7.** *Let $T : U \to V$ be a linear map, and suppose that $\{u_1, \ldots, u_k\}$ and $\{T(w_1), \ldots, T(w_l)\}$ are bases for $Ker(T)$ and $Im(T)$ respectively. Then $\{u_1, \ldots, u_k, w_1, \ldots, w_l\}$ is a basis for $U$. In particular,*

$$dim_K(U) \;=\; Nullity(T) \;+\; Rank(T)\,.$$

*Proof.* Exercise! There are two things to prove about the set of vectors $\{u_1, \ldots, u_k, w_1, \ldots, w_l\}$: it is a linearly independent set and it generates $U$.

**Linear independence.** Suppose

$$\alpha_1 u_1 + \cdots + \alpha_k u_k + \beta_1 w_1 + \cdots + \beta_l w_l = 0\,.$$

We have to prove that $\alpha_1 = 0, \ldots, \alpha_k = 0, \beta_1 = 0, \ldots, \beta_l = 0$. First apply $T$ to both sides of the equation above. Does this simplify at all? Why? What does the resulting equation tell you about the $\beta_i$? Why? Now plug this information about the $\beta_i$ back into the original equation above. What do you get now? What can you conclude about the $\alpha_i$? Why?

**Generating set.** Given $u \in U$ you have to find scalars $\alpha_i$ and $\beta_j$ so that

$$u \;=\; \alpha_1 u_1 + \cdots + \alpha_k u_k + \beta_1 w_1 + \cdots + \beta_l w_l \,.$$

First look at $T(u)$. What do you know about $T(u)$ and the $T(w_j)$? What does this tell you about $u$ and a linear combination of the $w_j$? (careful here!). How do the $u_i$ come in to play here? conclude the proof. □

**Corollary 1.2.8.** *Suppose that $T : U \to V$ is a linear map of finite dimensional vector spaces such that $dim_K(U) = dim_K(V)$. Then $T$ is injective if and only if $T$ is surjective.*

*Proof.* Use result above, together with 1.2.5. □

**Definition 1.2.9.** Let $U$ and $V$ be $K$-vector spaces with bases $\mathcal{B}_1 = \{u_1, \dots, u_n\}$ and $\mathcal{B}_2 = \{v_1, \dots, v_m\}$ respectively. Let $\psi_1 : U \to K^n$ and $\psi_2 : V \to K^m$ be the corresponding coordinate isomorphisms. Suppose that $T : U \to V$ is a linear map. We have seen in 1.2.3.8 that $T$ is uniquely determined by the vectors $\{T(u_i)\}_{i=1}^n$.

Define a matrix $A_T \in K^{m \times n}$ by setting $\psi_2(T(u_j))$ to be its $j$-th column for each $1 \leq j \leq n$. In other words, define

$$a_{ij} \text{ is the } i\text{-th coordinate (w.r.t. } \mathcal{B}_2\text{) of the vector } T(u_j).$$

We can visualize this in terms of matrices as follows

$$\begin{array}{cc} & \begin{array}{ccccc} \phantom{\alpha_{11}} & \phantom{\cdots} & [T(u_j)]_{\mathcal{B}_2} & \phantom{\cdots} & \phantom{\alpha_1 n} \\ & & \downarrow & & \end{array} \\ A_t \;=\; & \begin{pmatrix} \alpha_{11} & \cdots & \alpha_1 j & \cdots & \alpha_1 n \\ \vdots & & \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mj} & \cdots & \alpha_{mn} \end{pmatrix} \end{array}$$

Note that there is a commutative diagram

$$\begin{tikzcd} U \arrow[r, "T"] \arrow[d, "\psi_1"'] & V \arrow[d, "\psi_2"] \\ K^n \arrow[r, "A_T"'] & K^m \end{tikzcd}$$

That is,

$$\psi_2(T(u)) \;=\; A_T(\psi_1(u)) \qquad \text{for all } u \in U.$$

We call $A_T$ the *matrix of the linear map $T$ with respect to the bases $\mathcal{B}_1$ for $U$ and $\mathcal{B}_2$ for $V$.*

**Examples 1.2.10.** Compute the matrices of the following linear maps with respect to the given bases or verify the claims that are made as appropriate. Keep in mind that the $\psi$ isomorphisms take us from the realm of abstract vector spaces into the concrete coordinate world of $K^n$ spaces.

1. $D$ is the derivative linear map from the space $U$ of polynomials of degree at most $n$ into itself, and $\mathcal{B}_1 = \mathcal{B}_2 = \{1, x, x^2, \ldots, x^n\}$.

2. $U = V = \mathbb{R}^n$, $T$ is the identity map, $\mathcal{B}_1 = \mathcal{B} = \{v_1, \ldots, v_n\}$, and $\mathcal{B}_2$ is the standard basis: $\{e_1, \ldots, e_n\}$. This matrix is called *the change of basis matrix* from the basis $\mathcal{B}$ to the standard basis. We shall denote it by $P_{\mathcal{B}}$.

   The matrix which changes from the standard basis to $\mathcal{B}$ is just $P_{\mathcal{B}}^{-1}$.

3. Let $\mathcal{B}_1$ and $\mathcal{B}_2$ be two bases for $\mathbb{R}^n$. The change of basis matrix from $\mathcal{B}_1$ to $\mathcal{B}_2$ is given by the matrix
$$P = P_{\mathcal{B}_2}^{-1} P_{\mathcal{B}_1}$$

4. Let $T : U \to U$ be a linear transformation on a finite dimensional vector space $U$. Suppose that $T$ has matrix $A$ with respect to the basis $\mathcal{B}$ for $U$, and let $\mathcal{B}'$ be another basis for $U$.

   Then the matrix for $T$ with respect to the basis $\mathcal{B}'$ is given by
$$PAP^{-1}$$
   where is the change of basis matrix from $\mathcal{B}$ to $\mathcal{B}'$.

5. $T : \mathbb{R}^n \to \mathbb{R}^n$ is the linear map which is defined as a permutation of the standard basis of $\mathbb{R}^n$ and extended by linearity. Basis for $\mathbb{R}^n$ is standard.

   Do an explicit computation for the symmetric group $S_3$ of all permutations of a set of three elements which acts by linear transformations on $\mathbb{R}^3$. See that these give isometries of $\mathbb{R}^3$ which preserve the 2-simplex with vertices at $(1, 0, 0)$, $(0, 1, 0)$ and $(0, 0, 1)$. Generalize this to higher dimensions.

6. Let $U$, $V$, and $W$ be $K$-vector spaces with chosen bases $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$ respectively. Let $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$ have matrices $A$ and $B$ (respectively) with respect to the given bases. Then $ST$ has matrix $AB$ with respect to the bases $\mathcal{B}_1$ and $\mathcal{B}_3$.

7. Let $R_\theta$ denote the linear transformation of $\mathbb{R}^2$ which consists of a standard rotation of $\theta$ radians about the origin. Here standard means counterclockwise for positive $\theta$. Let $L_\theta$ denote the linear transformation of $\mathbb{R}^2$ which consists of a reflection in the line $l_\theta$ which contains the origin and makes an angle of $\theta$ radians in standard position (that is, $x$-axis is initial edge, and $l_\theta$ is terminal edge, and positive angles are measured in the usual counterclockwise direction).

   We have
$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \qquad L_\theta = \begin{pmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{pmatrix}$$
   Verify by matrix multiplication that $R_\theta R_\phi = R_{\theta+\phi}$ and that $L_\theta L_\phi = R_{2(\theta-\phi)}$. Interpret these results geometrically.

   Compute and interpret the following $L_\theta L_\theta$, $L_\phi L_\theta$, $L_\theta R_\phi$, and $R_\phi L_\theta$.

8. Realize the group of isometries of the Euclidean plane as a group of linear transformations of $\mathbb{R}^3$ to itself. What is the special form of these linear transformations. Geometric interpretations.

**Definition 1.2.11.** Let $U$ and $V$ be $K$-vector spaces. We denote by $\mathcal{L}(U, V)$ the set of all linear mappings $U \to V$, and by $\mathcal{L}(U)$ the set of all linear mappings $U \to U$.

**Theorem 1.2.12.** *Let $U$ and $V$ be $K$-vector spaces. Then we have*

1. *$\mathcal{L}(U, V)$ is a $K$-vector space, with operations defined by*

$$(S+T)(x)\,, = \; S(x) + T(x)\,,$$

*and*

$$(kT)(x) \; = \; k(T(x))$$

*for all $S, T \in \mathcal{L}(U, V)$, $x \in U$, and all $k \in K$.*

2. *Let $W$ be a $K$-vector space. Composition of maps gives a multiplication*

$$\mathcal{L}(V, W) \times \mathcal{L}(U, V) \; \to \; \mathcal{L}(U, W) \; : (S, T) \mapsto ST$$

*where $ST(u) = S(T(u))$ for all $u \in U$. This multiplication satisfies*

*(a)* $(RS)T = R(ST)$

*(b)* $R(S+T) = RS + RT$

*(c)* $(R+S)T = RT + ST$

*(d)* $(kS)T = k(ST) = S(kT)$

*provided each is well-defined.*

*Proof.* Easy! Exercise. □

**Corollary 1.2.13.** *Let $U$ be a $K$-vector space. Then $\mathcal{L}(U)$ is*

1. *a $K$-vector space*

2. *a ring under $S + T$ and $ST$*

3. *$(kS)T = k(ST) = s(kT)$ for all $k \in K$ and for all $S, T \in \mathcal{L}(U)$.*

**Definition 1.2.14.** Let $K$ be a field of scalars. A set $X$ on which there is an addition, a multiplication, and a scalar multiplication all satisfying 1–3 of the corollary above is called a $K$-algebra.

**Proposition 1.2.15.** *Let $U$ and $V$ be $K$-vector spaces of dimension $n$ and $m$ respectively. Then $\mathcal{L}(U, V)$ is isomorphic to $K^{m \times n}$ as $K$-vector spaces. Furthermore, $\mathcal{L}(U)$ is isomorphic to $K^{n \times n}$ as $K$-algebras.*

*Proof.* Exercise. □

**Corollary 1.2.16.** *Let $U$ and $V$ be $K$-vector spaces of dimension $n$ and $m$ respectively. Then*

$$dim_K(\mathcal{L}(U,V)) \,=\, mn$$

*and*

$$dim_K(\mathcal{L}(U)) \,=\, n^2\,.$$

*Proof.* Exercise. □

**Remark 1.2.17.** The *Einstein summation convention* requires one to sum over any repeated upper and lower indices in an expression (involving tensors). For example the expression $\sum_{j=1}^{n} \alpha_i^j v_j$ becomes the simpler expression $\alpha_i^j v_j$. We shall not make too much use of this tensor notation (upper and lower indices) in this course. Nevertheless, it's good to be aware of it.

## 1.3 Linear functionals and duality

**Definition 1.3.1.** Let $U$ be a $K$-vector space. A *linear functional on* $U$ is a linear transformation $U \to K$. The vector space $\mathcal{L}(U, K)$ of all linear functionals on $U$ is called the *dual space of* $U$ and is denoted by $U^*$.

**Examples 1.3.2.** Examples of linear functionals include.

1. Let $(\gamma_1, \dots, \gamma_n)^T \in K^n$. Then

$$K^n \to K : (k_1, \dots, k_n)^T \mapsto (\gamma_1, \dots, \gamma_n)(k_1, \dots, k_n)^T$$

   is a linear functional on $K^n$.

2. The definite integral $I(f) = \int_a^b f(t)dt$ is a linear functional on $\mathcal{C}([a, b])$.

3. The trace of a matrix defines a linear functional on $K^{n\times n}$.

**Remark 1.3.3.** By Corollary 1.2.16 if $U$ is a finite dimensional $K$-vector space, then

$$dim_K(U^*) \;=\; dim_K(\mathcal{L}(U, K)) \;=\; dim_K(U)dim_K(K) \;=\; dim_K(U)$$

so that $U$ and $U^*$ are both isomorphic to a given $K^n$, and hence to each other. However, this is not a *natural isomorphism* since it depends on a choice of bases for $U$ and $U^*$.

**Definition 1.3.4.** Bracket notation for linear functional $f$ acting on vector $u$, $\langle\, f \,,\, u \rangle$.

**Definition 1.3.5.** Kronecker delta is defined for $i, j \in \{1, \dots, n\}$

$$\delta_{ij} \;=\; \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

**Proposition 1.3.6.** *Let $U$ be a $K$-vector space with basis $\{u_1, \dots, u_n\}$. Define linear functionals $\{f_1, \dots, f_n\}$ on $U$ by*

$$\langle\, f_i \,,\, u_j \,\rangle \;=\; \delta_{ij}$$

*Then $\{f_1, \dots, f_n\}$ is a basis for $U^*$ which we call the dual basis to $\{u_1, \dots, u_n\}$.*

*Proof.* Exercise. □

**Remark 1.3.7.** If $V$ has basis $\{v_1, \dots, v_m\}$ and $\{f_1, \dots, f_m\}$ is a dual basis, then

$$\langle\, f_i \,,\, \sum_{j=1}^{m} c_j v_j \,\rangle \;=\; \sum_{j=1}^{m} c_j \langle\, f_i \,,\, m_j \,\rangle \;=\; \sum_{j=1}^{m} c_j \delta_{ij} \;=\; c_i$$

so that $f_i$ picks out the $i$-th coordinate (with respect to the basis $\{v_1, \dots, v_n\}$) of a vector $v \in V$. If $T : U \to V$ is a linear transformation, and $\{u_1, \dots, u_n\}$ is a basis for $U$, then the matrix $A = (a_{ij})$ of $T$ with respect to these two bases is given by

$$a_{ij} \;=\; \langle\, f_i \,,\, T(u_j) \,\rangle \,.$$

The bracket notation makes it clear that $U$ should act as a dual to $U^*$. This is content of next proposition.

**Proposition 1.3.8.** *If $U$ is s finite dimensional $K$-vector space, then $U$ is naturally isomorphic to $U^{**}$.*

*Proof.* Exercise. Given $v \in U$, let $L_v : U^* \to K$ be the *evaluation map* which sends a linear functional $f \in U^*$ to the scalar $f(v) \in K$. All you have to do is verify the following.

- $L_v$ is a linear functional on $U^*$.
- The map $v \mapsto L_v$ is linear.
- The map $v \mapsto L_v$ is bijective.

□

**Definition 1.3.9.** Let $S$ be a subset of a finite dimensional vector space $V$. The *annihilator of $S$* is denoted by $S^\circ$ and is defined as follows.

$$S^\circ \;=\; \{\, f \in V^* \mid f(v) = 0 \text{ for all } v \in S \,\}$$

**Properties 1.3.10.** The following should be intuitive properties. Give proofs of them all.

1. $S^\circ$ is a subspace of the dual space $V^*$.
2. If $S = \{0\}$, then $S^\circ = V^*$.
3. If $S = V$, then $S^\circ = \{0\} \subset V^*$.

**Definition 1.3.11.** Let $V$ be an $n$-dimensional vector space. A *hyperspace* is a subspace of $V$ which has dimension $(n-1)$.

**Remark 1.3.12.** Hyperspaces in the finite dimensional vector space $V$ are precisely the kernels of linear functionals on $V$.

**Lemma 1.3.13.** *Let $W$ be a subspace of the finite dimensional vector space $V$, then*

$$dim(W) \,+\, dim(W^\circ) \;=\; dim(V)$$

*Proof.* Exercise. Let $\{v_1, \ldots, v_k\}$ be a basis for $W$. Complete it to a basis $\{v_1, \ldots, v_k, \ldots, v_n\}$ for $V$. Let $\{f_1, \ldots, f_n\}$ be the corresponding dual basis for $V^*$. Prove that $\{f_{k+1}, \ldots, f_n\}$ is a basis for $W^\circ$. □

**Corollary 1.3.14.** *Let $V$ be an $n$-dimensional vector space. Then each $k$-dimensional subspace (here $k \leq n$) of $V$ is the intersection of $(n-k)$ hyperspaces of $V$.*

**Corollary 1.3.15.** *If $W_1$ and $W_2$ are subspaces of a finite dimensional vector space $V$. Then $W_1 = W_2$ if and only if $W_1^\circ = W_2^\circ$.*

**Definition 1.3.16.** Let $T : U \to V$ be a linear transformation of finite dimensional $K$-vector spaces. Then some earlier proposition (in l.t. section) ensures that there is a unique linear transformation $T' : V^* \to U^*$ satisfying

$$\langle\, T'(f)\,,\, u\,\rangle \;=\; \langle\, f\,,\, T(u)\,\rangle \qquad \text{for all } u \in U \text{ and all } f \in V^*.$$

$T'$ is called the *adjoint* of the operator $T$.

**Remark 1.3.17.** Note that for $f \in V^*$ we have just defined $T'(f)$ to be the composition

$$T'(f) \;=\; f \circ T$$

The adjoint construction gives rise to a homomorphism (verify this!)

$$Ad : \mathcal{L}(U, V) \;\to\; \mathcal{L}(V^*, U^*) \;\; : T \;\mapsto\; Ad(T) = T'$$

**Proposition 1.3.18.** *Let $T : U \to V$ above have matrix $A$ with respect to bases $\mathcal{B}_1$ for $U$ and $\mathcal{B}_2$ for $V$. Then $T'$ has matrix $A^T$ (transpose) with respect to the respective dual bases for $V^*$ and $U^*$.*

*Proof.* Exercise. □

**Proposition 1.3.19.** *Let $T : U \to V$ be a linear transformation of the vector spaces $U$ and $V$. Then the kernel of the adjoint $T'$ of $T$ is the annihilator of the image of $T$. In particular, if $U$ and $V$ are finite dimensional we have*

1. *$rank(T') = rank(T)$.*
2. *image of $T'$ is the annihilator of the kernel of $T$.*

**Corollary 1.3.20.** *For $(n \times n)$-matrices $A$ we have*

$$rank(A) = rowrank(A) = columnrank(A)$$

## 1.4 Determinants

**Definition 1.4.1.** Let $A \in M_{2\times 2}(K)$. Then the *determinant of* $A$ is denoted by $det(A)$ or by $|A|$ and is an element of the field $K$ defined by

$$det(A) \ = \ a_{11}a_{22} - a_{12}a_{21}$$

**Remark 1.4.2.** For $A \in M_{2\times 2}(K)$, we have $A$ is invertible if and only if $|A| \neq 0$, and in such a case we have

$$A^{-1} \ = \ \frac{1}{|A|}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

**Definition 1.4.3.** Here's an inductive definition of determinants of $(n \times n)$-matrices. Throughout this definition we shall assume that $A \in M_{n\times n}(K)$ for some field $K$.

- For $i, j \in \{1, \ldots, n\}$ let $\widehat{A}_{ij}$ denote the element of $M_{(n-1)\times(n-1)}(K)$ which is obtained by deleting the $i$-th row and the $j$-th column from $A$.
- If $n = 1$ we define $|A| = a_{11}$, otherwise we define $|A|$ inductively by *cofactor expansion along its first row* as follows
$$|A| \ = \ \sum_{i=1}^{n}(-1)^{i+1}a_{1i}|\widehat{A}_{1i}|$$
- The terms $(-1)^{i+j}|\widehat{A}_{ij}|$ are called the *ij-cofactors* of the matrix $A$. They shall appear below in a more general formula for expansion of determinants by any row or column.

**Examples 1.4.4.** Here are some determinants.

1. $det(L_\theta) = -1$
2. $det(R_\theta) = 1$
3. determinant of upper (lower) triangular matrix
4. Vandermonde determinant: $det\begin{pmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{pmatrix} \ = \ \prod_{i<j}(x_i - x_j)$

**Properties 1.4.5.** Next we list some properties of determinants which can be deduced from the definition above. You should verify that these properties also hold if we use instead the definition of the determinant by cofactor expansion along the $i$-th row. Namely,

$$det_i(A) \ = \ \sum_{j=1}^{n}(-1)^{i+j}a_{ij}|\widehat{A}_{ij}|$$

so that $|A|$ or $det(A)$ above is actually $det_1(A)$. In time, we shall deduce that all these expansions give the same result, and that you can expand by columns too.

1. The determinant is a linear function of each column when the other columns are kept fixed. That is, letting $A^i$ denote the $i$-th column of $A$ and $A'^i$ denote a column vector and $k, k' \in K$, we have

$$det(A^1, \ldots , kA^i + k'A'^i, \ldots , A^n) \; = \; kdet(A^1, \ldots , A^i, \ldots , A^n) + k'det(A^1, \ldots , A'^i, \ldots , A^n)$$

2. If $A$ has two adjacent columns which are equal, then $|A| = 0$.
3. Let $I_n$ denote the $(n \times n)$-identity matrix. Then $|I_n| = 1$.
4. If adjacent columns of $A$ are interchanged, then $|A|$ changes sign.
5. If two columns of $A$ are equal, then $|A| = 0$.
6. If one adds a scalar multiple of one column of $A$ to another, then $|A|$ does not change.

**Remark 1.4.6.** There are two neat things to note here.

1. Properties 4, 5, and 6 (as well as any deductions from them below) all follow from properties 1–3. So any function of $n$ column vectors which satisfies 1–3 will have to be the determinant function (from the uniqueness result in 1.4.13 below).
2. Property 6 is the starting point for speedy computations of determinants (recall the hateful exercises in Math 3333).

The following theorem is a classical tool used in solving systems of linear equations called *Cramer's Rule.*

**Theorem 1.4.7.** *Let $A^1, \ldots , A^n, B \in \mathbb{R}^n$ be column vectors and suppose that $det(A^{,} \ldots , A^n) \neq 0$. Then we can solve the linear system*

$$x_1A^1 + \cdots + x_nA^n \; = \; B$$

*as follows*

$$x_j \; = \; \frac{det(A^1, \ldots , B, \ldots , A^n)}{det(A^1, \ldots , A^j, \ldots , A^n)}$$

**Theorem 1.4.8.** *Let $A^1, \ldots , A^n \in \mathbb{R}^n$ be column vectors. If $det(A^1, \ldots , A^n) \neq 0$ then $\{A^1, \ldots , A^n\}$ is a linearly independent set.*

**Corollary 1.4.9.** *If the column vectors $A^1, \ldots , A^n \in \mathbb{R}^n$ satisfy $det(A^1, \ldots , A^n) \neq 0$ then the linear system*

$$x_1A^1 + \cdots + x_nA^n \; = \; B$$

*has a solution (which can be found by Cramer's rule) for any column vector $B \in \mathbb{R}^n$.*

**Definition 1.4.10.** A *permutation* of the set $J_n = \{1, \ldots , n\}$ is just a bijection of this set to itself. A *transposition* is a permutation which just interchanges two elements of the set. Cyclic notation for permutations. The symmetric group $S_n$.

**Lemma 1.4.11.** *A permutation of $J_n$ is a product of transpositions.*

**Proposition 1.4.12.** *To each permutation $\sigma \in S_n$ we can associate a number $\epsilon(\sigma) \in \{\pm 1\}$ such that*

1. *$\epsilon(\tau) = -1$ for any transposition $\tau$*

2. *$\epsilon(\sigma_1\sigma_2) = \epsilon(\sigma_1)\epsilon(\sigma_2)$ for all permutations $\sigma_1, \sigma_2 \in S_n$.*

*In particular, if $\sigma$ can be expressed as a product of transpositions*

$$\sigma \;=\; \tau_1 \cdots \tau_m$$

*then $m$ is odd (even) according as $\epsilon(\sigma) = -1$ or $+1$.*

This next result establishes the uniqueness of determinants. It will be useful in proving some fundamental results about determinants such as the fact that the determinant of a transpose is the same as the determinant of the original matrix (which in turn leads to the familiar row or column cofactor expansion formula). It is also used to establish a geometric interpretation of determinants in terms of areas and volumes.

**Theorem 1.4.13.** *Let $U_1, \dots, U_n \in K^n$ be column vectors, and let new vectors $A^1, \dots, A^n \in K^n$ be defined by $A^j = \alpha_{1j}U_1 + \cdots + \alpha_{nj}U_n$ for scalars $\alpha_{ij} \in K$. Then*

$$det(A^1, \dots, A^n) \;=\; \sum_{\sigma \in S_n} \epsilon(\sigma)\alpha_{\sigma(1)1}\alpha_{\epsilon(2)2}\dots\alpha_{\epsilon(n)n} det(U_1, \dots, U_n)$$

*In particular, the determinant of a matrix $A = (a_{ij}) \in K^{n\times n}$ is given by the formula*

$$det(A) \;=\; \sum_{\sigma \in S_n} \epsilon(\sigma)a_{\sigma(1)1}a_{\epsilon(2)2}\dots a_{\epsilon(n)n}$$

There is another way of stating this. We say that a function of $n$ vector variables (which are each $n$-dimensional column vectors!) is *$n$-linear* if it is linear in each variable (keeping other variables fixed). Way that the function is *alternating* if its output is zero whenever two input variables are equal and if its output changes sign whenever two input variables are interchanged. The theorem becomes *let $D$ be an alternating $n$-linear function whose value on the input $(e_1, \dots, e_n)$ is 1, then $D = det$.*

**Theorem 1.4.14.** *Let $A$ be an $(n \times n)$-matrix and let $A^T$ denote its transpose. Then*

$$det(A) \;=\; det(A^T)$$

**Corollary 1.4.15.** *The determinant of the $(n\times n)$-matrix $A$ can be evaluated by cofactor expansion via any row or any column. That is*

$$det(A) \;=\; \sum_{i=1}^{n}(-1)^{i+j}a_{ij}|\widehat{A}_{ij}| \;=\; \sum_{j=1}^{n}(-1)^{i+j}a_{ij}|\widehat{A}_{ij}|$$

**Theorem 1.4.16.** *Let $A$ and $B$ be $(n \times n)$-matrices. Then*

$$det(AB) \;=\; det(A)det(B)$$

**Corollary 1.4.17.** *Let $A$ be an invertible $(n \times n)$-matrix. Then*

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

*and*

$$A^{-1} \;=\; \frac{1}{det(A)} \left((-1)^{i+j}|\widehat{A}_{ij}|\right)^T$$

**Remark 1.4.18.** Mention volumes functions in $\mathbb{R}^n$ and their relationship with determinants.

# Chapter 2

# Structure of Linear Operators

In the next few sections we shall develop a structure theory for linear transformations on a finite dimensional vector space. The basic problem that we are faced with is this: given a linear transformation $T$ on a finite dimensional $K$-vector space $V$, choose a basis for $V$ with respect to which $T$ is very easy to understand. For example the matrix of $T$ with respect to our basis has a very simple form.

What is the simplest form we should hope for? Well, diagonal matrices are very easy to work with. So we start off in section 2.1 by discussing eigenvectors and diagonalization. You may recall from Math 3333, that the basic goal is to find a basis for $V$ composed of eigenvectors of $T$.

There are a number of problems that may occur with our attempts at diagonalization. For instance, we may not be able to solve the characteristic equation for $T$ over the field $K$ (so we cant find any eigenvalues for $T$ in $K$). Or we may find that the dimensions of the eigenspaces do not sum up to give the dimension of $V$ (so we cant find a basis for $V$ consisting of eigenvectors of $T$). We are led to consider more general $T$-invariant subspaces, and eventually to the primary decomposition theorem (section 2.2), and the Jordan (section 2.3) and rational (section 2.3) forms. This theory involves a beautiful interplay between $T$-invariant subspaces, and polynomial combinations of the operator $T$.

## 2.1 Eigenvalues and Eigenvectors, Diagonalization

**Definition 2.1.1.**

- Let $T : V \to V$ be a linear operator of a $K$-vector space $V$ to itself. An element $\lambda \in K$ is called an *eigenvalue* of $T$ if there exists a nonzero vector $v \in V$ such that

$$Tv = \lambda v$$

- Suppose $\lambda \in K$ is an eigenvalue of the linear operator $T$. Then the collection

$$\{v \in V \mid Tv = \lambda v\}$$

is a subspace of $V$ called the $\lambda$-*eigenspace* of $T$. It's elements are called $\lambda$-eigenvectors (or just *eigenvectors* if the context is clear) of $T$.

- Note that $\lambda \in K$ is an eigenvalue of $T$ if and only if $T - \lambda I$ is singular (has nontrivial kernel), and this is true if and only if $det(T - \lambda I) = 0$. In this case, $Ker(T - \lambda I)$ is precisely the $\lambda$-eigenspace of $T$.

**Examples 2.1.2.**

1. 
$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$$

There is a basis for $\mathbb{R}^2$ consisting of eigenvectors of the linear operator, $T$, given by matrix multiplication by $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. We can use this to transform this matrix into a diagonal matrix. This can be used to compute high powers of our matrix. See Fibonacci sequence applications (in class).

2. 
$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

This matrix has eigenvectors and eigenvalues only when $\theta$ is a multiple of $\pi$. In these cases the original matrix is already diagonal, and the eigenvalues are clearly $\pm 1$. In all other cases the matrix does not have any eigenvalues over $\mathbb{R}$. However, when viewed as a 2-by-2 complex matrix, this has eigenvalues $e^{\pm i\theta}$. See earlier homework exercise.

3. 
$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

This matrix has eigenvalue equal to 1, but $\mathbb{R}^2$ does not have a basis of eigenvectors.

**Definition 2.1.3.** The *characteristic polynomial* of the $n$-by-$n$ matrix $A$ is defined to be the following polynomial in $\lambda$

$$det(A - \lambda I)$$

Note that the eigenvalues of the linear operator of $K^n$ given by matrix multiplication by $A$ are precisely the roots of this characteristic polynomial.

**Lemma 2.1.4.** *Similar matrices have the same characteristic polynomials*

**Corollary 2.1.5.** *A linear transformation $T : V \to V$ of a finite dimensional vector space has a well-defined characteristic polynomial. The eigenvalues of $T$ are the roots of this polynomial.*

**Definition 2.1.6.** A linear transformation $T : V \to V$ is said to be *diagonalizable* if there exists a basis for $V$ comprised entirely of eigenvectors of $T$.

**Lemma 2.1.7.** *Non-zero eigenvectors of a linear operator $T$ which correspond to distinct eigenvalues of $T$ are linearly independent.*

*Proof.* Suppose $v_i$ is a non-zero eigenvector of the linear operator $T$ with corresponding eigenvalue $\lambda_i$ for $1 \leq i \leq m$, and suppose that the $\lambda_i$ are all distinct.

We have to prove that

$$\sum_i \alpha_i v_i = 0 \qquad \text{implies} \qquad \alpha_i = 0 \qquad \text{for all } i.$$

We do this by induction on $m$. This is clearly true for $m = 1$ since we're considering non-zero eigenvectors. Applying $T$ to $\sum_i \alpha_i v_i = 0$ gives

$$\sum_i \alpha_i \lambda_i v_i \;=\; 0$$

since $T(v_i) = \lambda_i v_i$ for all $i$. On the other hand, multiplying $\sum_i \alpha_i v_i = 0$ across by $\lambda_j$ gives

$$\sum_i \alpha_i \lambda_j v_i \;=\; 0\,.$$

Subtracting these two equations gives

$$\sum_i \alpha_i(\lambda_j - \lambda_i)v_i \;=\; 0$$

Note that this sum has really got $m - 1$ terms (the term $i = j$ vanishes), and so the inductive hypothesis tells us that the $v_i$ $(i \neq j)$ are already linearly independent. Thus $\alpha_i(\lambda_j - \lambda_i) = 0$ for each $i \in \{1, \dots, m\} \setminus \{j\}$. Since the $\lambda_i$ are all distinct, we conclude that $\alpha_i = 0$ for all $i \neq j$. Putting this back into the original equation gives

$$0 + \alpha_j v_j + 0 \;=\; 0$$

and $v_j \neq 0$ gives us that the remaining $\alpha_j$ must be 0. Done. □

**Theorem 2.1.8.** *$V$ a finite dimensional vector space, $T : V \to V$ a linear transformation, $\{\lambda_1, \dots, \lambda_m\}$ the set of all distinct eigenvalues of $T$, and $W_i = Ker(T - \lambda_i I)$ for $1 \leq i \leq m$. Then the following are equivalent.*

1. *$T$ is diagonalizable*

2. *char poly of $T$ is of the form*

$$f \;=\; \prod_{i-1}^{m} (x - \lambda_i)^{d_i}$$

*and $dim_K(W_i) = d_i$ for $1 \leq i \leq m$.*

*3.* $dim_K(V) = dim_K(W_1) + \cdots + dim_K(W_m)$

*4.* $V = W_1 \oplus \cdots \oplus W_m$

*Proof.* Depending on how you try to answer this, there are at least four implications to prove. Here we'll give the minimum of four.

[1 → 2] By definition, $T$ diagonalizable implies that $V$ has a basis $\{v_1, \ldots, v_n\}$ of eigenvectors of $T$. The matrix representation of $T$ with respect to this basis is simply the diagonal matrix with the eigenvalues on the diagonal. Thus, the characteristic polynomial of $T$ is of the form

$$f_T(x) \;=\; \prod_{i=1}^{m}(x-\lambda_i)^{d_i}$$

where $d_i$ is the number of times $\lambda_i$ appears on the diagonal. That is $d_i$ is the number of $\lambda_i$-eigenvectors present in the basis $\{v_1, \ldots, v_n\}$. We just have to verify that each $\lambda_i$ appears, and that it appears precisely $dim_K(W_i)$ times.

- First we verify that each $\lambda_i$ appears. Suppose some $\lambda_i$ does not appear. This means that we have a basis for $V$ of eigenvectors of $T$ without ever having to use an eigenvector with eigenvalue $\lambda_i$. But such an eigenvector is an element of $V$, and so can be expressed as a linear combination of the basis elements. That is, a non-zero eigenvector can be expressed as a linear combination of eigenvectors with different eigenvalues. But this contradicts the previous lemma!

- Now we have to show that the subspace, $S$, spanned by all those basis vectors in $\{v_1, \ldots, v_n\}$ which correspond to a given eigenvalue, $\lambda_i$ say, is equal to the eigenspace $W_i$. Clearly (exercise!) this subspace is contained in $W_i$. To prove the reverse inclusion, suppose that $v \in V$ is a $\lambda_i$-eigenvector of $T$. Since $\{v_1, \ldots, v_n\}$ is a basis, $v$ can be expressed as a linear combination of the $v_i$. We have to show that this combination only involves the $v_j$ which are $\lambda_i$-eigenvectors. Well, if not then we get a nontrivial linear dependence relation between eigenvectors with distinct eigenvalues. Again, this contradicts the previous lemma.

[2 → 3] This is just a dimension count! And it's trivial. We know from the definition, that the characteristic polynomial has degree $n = dim_K(V)$. Property 2 tells us that $n = \sum_{i=1}^{m} d_i$, and it also tells us that $d_i = dim_K(W_i)$. So we're done!!

[3 → 4] First we show that the sum $W_1 + \cdots + W_m$ is direct. To do this it suffices (remember Midterm I) to prove that $W_j \cap \sum_{i \neq j} W_i = \{0\}$ for each $j$. But if this intersection contained a nonzero vector, then it could be expressed in two ways as follows

$$w_j \;=\; \sum_{i \neq j} w_i \,.$$

Since $w_j \neq 0$, then at least one of the $w_i$ must also be non-zero, and so we obtain another relation of linear dependence among eigenvectors with distinct eigenvalues, thus contradicting the previous lemma.

Now that we know this sum is direct we can say that $W_1 \oplus \cdots \oplus W_m$ is isomorphic to $\sum_{i=1}^m W_i$ which is a subspace of $V$ of dimension $\sum_{i=1}^m dim_K(W_i)$. But property 3 says that this is just $dim_K(V)$. Thus $\sum_{i=1}^m W_i$ is a subspace of $V$ of the same dimension as $V$. Thus $\sum_{i=1}^m W_i = V$ and we're done.

[$4 \to 1$] If $V$ is a direct sum of eigenspaces, then we can combine bases for these eigenspaces together to obtain a basis for $V$. Thus $T$ is diagonalizable by definition. □

**Remark 2.1.9.** We will establish a neat algorithm for checking if a linear operator is diagonalizable as a corollary of the Primary Decomposition Theorem in the next section.

## 2.2 Annihilating Polynomials, Hamilton-Cayley Theorem, Primary Decomposition Theorem

The main theme of this section is that one can understand a linear operator acting on a finite dimensional vector space by analyzing the polynomials which annihilate it. A beautiful practical existence result for annihilating polynomials is the Hamilton-Cayley Theorem. The main result which relates the structure of a linear operator on a finite dimensional vector space to the algebra of one of its annihilating polynomials is the Primary Decomposition Theorem. With this tool in hand, it will be a very short step to the neat classification of diagonalizable operators result which was promised in the previous section, and also to the Jordan Normal Form Theorem.

First we define what we mean by an annihilating polynomial.

**Definition 2.2.1.** Let $V$ be a $K$-vector space and let $T \in \mathcal{L}(V)$. An *annihilating polynomial* for $T$ is a polynomial $p \in K[x]$ such that $p(T) = 0$.

**Examples 2.2.2.** For example, the simplest annihilating polynomial for the identity operator is just $p(x) = x - 1$, for then $P(T) = 0$ means $T - I = 0$ which is true since $T = I$.

Here is an existence theorem for annihilating polynomials. It's proof is easy (just remember that $K^{n\times n}$ is $n^2$-dimensional, so that the $n^2+1$ matrices: $I, A, A^2, \ldots , A^{n^2}$ are linearly dependent).

**Lemma 2.2.3.** *Let $A \in K^{n\times n}$. Then there is a polynomial $p \in K[x]$ such that $p(A) = 0$. In fact we can choose $p$ to have degree at most $n^2$.*

*Same works for linear transformations of an $n$-dimensional $K$-vector space.*

**Examples 2.2.4.** For example, we know (from the proof of the theorem above) that $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ must satisfy a polynomial equation of degree at most 4. In fact, it satisfies $A^2 - 3A + I = 0$. You may also recall from the previous section that $x^2 - 3x + 1$ is the characteristic polynomial of $A$. That this is not just a coincidence is the subject of the Hamilton-Cayley Theorem. Before stating and proving it, we develop some intuition about matrix polynomials.

**Definition 2.2.5.** A *matrix polynomial* over the field $K$ is a matrix whose entries are polynomials (in $x$ say) with coefficients in the field $K$. It may be written as either

$$P(x) \; = \; \begin{pmatrix} p_{11}(x) & \cdots & p_{1n}(x) \\ \vdots & & \vdots \\ p_{m1}(x) & \cdots & p_{mn}(x) \end{pmatrix}$$

or as

$$P(x) \; = \; P_0 + xP_1 + x^2P_2 + \cdots + x^dP_d$$

where $P_j \in K^{m\times n}$.

**Examples 2.2.6.** Here is an example.

$$\begin{pmatrix} 1+x^2 & x \\ 2x+1 & 1-x^2 \end{pmatrix} \; = \; \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + x \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} + x^2 \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

**Definition 2.2.7.** If $P$ is an $m$-by-$m$ matrix polynomial written as

$$P(x) \;=\; P_0 + xP_1 + x^2P_2 + \cdots + x^dP_d$$

and $A \in K^{m\times m}$ then we may define $P(A)$ to be the $m$-by-$m$ matrix

$$P(A) \;=\; P_0 + AP_1 + A^2P_2 + \cdots + A^dP_d$$

In this case you have to be careful about multiplication. Note that $(P+Q)(A) = P(A)+Q(A)$ but that $PQ(A)$ need not be equal to $P(A)Q(A)$. For example, if $P(x) = xB$ and $Q(x) = xC$ then $PQ(x) = x^2BC$ and so we have: $P(A)Q(A) = ABAC$ while $PQ(A) = A^2BC$. These are not necessarily equal (if $A$ and $B$ do not commute).

**Lemma 2.2.8.** *Let $P(x)$ be a matrix polynomial of size $n \times n$ over the field $K$, and let $A \in K^{n\times n}$. Then $P(A) = 0$ if and only if there exists a matric polynomial $Q(x)$ of size $n\times n$ over $K$ such that*

$$P(x) \;=\; (xI - A)Q(x)$$

*Proof.* This seems completely intuitive. The thing to be careful about is the fact that these matrix polynomials do not have a commutative multiplication. We just have to remember the definition of evaluation of a matrix polynomial at a $n \times n$ matric $A$ given above.

Suppose that $P(A) = 0$. Then writing $P(x)$ out as

$$P(x) \;=\; P_0 + xP_1 + \cdots + x^dP_d$$

we get

$$\begin{aligned} P(x) = P(x) - 0 \;&=\; P(x) - P(A) \\ &=\; P_0 + xP_1 + \cdots + x^dP_d - P_0 - AP_1 - \cdots - A^dP_d \\ &=\; (xI - A)P_1 + (x^2I - A^2)P_2 + \cdots + (x^dI - A^d)P_d \\ &=\; (xI - A)Q(x) \end{aligned}$$

since each $(x^jI - A^j)$ term can be written as

$$(x^{j-1}I + x^{j-2}A + \cdots + xA^{j-2} + A^{j-1})$$

and so the $(xI - A)$ can be completely factored out on the left.

Conversely, suppose that $P(x) = (xI - A)Q(x)$ for some matrix polynomial

$$Q(x) = Q_0 + xQ_1 + \cdots + x^eQ_e\,.$$

Thus $P(x) = xQ_0 + x^2Q_1 + \cdots + x^{e+1}Q_e - AQ_0 - xAQ_1 - \cdots - x^eAQ_e$ and so we get

$$P(A) \;=\; AQ_0 + A^2Q_1 + \cdots + A^{e+1}Q_e - AQ_0 - AAQ_1 - \cdots - A^eAQ_e \;=\; 0$$

as required. □

**Theorem 2.2.9 (Hamilton-Cayley).** *Let $T$ be a linear transformation on a finite dimensional vector space $V$. If $f$ is the characteristic polynomial of $T$, then $f(T) = 0$.*

*Proof.* Let $f \in K[x]$ be the characteristic polynomial of $T$. We have to show that $f(T) = 0$ or equivalently, that $f(A) = 0$ where $A \in K^{n \times n}$ is the matrix of $T$ with respect to some basis for $V$. Here $n = dim_K(V)$. To do this let $P(x) = diag(f(x), \dots, f(x))$ be the $n \times n$ matrix polynomial consisting of $f(x)$'s along the diagonal and zeros elsewhere. Note that $P(x) = f(x)I$ so that $P(A) = f(A)I$ and so $P(A)$ is zero precisely when $f(A)$ is zero.

By the previous result, it suffices to find a polynomial matrix $Q(x)$ such that

$$P(x) \ = \ (xI - A)Q(x)$$

But we already know this (from the section on determinants and inverses!), namely the $ij$-entry of $Q(x)$ is simply $(-1)^{i+j} det(\widehat{xI - A})_{ji}$. Note the $ij$ $ji$ switch (accounts for the transpose in computing inverses). Note that the entries of the $(n-1) \times (n-1)$ matrix $(\widehat{xI - A})_{ji}$ are all polynomials in $x$ of degree at most 1, and so the determinant is a polynomial in $x$ of degree at most $n - 1$. Thus $Q(x)$ is clearly a matrix polynomial. □

So we have seen that linear transformations in $\mathcal{L}(V)$ (and so square matrices) satisfy polynomial equations. In particular, they are roots of their characteristic polynomials.

Now that we have found a good source of annihilating polynomials, we wish to develop a structure theorem for linear transformations based on properties of their annihilating polynomials. To do this, we need some definitions and results from polynomial algebra.

**Definition 2.2.10.** A polynomial $p \in K[x]$ is *reducible* if there exists a factorization

$$p \ = \ p_1 p_2$$

where $p_i \in K[x]$ are polynomials of strictly smaller degree than $p$. If no such factorization exists then $p$ is said to be *irreducible*. Note that irreducibility depends on the base field $K$ (eg. $x^2 + 1$ is irreducible in $\mathbb{R}[x]$ but is reducible in $\mathbb{C}[x]$).

A *greatest common divisor* (g.c.d.) of polynomials $p_1, \dots, p_n \in K[x]$ is a polynomial $p \in K[x]$ of maximal degree which divides evenly into all of the $p_i$. If the g.c.d. of the polynomials $p_1, \dots, p_n$ is 1 (or a scalar), then we say that the polynomials are *relatively prime*.

**Theorem 2.2.11.** *Every poly $p \in K[x]$ has a decomposition into irreducible factors $p_1 \dots p_n$. The number $n$ and the factors are uniquely determined (upto ordering and multiplication by non-zero scalars).*

*If the g.c.d. of polynomials $p_1, \dots, p_m$ is 1, then there exists polynomials $q_1, \dots, q_m \in K[x]$ such that*

$$q_1 p_1 + \cdots + q_m p_m \ = \ l$$

*Proof.* We refer the reader to an abstract algebra book for the first part, and give a proof of the second part here.

Let

$$\mathcal{S} \ = \ \{q_1 p_1 + \cdots + q_m p_m \ : \ q_i \in K[x]\}$$

be the set of all linear combinations (with polynomial coefficients!) of the $p_i$. Let $d \in \mathcal{S}$ have minimal degree. Note that we can write

$$d = q_1 p_1 + \cdots + q_m p_m$$

We claim that $d$ divides evenly into all the $p_i$. If not, then we can divide some $p_i$ by $d$ to get a nonzero remainder $r$ which necessarily has smaller degree than $d$. Say

$$p_i = qd + r$$

But we can rearrange this to get $r = qd - p_i$ and so $r \in \mathcal{S}$. But this contradicts the minimality of the degree of $d$.

So we have seen that $d$ is a common divisor of all the $p_i$. Since the g.c.d. of all the $p_i$ is 1, then $d$ must have degree 0. That is $0 \neq d \in K$. Replacing all the $q_i$ above by $q_i/d$ gives the desired expression for the constant polynomial 1 as an element of $\mathcal{S}$. □

**Examples 2.2.12.** Find polynomials $q_i$ $(i = 1, 2, 3)$ such that

$$q_1(x-1)(x-2) + q_2(x-2)(x-3) + q_3(x-3)(x-1) = 1$$

Hint: Thinking about Lagrange polynomials from Midterm II will help!

Now we are ready to state and prove the Primary Decomposition Theorem.

**Definition 2.2.13.** Let $V$ be a $K$-vector space and let $T \in \mathcal{L}(V)$. A subspace $U \subset V$ is said to be an *$T$-invariant subspace* if $T(U) \subset U$.

**Theorem 2.2.14 (Primary Decomposition).** *Let $V$ be a $K$-vector space, and let $T \in \mathcal{L}(V)$. Suppose that $p \in K[x]$ is an annihilating polynomial for $T$ which has a decomposition as*

$$p = p_1 \cdots p_k$$

*where the $p_j$ are relatively prime. Then we have:*

1. *$V = ker(p_1(T)) \oplus \cdots \oplus ker(p_k(T))$, and each of these are $T$-invariant subspaces of $V$.*
2. *The projections $\pi_i : V \to ker(p_i(T))$ is a polynomial in $T$.*
3. *If $U \subset V$ is $T$-invariant, then*

$$U = \oplus_{i=1}^{k}(U \cap ker(p_i(T)))$$

*Proof.* We begin with a few definitions and some notation. Define

$$\widehat{p_i} = \prod_{j=1, j\neq i}^{k} p_j$$

Note that, since the $p_i$ are relatively prime, the $\widehat{p_i}$ are relatively prime. Thus, there exist polynomials $q_i \in K[x]$ such that

$$q_1\widehat{p}_1 + \cdots + q_k\widehat{p}_k = 1$$

Now we are ready to establish the points of the theorem.

- The $ker(p_i(T))$ are $T$-invariant, since if $v \in ker(p_i(T))$ then $p_i(T)Tv = Tp_i(T)v = T0 = 0$, and so $Tv \in ker(p_i(T))$ too.

- $V$ is a sum of the $ker(p_i(T))$. Note that

  $$q_1(T)\widehat{p}_1(T) + \cdots + q_k(T)\widehat{p}_k(T) \;=\; I$$

  Thus, given any $v \in V$ we can write

  $$v \;=\; Iv \;=\; \sum_{i=1}^{k} q_i(T)\widehat{p}_i(T)v$$

  All we have to do now is to verify that $q_i(T)\widehat{p}_i(T)v \in ker(p_i(T))$. Well,

  $$p_i(T)q_i(T)\widehat{p}_i(T)v \;=\; q_i(T)p_i(T)\widehat{p}_i(T)v \;=\; q_i(T)p(T)v \;=\; q_i(T)0 \;=\; 0$$

  and we have shown $V = \sum_i ker(p_i(T))$.

- Now we have to show that the sum above is a direct sum. This involves showing (recall Midterm I) that the only element common to each $ker(p_i(T))$ and the sum of the remaining $ker(p_j(T))$'s is 0. **Equivalently**, (verify this!) one only has to see that

  $$v_1 + \cdots + v_k \;=\; 0$$

  and $v_i \in ker(p_i(T))$ implies that $v_i = 0$ for all $i$.

  We see this by the following pretty argument. Apply $\sum q_j(T)\widehat{p}_j(T) = I$ to $v_i$ to get

  $$q_i(T)\widehat{p}_i(T)v_i \;=\; Iv_i \;=\; v_i$$

  The other $q_j(T)\widehat{p}_j(T)v_i$ terms on the left side vanish since $v_i \in ker(p_i(T))$ and $p_i$ is a factor of $\widehat{p}_j$ when $j \neq i$.

  Now use the equation $v_1 + \cdots + v_k \;=\; 0$ to substitute in for $v_i$ as follows.

  $$v_i \;=\; q_i(T)\widehat{p}_i(T)v_i \;=\; q_i(T)\widehat{p}_i(T)\left(-\sum_{s\neq i} v_s\right)$$

  But this gives

  $$v_i \;=\; -\sum_{s\neq i} q_i(T)\widehat{p}_i(T)v_s \;=\; -\sum_{s\neq i} q_i(T)0 \;=\; -\sum_{s\neq i} 0 \;=\; 0$$

  since (as above) $\widehat{p}_i(T)v_s = 0$ whenever $i \neq s$. Therefore we have shown $v_i = 0$, and so the sum is direct. At this stage we have established point 1.

- 2 will follow once we convince ourselves that $Im(q_i(T)\widehat{p}_i(T))$ is the same as $ker(p_i(T))$ (since $q_i(T)\widehat{p}_i(T)$ is a polynomial in $T$). We have clearly seen above that $Im(q_i(T)\widehat{p}_i(T)) \subset$

$ker(p_i(T))$. We have also seen the reverse inclusion (where??) implicitly, but let's make it explicit here. If $v_i \in ker(p_i(T))$ then

$$v_i \;=\; Iv_i \;=\; \sum_j q_j(T)\widehat{p}_j(T)v_i$$

But we remember that if $i \neq j$ then $\widehat{p}_j(T)v_i \;=\; 0$, and so the sum on the right hand side reduces down to $q_i(T)\widehat{p}_i(T)v_i$. That is

$$v_i \;=\; q_i(T)\widehat{p}_i(T)v_i$$

and so $v_i \in Im(q_i(T)\widehat{p}_i(T))$.

- Finally for 3, suppose that $U$ is $T$-invariant. This means if $u \in U$, then $Tu \in U$, and more generally $f(T)u \in U$ for any polynomial $f \in K[x]$. By part 1 we have seen that each vector $u \in U \subset V$ can be expressed as a sum $u_1 + \cdots + u_k$ where each $u_i \in ker(p_i(T))$. By part 2, $u_i$ can be expressed as a polynomial in $T$ times $u$, and so also belongs to $U$ by $T$-invariance. Done!

□

**Examples 2.2.15.** Let's look at our motivating examples (class notes!) again. These are the projection operators. They satisfy $T^2 = T$, or in other words $T(T - I) = 0$. The Primary Decomposition Theorem tells us that the finite dimensional vector space $V$ on which $T$ acts decomposes as a sum

$$V \;=\; V_0 \oplus V_1 \;=\; ker(T) \oplus ker(T - I)$$

of 0- and 1-eigenspaces of $T$. Note that since

$$1 \;=\; 1(x) - 1(x - 1)$$

we have that $IT = T$ is the projection onto the 1-eigenspace, and that $-I(T - I) = (I - T)$ is the projection onto the 0-eigenspace. Thus the 1-eigenspace is the image of $T$ and is also the kernel of $(T - I)$, while the 0-eigenspace is the kernel of $T$ and is the image of $(I - T)$.

Here is another example. Suppose that $T^2 \;=\; I$. Then $(T - I)(T + I) \;=\; 0$ and so Primary Decomposition tells us that $V$ is a sum of the 1- and the $-1$-eigenspaces of $T$. There are three cases.

**Case 1.** The 1-eigenspace is all of $V$. Then $T = I$.

**Case 2.** The $-1$-eigenspace is all of $V$. Then $T = -I$ is a central symmetry through the origin.

**Case 3.** Both the 1- and the $-1$-eigenspaces are nonzero. Then $T$ is a reflection in the 1-eigenspace.

In view of the Primary Decomposition Theorem it makes good sense to look for the simplest possible polynomials (eg of lowest degree) which annihilate $T$. This motivates the following.

**Definition 2.2.16.** Let $V$ be a finite dimensional $K$-vector space and let $T \in \mathcal{L}(V)$. The *minimal polynomial* of $T$ is the unique monic (leading coefficient = 1) polynomial of minimal degree which annihilates $T$.

The next result says that this concept is indeed well-defined, and it establishes a nice relationship between the minimal polynomial and the characteristic polynomial.

**Lemma 2.2.17.** *Let $m$ be an annihilating polynomial of $T$ which has minimal degree. Then*

1. *$m$ divides evenly into every other annihilating polynomial of $T$. In particular, $m$ divides evenly into the characteristic polynomial of $T$. Also, the notion of* minimal polynomial *is well-defined.*

2. *The minimal polynomial of $T$ and the characteristic polynomial of $T$ have the same roots.*

*Proof.* Let $f$ be an annihilating polynomial for $T$. If $m$ does not divide evenly into $f$ we can find a polynomial $q$ and a nonzero polynomial $r$ such that

$$f \;=\; mq + r$$

Moreover, the degree of $r$ is strictly less than that of $m$. But $r(T) = f(T) - m(T)q(T) = 0$ and so $r$ is an annihilating polynomial of $T$ which has strictly smaller degree than $m$. This contradicts the minimality of the degree of $m$.

This has two neat consequences. The first is that $m$ divides the characteristic polynomial of $T$, since the characteristic polynomial annihilates $T$ by Hamilton-Cayley. The second consequence is the uniqueness of the minimal polynomial. If $m$ and $m'$ are two annihilating polynomials of $T$ with minimal degree, then $m$ divides $m'$ and $m'$ divides $m$ by the argument above. Thus, $m$ and $m'$ can only differ by at most a scalar multiple. Therefore we can uniquely define the minimal polynomial by deciding how to choose a scalar multiple. We do this by requiring that the leading coefficient of $m$ should be 1 ($m$ is called *monic*).

Now for part two. We've seen in part one that the minimal poly divides the char poly. Thus every root of the minimal poly is automatically a root of the char poly. So we have only to prove the reverse implication.

Suppose $\lambda \in K$ is a root of the char poly. Thus, $\lambda$ is an eigenvalue of $T$. That is, there exists a nonzero vector $v \in V$ such that $Tv = \lambda v$. Thus $T^j v = \lambda^j v$ and, more generally,

$$m(T)v \;=\; m(\lambda)v$$

where $m$ is the minimal poly. But $m(T) = 0$ and $v \neq 0$. Thus we must have $m(\lambda) = 0$, and so $\lambda$ is a root of $m$. □

**Remark 2.2.18.** Here is a direct proof of the fact that $m(k) = 0$ implies that $k$ is an eigenvalue of $T$ (and hence is a root of the char poly). Since $k$ is a root of $m$ we can write

$$m(x) \;=\; (x - k)q(x)$$

where $q$ has degree strictly less that the degree of $m$. By definition of minimal polynomial, this means that $q$ cannot annihilate $T$. Thus there exists a nonzero vector $w \in V$ such that $q(T)w \neq 0$. But

$$(T - kI)q(T)w \;=\; m(T)w \;=\; 0$$

and so $k$ is indeed an eigenvalue of $T$ with eigenvector $q(T)w$. □

Here is the characterization of diagonalizable operators as promised earlier.

**Theorem 2.2.19 (Characterization of diagonalizable).** *Let $V$ be a finite dimensional $K$-vector space and let $T \in \mathcal{L}(V)$. Then $T$ is diagonalizable if and only if the minimal polynomial of $T$ is a product of distinct linear factors.*

*Proof.* Suppose $T$ is diagonalizable. This means that there is a basis for $V$ with respect to which the matrix of $T$ is diagonal, with (repeated) eigenvalues along the diagonal. Thus the char poly is of the form

$$\prod_j (x - \lambda_j)^{d_j}$$

where the $j$ index runs over the set of distinct eigenvalues of $T$, and the $d_j$ are the number of times the $\lambda_j$ appears on the diagonal of $A$ which is the same as the dimension of the $\lambda_j$-eigenspace of $T$. It is clear (do the matrix multiplication!) that $A$ satisfies the polynomial

$$\prod_j (x - \lambda_j)$$

where the $j$ index is as above, but that it wont satisfy a polynomial of the form above which omits one of the $j$ indices. This must be the minimal poly of $T$, since the minimal poly is the monic poly of minimal degree which divides the char poly.

Conversely, Suppose the minimal poly of $T$ is a product

$$\prod_j (x - \lambda_j)$$

where the $\lambda_j$ are all distinct. The Primary Decomposition Theorem tells us that $V$ is a direct sum of the $ker(T - \lambda_j I)$. But each of these is a $\lambda_j$-eigenspace of $T$. Picking bases for each of these direct summands gives a basis of eigenvectors of $T$ for $V$. Thus $T$ is diagonalizable. □

Finally, here's a result about simultaneous diagonalization.

**Theorem 2.2.20 (Simultaneous diagonalization).** *Let $V$ be a finite dimensional $K$-vector space, and let $S, T \in \mathcal{L}(V)$ be diagonalizable. Then $S$ and $T$ are simultaneously diagonalizable if and only if $ST = TS$.*

*Proof.* Suppose $S$ and $T$ are simultaneously diagonalizable. This means there exists a basis $\mathcal{B}$ for $V$ with respect to which $S$ has matrix $A = diag(\lambda_1, \ldots, \lambda_n)$ and $T$ has matrix $B = diag(\mu_1, \ldots, \mu_n)$. Now

$$AB \;=\; \begin{pmatrix} \lambda_1\mu_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n\mu_n \end{pmatrix} \;=\; \begin{pmatrix} \mu_1\lambda_1 & & 0 \\ & \ddots & \\ 0 & & \mu_n\lambda_n \end{pmatrix} \;=\; BA$$

and so $ST = TS$.

On the other hand, suppose that $S$ and $T$ are diagonalizable, and that $ST = TS$. Since $S$ is diagonalizable, we may write

$$V \;=\; V_1 \oplus \cdots \oplus V_k$$

where the $V_i$ are eigenspaces of $S$. Since $ST = TS$, each of the $V_i$ are $T$-invariant. Here's the proof: $v \in V_i$ implies

$$S(Tv) \;=\; T(Sv) \;=\; T(\lambda_i v) \;=\; \lambda_i Tv$$

and so $Tv \in V_i$.

Now $T$ diagonalizable implies that

$$V \;=\; W_1 \oplus \cdots \oplus W_l$$

where each $W_i$ is an eigenspace of $T$, and the decomposition corresponds to a decomposition of the minimal polynomial for $T$ into linear factors as shown

$$(x - \mu_1) \cdots (x - \mu_l)\,.$$

Now, since each $V_i$ is $T$-invariant, the last part of the Primary Decomposition Theorem states that

$$V_i \;=\; \bigoplus_{j=1}^{l} (V_i \cap W_j)\,.$$

Thus

$$V \;=\; \bigoplus_{i=1}^{k} \bigoplus_{j=1}^{l} (V_i \cap W_j)$$

is a direct sum of intersections of eigenspaces of $S$ and $T$. Choosing bases for each of the $V_i \cap W_j$ and combining these together yields a basis for $V$ which consists of simultaneous eigenvectors of $S$ and of $T$. $\square$

## 2.3 Jordan and Rational Canonical Forms

In this section we address two problems that may prevent a linear operator from being diagonalizable. First, even though the characteristic polynomial may factor into a product of linear terms, the minimal polynomial may have some repeated roots. Thus the operator is not diagonalizable, and so does not have a diagonal matrix representative. The next best thing to a diagonal matrix is the so-called Jordan Form matrix. This is a lower-triangular matrix, consisting of eigenvalues on the main diagonal, 1's and 0's just below the diagonal, and 0's elsewhere. Secondly, the characteristic polynomial may not even factor into linear terms over the field of scalars $K$, and the minimal polynomial may have some high degree irreducible factors. In this case we can obtain a canonical matrix representation for the operator called the Rational Form.

Throughout this section $K$ is a field, $V$ is a finite dimensional $K$-vector space, and $T \in \mathcal{L}(V)$. We begin with the case where the characteristic polynomial of $T$ factors into linear terms, but $T$ is not diagonalizable. Before stating the existence result for the Jordan Canonical Form, we need a definition.

**Definition 2.3.1.** Let $\lambda \in K$ be a scalar. A *Jordan block of size* $m$ is an $m \times m$ matrix in $K^{m\times m}$ of the form

$$J(\lambda) \;=\; \begin{pmatrix} \lambda & & & 0 \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ 0 & & 1 & \lambda \end{pmatrix}$$

with $\lambda$'s on the diagonal, 1's just below the diagonal and 0's elsewhere.

**Examples 2.3.2.** Here are some examples.

1. A Jordan block $J(\lambda)$ of size 1 is just the following:

$$(\lambda)$$

2. Here is a Jordan block $J(3)$ of size 2:

$$\begin{pmatrix} 3 & 0 \\ 1 & 3 \end{pmatrix}$$

3. Here is a Jordan block $J(5)$ of size 3:

$$\begin{pmatrix} 5 & 0 & 0 \\ 1 & 5 & 0 \\ 0 & 1 & 5 \end{pmatrix}$$

**Remark 2.3.3.** Note that if $J(\lambda)$ is a Jordan block of size $m$ then $J(\lambda) - \lambda I_m$ is a nilpotent operator (whose $m$-th power is 0, but whose $(m-1)$-st power is non-zero).

**Theorem 2.3.4 (Jordan Forms: Existence).** *Let $K$ be a field and let $V$ be a finite dimensional $K$-vector space. Let $T \in \mathcal{L}(V)$. Then the following are equivalent.*

1. *the characteristic polynomial, p, of T factors into linear terms*
$$p(x) \;=\; \prod_{i=1}^{k}(x-\lambda_i)^{d_i}$$

2. *$V$ decomposes as a direct sum $V^{\lambda_1}\oplus\cdots\oplus V^{\lambda_k}$ where*
$$V^{\lambda_i} \;=\; ker(T-\lambda_i I)^{d_i}$$

3. *There is a basis for $V$ with respect to which $T$ has a block-diagonal matrix representation as shown:*
$$\begin{pmatrix} J(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & J(\lambda_k) \end{pmatrix}$$
*where for each eigenvalue $\lambda_i$ there may be several Jordan blocks $J(\lambda_i)$ of various sizes all along the diagonal.*

4. *There is a basis of $V$ with respect to which $T$ has a lower-triangular matrix as shown:*
$$\begin{pmatrix} * & & 0 \\ & \ddots & \\ * & & * \end{pmatrix}$$

The matrix representation in part 3 above is called the *Jordan canonical form* of $T$. We abbreviate this to JCF.

**Examples 2.3.5.** Here are some examples of JCF matrices.

1. The form
$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 1 & 3 \end{pmatrix}$$
has one Jordan block of size 1 and one of size 2. The char poly is
$$(x-3)^3$$
and the minimal poly is
$$(x-3)^2$$
The dimension of the 3-eigenspace is 2. This is not diagonalizable.

2. The form
$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

has 3 Jordan blocks of size 1. The char poly is

$$(x-3)^3$$

and the minimal poly is

$$(x-3)$$

The dimension of the 3-eigenspace is 3. This is diagonalizable.

3. The form

$$\begin{pmatrix} 3 & 0 & 0 \\ 1 & 3 & 0 \\ 0 & 1 & 3 \end{pmatrix}$$

has 1 Jordan block of size 3. The char poly is

$$(x-3)^3$$

and the minimal poly is

$$(x-3)^3$$

The dimension of the 3-eigenspace is 1. This is not diagonalizable.

4. The form

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 1 & 7 \end{pmatrix}$$

has 1 $J(3)$-block of size 2 and 1 $J(7)$ block of size 2. The char poly is

$$(x-3)^2(x-7)^2$$

and the minimal poly is

$$(x-3)^2(x-7)^2$$

Each of the 7- and 3-eigenspaces are 1-dimensional. This is not diagonalizable.

5. The form

$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 1 & 3 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}$$

has 1 $J(3)$-block of size 2 and 2 $J(7)$ blocks of size 1. The char poly is

$$(x-3)^2(x-7)^2$$

and the minimal poly is

$$(x-3)^2(x-7)$$

The 3-eigenspace is 1-dimensional, while the 7-eigenspace is 2-dimensional. This is not diagonalizable.

6. The form
$$\begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 7 \end{pmatrix}$$
has 2 $J(3)$-blocks of size 1 and 2 $J(7)$-blocks of size 1. The char poly is
$$(x-3)^2(x-7)^2$$
and the minimal poly is
$$(x-3)(x-7)$$
Each of the 7- and 3-eigenspaces are 2-dimensional. This is diagonalizable.

Here's a corollary of the JCF theorem.

**Definition 2.3.6.** Say that an operator $T \in \mathcal{L}(V)$ is *triangulable* if there exists a basis for $V$ with respect to which the matrix of $T$ is (lower) triangular.

**Corollary 2.3.7 (Triangulable operators).** *Let $V$ be a finite dimensional $K$-vector space. An operator $T \in \mathcal{L}(V)$ is triangulable if and only if its characteristic polynomial factors as a product of linear terms. In particular, if the field $K$ is algebraically closed, then every $T \in \mathcal{L}(V)$ is triangulable.*

The JCF matrix is uniquely determined by the operator $T$ (at least upto ordering of the $J(\lambda)$ along the diagonal) as the next result states.

**Theorem 2.3.8 (Jordan Forms: Uniqueness).** *Suppose $K$ is a field and $V$ is a finite dimensional $K$-vector space. Suppose also that $T \in \mathcal{L}(V)$ has a Jordan form matrix representative. Then the number and size of the Jordan blocks in this representative are determined by $T$. Specifically we have:*

1. *The types $J(\lambda)$ of the Jordan blocks are determined by the characteristic polynomial of $T$. The $\lambda$ are precisely the roots of this polynomial.*
2. *The number of Jordan blocks of type $J(\lambda)$ of size $m$ is equal to*
$$rank((T-\lambda I)^{m-1}) + rank((T-\lambda I)^{m+1}) - 2rank((T-\lambda I)^m)$$

*So, the operator $T$ uniquely determines the number, type and size of its Jordan blocks. That is, $T$ uniquely determines its JCF upto the order in which the Jordan blocks appear along the diagonal.*

The next result records some obvious connections between the Jordan canonical form of $T$ and properties of $T$.

**Theorem 2.3.9.** *Let $V$ be a finite dimensional $K$-vector space, and let $T \in \mathcal{L}(V)$ have a characteristic polynomial which factors into linear terms over $K$. Then the following are true for an eigenvalue $\lambda$ of $T$.*

1. *The number of Jordan blocks of type $J(\lambda)$ equals the dimension of the $\lambda$-eigenspace of $T$.*

2. *The size of the largest Jordan block of type $J(\lambda)$ equals the multiplicity of $\lambda$ as a root of the minimal polynomial.*

3. *The total number of occurrences of $\lambda$ in the JCF of $T$ equals the multiplicity of $\lambda$ as a root of the characteristic polynomial of $T$.*

If the characteristic polynomial of an operator $T \in \mathcal{L}(V)$ **does not** factor over the field $K$ into a product of linear terms (for example, $x^2 + 1$ over the field of real numbers) then we can still obtain a useful canonical matrix form for $T$ called the Rational Canonical Form. The idea is to use the polynomial to determine the matrix as follows.

Talk about this at a later date.

# Chapter 3

# Inner Product Spaces

Recall from Calculus III or from Math 3333 that you can compute the angle between two vectors $u = (u_1, \dots , u_n)$ and $v = (v_1, \dots , v_n)$ in $\mathbb{R}^n$ using the law of cosines as follows.

Start by observing that $u$, $v$ and $u - v$ are the sides of a triangle in $\mathbb{R}^n$. The law of cosines tells us that

$$||u - v||^2 \;=\; ||u||^2 + ||v||^2 - 2||u||||v|| \cos\theta$$

where $\theta$ is the angle between $u$ and $v$. We use the Pythagorean formula for the length of a vector in $\mathbb{R}^n$, and so get

$$(u_1 - v_1)^2 + \cdots + (u_n - v_n)^2 \;=\; (u_1^2 + \cdots + u_n^2) + (v_1^2 + \cdots + v_n^2) - 2\sqrt{(u_1^2 + \cdots + u_n^2)}\sqrt{(v_1^2 + \cdots + v_n^2)} \cos\theta$$

Squaring out the terms on the LHS, and simplifying gives us

$$-2(u_1v_1 + \cdots + u_nv_n) \;=\; -2||u||||v|| \cos\theta$$

or

$$\cos\theta \;=\; \frac{(u_1v_1 + \cdots + u_nv_n)}{||u||||v||}$$

So we see that the term in the numerator is very useful because

- It is easy to compute.
- It has a cool geometric interpretation: $||u||||v|| \cos\theta$.

It is called the *dot product* of the vectors $u$ and $v$, and is often denoted by $u \cdot v$. Some cool properties that it enjoys include:

- $u \cdot v = v \cdot u$ for all vectors $u$ and $v$.
- $(ku) \cdot v = k(u \cdot v)$ for all vectors $u$ and $v$ and all real numbers $k$.
- $(u + v) \cdot w = u \cdot w + v \cdot w$ for all vectors $u$, $v$ and $w$.
- $u \cdot u = ||u||^2 \geq 0$ and equals 0 if and only if $u = 0$.

We take this as our starting point for defining a *real inner product* on a real vector space and, by analogy, a *hermitian product* on a complex vector space.

## 3.1 Inner Product Spaces

**Definition 3.1.1.** Let $V$ be a real vector space. A *real inner product* on $V$ is a function

$$\langle\ ,\ \rangle : V \times V \to \mathbb{R} : (v, w) \mapsto \langle v, w\rangle$$

which satisfies

(i) $\langle u, v\rangle = \langle v, u\rangle$ for all $u, v \in V$

(ii) $\langle ku, v\rangle = k\langle u, v\rangle$ for all $u, v \in V$ and all $k \in \mathbb{R}$.

(iii) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$.

(iv) $\langle v, v\rangle \geq 0$, and $\langle v, v\rangle = 0$ if and only if $v = 0$, for all $v \in V$.

**Definition 3.1.2.** Let $V$ be a complex vector space, and let $\overline{z}$ denote the complex conjugate of $z \in \mathbb{C}$. A *hermitian product* on $V$ is a function

$$\langle\ ,\ \rangle : V \times V \to \mathbb{C} : (v, w) \mapsto \langle v, w\rangle$$

which satisfies

(i) $\langle u, v\rangle = \overline{\langle v, u\rangle}$ for all $u, v \in V$

(ii) $\langle ku, v\rangle = k\langle u, v\rangle$ for all $u, v \in V$ and all $k \in \mathbb{C}$.

(iii) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$.

(iv) $\langle v, v\rangle \geq 0$, and $\langle v, v\rangle = 0$ if and only if $v = 0$, for all $v \in V$.

**Examples 3.1.3.** Here are some examples of real and complex inner product spaces. Verify that they are indeed so.

1. The usual dot product on $\mathbb{R}^n$

2. The hermitian product on $\mathbb{C}^n$, defined by

$$\langle (z_1, \ldots, z_n), (w_1, \ldots, w_n)\rangle \ = \ \sum_{i=1}^{n} z_i \overline{w_i}$$

3. Let $\mathcal{C}([a, b], \mathbb{C})$ (respectively $\mathcal{C}([a, b], \mathbb{R})$) denote the complex (respectively real) vector space of continuous complex-valued (respectively real-valued) functions on the interval $[a, b]$. Then

$$\langle f, g\rangle \ = \ \int_a^b f(x)\overline{g(x)}\, dx$$

is a hermitian product (respectively real inner product).

4. $\langle (x, y, z), (a, b, c)\rangle = xa + 5yb + 3zc - 2(xb + ya)$ is a real inner product on $\mathbb{R}^3$!!

**Definition 3.1.4.** Let $V$ be a real or complex inner product space. We say that $u, v \in V$ are *orthogonal* if $\langle u, v\rangle = 0$.

**Lemma 3.1.5.** *If $v_1, \ldots, v_n$ are non-zero mutually orthogonal vectors in a real (or complex) inner product space $V$, then they are linearly independent.*

*Proof.* Exercise. □

The following definitions are direct generalizations of the Calc III definitions.

**Definition 3.1.6.** Define the *length of norm* of $v \in V$ a real (or complex) inner product space to be

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

**Definition 3.1.7.** Define the *angle* between two vectors $v, w \in V$ a real inner product space to be

$$\cos\theta \;=\; \frac{\langle v, w\rangle}{||u||||v||}$$

**Definition 3.1.8.** Let $u$ and $v$ be vectors in a real or complex inner product space $V$. The *projection* of $u$ on $v$ is denoted by $proj_v(u)$ and is defined as

$$proj_v(u) \;=\; \frac{\langle u, v\rangle}{\langle v, v\rangle}v$$

**Lemma 3.1.9.** *Let $u, v \in V$ be vectors in a real or complex inner product space, then $v$ and $u - proj_v(u)$ are orthogonal.*

*Proof.* Exercise. □

**Theorem 3.1.10 (Cauchy-Schwarz Inequality).** *Let $(V, \langle\,,\,\rangle)$ be a real or complex inner product space. For all $u, v \in V$ we have*

$$|\langle u, v\rangle|^2 \;\leq\; ||u||^2|v||^2$$

*In the real case, the equality*

$$\langle u, v\rangle \;=\; ||u|||v||$$

*holds if and only if $u = kv$ for some $k \geq 0$.*

*Proof.* Note that the equality holds if $v = 0$. So suppose that $v \neq 0$. Then the vector $proj_v(u)$ is well-defined and we can say

$$\begin{aligned}
0 \;&\leq\; \langle u - proj_v(u), u - proj_v(u)\rangle \\
&=\; \langle u, u\rangle - \langle u, proj_v(u)\rangle \\
&=\; \langle u, u\rangle - \langle u, \frac{\langle u, v\rangle}{\langle v, v\rangle}v\rangle \\
&=\; \langle u, u\rangle - \frac{\overline{\langle u, v\rangle}}{\langle v, v\rangle}\langle u, v\rangle \\
&=\; \langle u, u\rangle - \frac{|\langle u, v\rangle|^2}{\langle v, v\rangle}
\end{aligned}$$

Rearranging (remembering that $||u||^2 = \langle u, u\rangle$ and $||v||^2 = \langle v, v\rangle$) gives the desired result.

In the real case, if $u = kv$ for some $k \geq 0$ then we have

$$\langle u, v\rangle \;=\; \langle kv, v\rangle \;=\; k\langle v, v\rangle \;=\; k||v||||v|| \;=\; ||u||||v||$$

and so equality holds. Conversely, if equality holds, then we see that (from the proof) $u - proj_v(u) = 0$ and so $u = proj_v(u)$ is indeed a multiple of $v$. If $u = kv$ where $k < 0$ then

$$\langle u, v\rangle \;=\; \langle kv, v\rangle \;=\; k||v||^2 \;<\; 0 \;\leq ||u||||v|||$$

and so equality would not hold. Therefore $u$ must be a positive multiple of $v$. □

**Remark 3.1.11.** Note that the C-S inequality is not trivial to prove in the special cases of $\mathbb{R}^n$, $\mathbb{C}^n$ and $\mathcal{C}([a, b], \mathbb{C})$ with the standard inner products defined above. So you should definitely appreciate the generality, beauty and simplicity of the proof given above. Here are the three versions of C-S.

- $\left(\sum_{i=1}^n x_i y_i\right)^2 = \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{i=1}^n y_i^2\right)$ for $x_i, y_i \in \mathbb{R}$.
- $\left(\sum_{i=1}^n x_i \overline{y_i}\right)^2 = \left(\sum_{i=1}^n |x_i|^2\right)\left(\sum_{i=1}^n |y_i|^2\right)$ for $x_i, y_i \in \mathbb{C}$.
- $\left(\int_a^b f(x)\overline{g(x)}\, dx\right)^2 = \left(\int_a^b |f(x)|^2\, dx\right)\left(\int_a^b |g(x)|^2\, dx\right)$

**Theorem 3.1.12.** *Let $(V, \langle\, ,\, \rangle)$ be a real or complex inner product space. Then $||\,|| : V \to \mathbb{R} : v \mapsto \sqrt{\langle v, v\rangle}$ is a norm on $V$. That is it satisfies the following properties:*

1. *$||v|| \geq 0$ for all $v \in V$, and equality holds if and only if $v = 0$.*
2. *$||kv|| = |k|||v||$ for all $k \in \mathbb{R}$ (or $\mathbb{C}$) and all $v \in V$.*
3. *$||v + w|| \leq ||v|| + ||w||$ for all $v, w \in V$.*

*Proof.* Clearly, $||v|| \geq 0$. Now $||v|| = 0$ if and only if $\langle v, v\rangle = 0$ and this is true if and only if $v = 0$ by definition of inner product (positive definiteness).

For $k \in \mathbb{C}$ and $v \in V$ we have

$$||kv|| \;=\; \sqrt{\langle kv, kv\rangle} \;=\; \sqrt{k\overline{k}\langle v, v\rangle} \;=\; \sqrt{|k|^2||v||^2} \;=\; |k|||v||\,.$$

Finally, for $v, w \in V$ we have

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

where the last inequality follows from Cauchy-Schwarz. This proves the triangle inequality, and the theorem. □

**Definition 3.1.13.** Let $V$ be an inner product space. A basis $\{v_1, \ldots, v_n\}$ for $V$ is said to be *orthogonal* if

$$\langle v_i, v_j \rangle \;=\; 0 \text{ whenever } i \neq j$$

and is said to be *orthonormal* if

$$\langle v_i, v_j \rangle \;=\; \delta_{ij}$$

**Examples 3.1.14.** Standard basis on $\mathbb{R}^n$ is orthonormal.

**Theorem 3.1.15 (Gram-Schmidt).** *A finite dimensional inner product space has an orthonormal basis.*

*Proof.* See Math 3333 for the usual G-S orthonormalization process. Start from an arbitrary basis $\{v_1, \ldots, v_n\}$ and define

$$u_1 \;=\; \frac{v_1}{||v_1||}$$

and, inductively,

$$u_j \;=\; \frac{v_j - \sum_{i=1}^{j-1} proj_{u_i}(v_j)}{||v_j - \sum_{i=1}^{j-1} proj_{u_i}(v_j)||}$$

□

**Examples 3.1.16.** There are many instances of this in the literature.

1. Let $V$ be the subspace of $\mathcal{C}([-1,1],\mathbb{C})$ spanned by the polynomials $\{1, x, x^2, \ldots, x^n\}$, and equipped with the inner product

$$\langle f, g \rangle \;=\; \int_{-1}^{1} f(x)\overline{g(x)}\, dx$$

   Then the G-S process applied to the ordered basis $\{1, x, x^2, \ldots, x^n\}$ produces an orthonormal basis of polynomials called *Legendre polynomials*. Compute them!

2. Let $W$ be the subspace of $\mathcal{C}([-\pi,\pi],\mathbb{C})$ spanned by $\{e^{ikx} \mid -n \leq k \leq n\}$ and equipped with the inner product

$$\langle f, g \rangle \;=\; \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}\, dx$$

   Then $\{e^{ikx} \mid -n \leq k \leq n\}$ is an orthonormal basis for $W$. If $f \in \mathcal{C}([-\pi,\pi],\mathbb{C})$ then the orthogonal projection of $f$ onto $W$ is given by

$$\sum_{k=-n}^{n} c_k e^{ikx}$$

   where the coefficients $c_k = \int_{-\pi}^{\pi} f(x)e^{ikx}\, dx$ are called the *Fourier coefficients* of $f$.

The second example above generalizes. First we give a definition.

**Definition 3.1.17.** Let $V$ be an inner product space, and let $S \subset V$. The *orthogonal complement* of $S$ in $V$ is denoted by $S^{\perp}$ and is defined as

$$S^{\perp} \;=\; \{v \in V \mid \langle v, s\rangle = 0 \text{ for all } s \in S\,\}$$

If $W \subset V$ is a finite dimensional subspace, then the *orthogonal projection* $Pr_W$ is the unique linear operator in $\mathcal{L}(V)$ such that

$$Pr_W(w) \;=\; w \text{ for all } w \in W,$$

and

$$Pr_W(v) \;=\; 0 \text{ for all } v \in W^{\perp}.$$

**Lemma 3.1.18.** *Let $V$ and $W$ be as in the definition above. Then there exists a unique projection operator as asserted in the definition. Moreover, if $\{v_1, \dots, v_k\}$ is an orthonormal basis for $W$, then $Pr_W$ is given by*

$$Pr_W(v) \;=\; \langle v, v_1\rangle v_1 + \cdots + \langle v, v_k\rangle v_k$$

*Proof.* $Pr_W$ (as defined above) is clearly linear, and clearly acts as the identity on $W$ and as the zero transformation on $W^{\perp}$. So we see that *projection operators* exist.

Now for uniqueness. Let $T \in \mathcal{L}(V)$ be such that $T|_W = \mathbb{I}_W$ and $T|_{W^{\perp}} = 0$. Given any $v \in V$ we can write

$$v \;=\; w + (v - w) \;\in\; W + W^{\perp}$$

where $w = Pr_W(v)$. Thus (by linearity of $T$) we get

$$T(v) \;=\; T(w) + T(v-w) \;=\; w + 0 \;=\; w \;=\; Pr_W(v)$$

and we're done. □

**Remark 3.1.19.** The *Least Squares Approximation* technique of Math 3333 may be neatly phrased in terms of orthogonal complements and projection operators. Recall the setup. Suppose that the system

$$Ax \;=\; b$$

where $A \in \mathbb{C}^{m\times n}$, $x \in \mathbb{C}^{n\times 1}$, and $b \in \mathbb{C}^{m\times 1}$, does **not** have a solution. This means that $b \notin Im(A) = Col(A)$. So we find the nearest (or orthogonal) projection of $b$ onto $Im(A)$ and solve for that! This is called the least squares solution of the system $Ax = b$. It's not actually a solution, but it's the next best thing!

Here's a trick for finding the least squares solution. It relies on the following observation:

$$Im(A)^{\perp} \;=\; ker(A^T)$$

Now $v$ is a lest squares solution if and only if $Av = Pr_{Im(A)}b$. That is, if and only if $Av - b \in Im(A)^{\perp}$. Now $Av - b \in Im(A)^{\perp}$ if and only if $A^T(Av - b) = 0$ and this is true if and only if $v$ is a solution of the *consistent system*

$$A^TAx \;=\; A^Tb$$

So that's it. Simply multiply your inconsistent (no solutions) equation $Ax = b$ across by $A^T$ on the left, and solve the resulting consistent system.

## 3.2 Diagonalization and Spectral Theorem

**Definition 3.2.1.** There are special names given to the change of basis matrices between orthonormal bases in inner product spaces.

1. $A \in \mathbb{R}^{n\times n}$ is said to be *orthogonal* if
$$A^T A = I_n$$
That is, $A$ is invertible and $A^{-1} = A^T$.

2. $A \in \mathbb{C}^{n\times n}$ is said to be *unitary* if
$$\overline{A}^T A = I_n$$
That is, $A$ is invertible and $A^{-1} = \overline{A}^T$. We usually denote $\overline{A}^T$ by $A^*$.

It is easy to see that $A$ is orthogonal (resp. unitary) if and only if its rows (and likewise its columns) form an orthonormal basis for $\mathbb{R}^n$ with the usual dot product (resp. $\mathbb{C}^n$ with the usual hermitian product).

**Definition 3.2.2.** Let $V$ and $W$ be inner product spaces (either both real or both complex). We say that
$$T : V \to W$$
is an *isometry* if

- $T$ is an isomorphism of vector spaces
- $\langle Tu, Tv\rangle = \langle u, v\rangle$ for all $u, v \in V$.

**Lemma 3.2.3.** *Let $T : V \to W$ be a linear transformation of finite dimensional inner product spaces. Then the following are equivalent.*

1. *$T$ is an isometry*
2. *For any orthonormal basis $\{u_1, \dots, u_n\}$ for $V$, the set $\{Tu_1, \dots, Tu_n\}$ is an orthonormal basis for $W$.*
3. *There exists an orthonormal basis $\{u_1, \dots, u_n\}$ for $V$, such that the set $\{Tu_1, \dots, Tu_n\}$ is an orthonormal basis for $W$.*

*Proof.* $1 \to 2$ is immediate, as is $2 \to 3$ (just use G-S to obtain an orthonormal basis for $V$ and then apply 2). The work is in proving $3 \to 1$.

Given any $u, v \in V$ then we can write
$$u = \alpha_1 u_1 + \cdots + \alpha_n u_n$$
and
$$v = \beta_1 u_1 + \cdots + \beta_n u_n$$

where $\{u_1, \dots, u_n\}$ is our given orthonormal basis for $V$ with the property that $\{Tu_1, \dots, Tu_n\}$ is an orthonormal basis for $W$. Then we have

$$\begin{aligned}
\langle Tu, Tv\rangle &= \langle T(\alpha_1 u_1 + \cdots + \alpha_n u_n), T(\beta_1 u_1 + \cdots + \beta_n u_n)\rangle \\
&= \langle \sum_i \alpha_i Tu_i, \sum_j \beta_j Tu_j\rangle \\
&= \sum_{ij} \alpha_i \overline{\beta_j} \langle Tu_i, Tv_j\rangle \\
&= \sum_{ij} \alpha_i \overline{\beta_j} \delta_{ij} \\
&= \sum_{i,j} \alpha_i \overline{\beta_j} \langle u_i, v_j\rangle \\
&= \langle \sum_i \alpha_i u_i, \sum_j \beta_j u_j\rangle \\
&= \langle u, v\rangle
\end{aligned}$$

and so we see that $T$ is indeed an isometry. □

**Corollary 3.2.4.** *These are all immediate corollaries.*

1. *$T \in \mathcal{L}(\mathbb{R}^n)$ is an isometry with respect to the usual dot product if and only if its standard basis matrix representation $[T]_{St}$ is orthogonal.*
2. *$T \in \mathcal{L}(\mathbb{C}^n)$ is an isometry with respect to the usual hermitian product if and only if its standard basis matrix representation $[T]_{St}$ is unitary.*
3. *Two finite dimensional real inner product spaces are isometric if and only if they have the same dimension.*
4. *Two finite dimensional complex inner product spaces are isometric if and only if they have the same dimension.*

As promised we shall go after a diagonalization theorem. First of all we say what it means for an operator $T \in \mathcal{L}(V)$ on an inner product space to interact nicely with the inner product. We shall do this by first talking about the adjoint of an operator $T \in \mathcal{L}(V)$ on an inner product space.

**Definition 3.2.5.** Let $T \in \mathcal{L}(V)$ be an operator on an inner product space. The *adjoint* of $T$ is an operator in $\mathcal{L}(V)$, denoted by $T^*$, which is defined by

$$\langle T(u), v\rangle = \langle u, T^*(v)\rangle$$

for all $u, v \in V$.

**Lemma 3.2.6.** *The notion given above of the adjoint of an operator $T$ on an inner product space $V$ is well-defined. Moreover, if $T$ has matrix representative $A$ with respect to some orthonormal basis for $V$, then $T^*$ has the matrix representative $A^*$ with respect to the same basis.*

*Proof.* Let's show linearity of $T^*$. If $v_1, v_2 \in V$ then we have

$$\begin{aligned} \langle u, T^*(v_1 + v_2)\rangle &= \langle T(u), v_1 + v_2\rangle \\ &= \langle T(u), v_1\rangle + \langle T(u), v_2\rangle \\ &= \langle u, T^*(v_1)\rangle + \langle u, T^*(v_2)\rangle \\ &= \langle u, T^*(v_1) + T^*(v_2)\rangle \end{aligned}$$

holding for all $u \in V$. In particular, this holds for all $u$ in some orthonormal basis $\mathcal{B}$ for $V$. This means that the coefficients of $T^*(v_1 + v_2)$ with respect to $\mathcal{B}$ all agree with the coefficients of $T^*(v_1) + T^*(v_2)$ with respect to $\mathcal{B}$, since we've just shown that the complex conjugates of these coefficients agree. Thus

$$T^*(v_1 + v_2) \;=\; T^*(v_1) + T^*(v_2)$$

and so $T^*$ respects addition. Likewise (exercise) you can show that $T^*$ respects scalar multiplication. Thus, $T^*$ is linear.

Finally, let $\mathcal{B} = \{u_i\}$ be an orthonormal basis for $V$. Suppose that the matrix of $T$ w.r.t. $\mathcal{B}$ is $A$. This means

$$A_{ij} \;=\; \langle T(u_j), u_i\rangle$$

(why not $\langle u_i, T(u_j)\rangle$?) and so if $B$ is the matrix for $T^*$ we have

$$B_{ij} \;=\; \langle T^*(u_j), u_i\rangle \;=\; \overline{\langle u_i, T^*(u_j)\rangle} \;=\; \overline{\langle T(u_i), u_j\rangle} \;=\; \overline{A_{ji}}$$

and so $B = A^*$ as required. □

**Definition 3.2.7.**

- Say that a linear operator $T : V \to V$ of a real or complex inner product space is *self-adjoint* if

  $$\langle Tu, v\rangle \;=\; \langle u, Tv\rangle$$

  for all $u, v \in V$. By the previous lemma/definition, this is just the same as saying that $T$ is equal to its own adjoint.

- Say that a matrix $A \in K^{n\times n}$ ($K = \mathbb{R}$ or $\mathbb{C}$) is *self-adjoint* if

  $$A^* \;=\; A$$

  In the real case this becomes $A^T = A$ and we call the matrix *symmetric*, and in the complex case this is still $A^* = A$ and we call the matrix *hermitian*.

**Remark 3.2.8.** Note that an operator $T$ is self-adjoint if and only if its matrix w.r.t. an orthonormal basis is a self-adjoint matrix.

Here's our first cool theorem.

**Theorem 3.2.9.** *Every self-adjoint operator $T \in \mathcal{L}(V)$ on a finite dimensional inner product space $V$ has a real eigenvalue. In fact, all eigenvalues of $T$ are real. Moreover, eigenspaces with distinct eigenvalues are orthogonal.*

*Proof.* Since finite dimensional inner product spaces of the same dimension are isometric, it suffices to consider a self-adjoint operator on some $\mathbb{C}^n$ or $\mathbb{R}^n$.

Given a self-adjoint operator $T$ on $\mathbb{R}^n$ we can consider it as a self-adjoint operator on $\mathbb{C}^n$. In this case we know that the characteristic polynomial of $T$ factors as a product of linear terms. Thus there are eigenvalues. We just have to show that they all must be real.

Let $\lambda \in \mathbb{C}$ be an eigenvalue of $T$ with nonzero eigenvector $v$. Then we have

$$\begin{aligned}
\lambda\langle v, v\rangle &= \langle \lambda v, v\rangle \\
&= \langle T(v), v\rangle \\
&= \langle v, T^*(v)\rangle \\
&= \langle v, T(v)\rangle \\
&= \langle v, \lambda v\rangle \\
&= \overline{\lambda}\langle v, v\rangle
\end{aligned}$$

Now, $v \neq 0$ implies that $\langle v, v\rangle \neq 0$ (positive definiteness), and so $\lambda = \overline{\lambda}$. Thus $\lambda \in \mathbb{R}$ and we're done.

Finally, if $u$ and $v$ are eigenvectors of $T$ corresponding to distinct eigenvalues $\lambda$ and $\mu$ respectively, then we have

$$\lambda\langle u, v\rangle \;=\; \langle \lambda u, v\rangle \;=\; \langle T(u), v\rangle \;=\; \langle u, T(v)\rangle \;=\; \langle u, \mu v\rangle \;=\; \mu\langle u, v\rangle$$

Thus $(\lambda - \mu)\langle u, v\rangle = 0$ and, since $\lambda \neq \mu$, we get $\langle u, v\rangle = 0$. Done! □

**Examples 3.2.10.** Note that the assumption of finite dimensionality is crucial here. For example the *multiplication by x* operator

$$M : \mathcal{C}([a, b], \mathbb{C}) \to \mathcal{C}([a, b], \mathbb{C}) : f \mapsto M(f)$$

where $M(f)(x) = xf(x)$ for all $x \in [a, b]$, is clearly self-adjoint with respect to the usual inner product

$$\langle f, g\rangle \;=\; \int_a^b f(x)\overline{g(x)}\, dx$$

but does not have any eigenvalues.

Now we're ready for our main diagonalization theorem. It is one form of the spectral theorem.

**Theorem 3.2.11 (Spectral Theorem I).** *Let $T \in \mathcal{L}(V)$ be a self-adjoint operator on a finite dimensional inner product space. Then $V$ has an orthonormal basis of eigenvectors of $T$ with real eigenvalues.*

**Corollary 3.2.12.** *If $A \in \mathbb{R}^{n\times n}$ is symmetric, then there exists an orthogonal matrix $P$ such that $PAP^{-1} = PAP^T$ is diagonal.*

*If $A \in \mathbb{C}^{n\times n}$ is hermitian, then there exists a unitary matrix $U$ such that $UAU^{-1} = UAU^*$ is diagonal with real entries.*

*Proof of Theorem.* Proof is by induction on $n = dim_K(V)$. The case $n = 1$ is trivial. Suppose theorem holds for self-adjoint operator on inner product spaces of dimension $n-1$, and let $V$ have dimension $n$.

Now we know $T$ has a non-zero eigenvector $u$ with real eigenvalue $\lambda$. Let $W$ be the one-dimensional space spanned by $u$ and let $W^{\perp}$ be its $(n-1)$-dimensional orthogonal complement. If $v \in W^{\perp}$, then we have

$$\langle u, T(v)\rangle \;=\; \langle T(u), v\rangle \;=\; \langle \lambda u, v\rangle \;=\; \lambda\langle u, v\rangle \;=\; \lambda 0 \;=\; 0$$

and so $T(v) \in W^{\perp}$. Thus the operator $T$ restricts to $W^{\perp}$ to give a self-adjoint operator on an $(n-1)$-dimensional inner product space. By the inductive hypothesis we know that $W^{\perp}$ has an orthonormal basis of eigenvectors of $T|_{W^{\perp}}$ with real eigenvalues. Adding in the vector $\frac{u}{||u||}$ gives an orthonormal basis for $V$ which is comprised of eigenvectors of $T$ with real eigenvalues. Done. □

Here's the more standard statement of the Spectral Theorem (for self-adjoint operators).

**Theorem 3.2.13 (Spectral Theorem).** *Let $T \in \mathcal{L}(V)$ be a self-adjoint operator on an inner product space $V$. Then there exist mutually orthogonal subspaces $W_1, \dots, W_k$ of $V$ together with real numbers $\lambda_1, \dots, \lambda_k$ such that*

$$T \;=\; \sum_i \lambda_i Pr_{W_i}$$

*and*

$$\mathbf{I} \;=\; \sum_i Pr_{W_i}$$

*Proof.* Let the $\lambda_i$ be the eigenvalues of $T$, and let the $W_i$ be the corresponding eigenspaces. □

**Remark 3.2.14.** If you are just interested in diagonalization of operators on inner product spaces, and do not require that the eigenvalues be real, then we see that a necessary condition for diagonalization is that $T$ should commute with its adjoint $T^*$. In fact, this is also a sufficient condition for diagonalization of $T$. We call an operator $T$ which satisfies this condition

$$TT^* \;=\; T^*T$$

a *normal operator.* The more general form of the Spectral theorem then reads as follows.

**Theorem 3.2.15.** *Let $T$ be a normal operator on a finite dimensional complex inner products space, or a self-adjoint operator on a finite dimensional real inner product space. Then $V$ has an orthonormal basis of eigenvectors of $T$.*

# Chapter 4

# Miscellaneous Topics

## 4.1 Introduction to Linear Groups and Geometry

**Definition 4.1.1.** Let $K$ be a field. The *general linear group* of $(n \times n)$-matrices over $K$ consists of the set of all invertible $(n \times n)$-matrices with values in $K$ under multiplication. The *special linear group* $SL(n, K)$ is the subgroup of $GL(n, K)$ consisting of all matrices with determinant 1.

**Definition 4.1.2.** The *projective special linear groups* $PSL(n, K)$ are defined by projectivization as follows.

GIVE DEFINITION

What has this all got to do with *projective geometry*?

**Definition 4.1.3.** A *linear group* is a subgroup of the general linear group $GL(n, K)$.

**Definition 4.1.4.** The *classical groups* consist of the *orthogonal*, *unitary*, and *symplectic* groups.

These groups are defined as stabilizers of various elements of $K^{n\times n}$ under various actions of $GL(n, K)$ or $SL(n, K)$.

First we consider the *change of basis in a bilinear form* action, which is defined by

$$GL(n,K) \times K^{n\times n} \;\rightarrow\; K^{n\times n}$$

$$(P, A) \;\mapsto\; P^T AP$$

**Definition 4.1.5.** The *orthogonal group* $O(n, K)$ is defined to be the stabilizer of the identity matrix $I \in K^{n\times n}$

$$O(n,k) \;=\; Stab(I) \;=\; \{P \in GL(n,K) \,|\, P^T P = I\}$$

The *special orthogonal group* $SO(n, K)$ is just defined as

$$SO(n,K) \;=\; SL(n,K) \cap O(n,K)\,.$$

Important cases when $K = \mathbb{R}$ or $\mathbb{C}$.

**Definition 4.1.6.** $O(p, q, K)$ is the stabilizer of the signature $(p, q)$ form, $I_{p,q}$.

$$O(p,q,K) \;=\; Stab(I_{p,q}) \;=\; \{P \in GL(p+q,K) \,|\, P^T I_{p,q} P = I_{p,q}\}$$

and $SO(p, q, K) = O(p, q, K) \cap SL(p + q, K)$.

In the case $K = \mathbb{R}$, $p = n$ and $q = 1$, we get the Lorentz groups denoted by $O(n, 1)$ for short.

**Definition 4.1.7.** The *symplectic group* $SP(2n, K)$ (usually $K = \mathbb{R}$ or $\mathbb{C}$) is defined as

$$SP(2n,K) \;=\; Stab(J) \;=\; \{P \in GL(2n,K) \,|\, P^T JP = J\}$$

where

$$J \;=\; \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}.$$

It is an exercise to see that the symplectic matrices already have determinant 1, so we don't get anything new by intersecting with $SL(n, K)$.

For the next class of groups we shall restrict to the case $K = \mathbb{C}$ and consider that action of $GL(n, \mathbb{C})$ on $\mathbb{C}^{n\times n}$ by the change of basis for hermitian forms

$$\begin{aligned} GL(n, \mathbb{C}) \times \mathbb{C}^{n\times n} &\to \mathbb{C}^{n\times n} \\ (P, A) &\mapsto P^* AP \end{aligned}$$

where $P^*$ denotes the conjugate-transpose of $P$.

**Definition 4.1.8.** The *unitary groups* $U(n)$ are defined as

$$U(n) = Stab(I) = \{P \in GL(n, \mathbb{C}) \,|\, P^* P = I\}$$

and the *special unitary groups* are defined as one would expect $SU(n) = U(n) \cap SL(n, \mathbb{C})$.

We shall discover some beautiful relationships between $SU(2)$, $SO(3)$, the 3-sphere $S^3$ and the quaternions in section **??**, and between $PSL(2, \mathbb{R})$, $PSL(2, \mathbb{C})$, $SO(n, 1)$ and hyperbolic geometry in section **??**.