Linear Algebra

Muchang Bahng

Spring 2020

Contents

A multiple-year course in linear algebra at the advanced undergraduate and graduate level. The notes for this section can be a bit too abstract for someone learning linear algebra for the first time, so I suggest learning about groups, rings, and fields first.

1 Vector Spaces

We begin with an axiomatic treatment of a vector space, which is one of the primary objects of study in linear algebra. Recall what a field is from algebra.

Before we start, note that when students learn about vectors for the first time, they are presented with two interpretations of vectors. First, we think of a vector as an arrow pointing in space, encoding both a *magnitude* and a *direction*. This is more of the physicists interpretation and is useful in visualizing these vectors in low dimensions (notably, 2 and 3).

(a) Addition is visualized using the "parallelogram rule." (b) Scalar multiplication is simply an stretch of the vector.

The other interpretation is to simply see it as a tuple of numbers, which is useful when dealing with arbitrary dimensions. This interpretation is more suited to the intuition of a computer scientist, who can interpret vectors as arrays with element-wise operations.

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}$$

(a) Addition is done component-wise.

$$c \cdot \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} ca_1 \\ ca_2 \\ \vdots \\ ca_n \end{pmatrix}$$

(b) Scalar multiplication is done component-wise.

Figure 2

By introducing a coordinate basis (also called a frame), we can unify the two interpretations above by decomposing each arrow into its component vectors, with magnitudes equal to the elements of the corresponding tuple. There are other ways that vectors are introduced.[1] Mathematicians try to unify the two[2] by generalizing both of these interpretations into the following definition.

Definition 1.1 (Vector Space)

A **vector space** V **over a field** $\mathbb{F}$ is a set endowed with two operations—addition $+ : V \times V \to V$ and scalar multiplication $\cdot : \mathbb{F} \times V \to V$. It satisfies the following properties.

1. *Group Structure.* $(V, +)$ is an abelian group.
2. *Scalar Distributivity.* For $\lambda, \mu \in \mathbb{F}$ and $v, u \in V$, $(\lambda + \mu)v = \lambda v + \mu v$
3. *Vector Distributivity.* For $\lambda \in \mathbb{F}$ and $v, u \in V$, $\lambda(v + u) = \lambda v + \lambda u$
4. *Associativity of Scalar Multiplication.* For $\lambda, \mu \in \mathbb{F}$ and $v \in V$, $(\lambda\mu)v = \lambda(\mu v) = \mu(\lambda v)$

A **vector** is an element of a vector space.

We define some more vector spaces that are often used.

[1] Aspinwall mentioned that there are 4 or 5 ways? Another being in terms of derivations?
[2] similar to what 3Blue1Brown says in his Linear Algebra Youtube series.

Example 1.1 ($\mathbb{R}^n$)

The set of all n-tuples of real numbers $\mathbb{R}^n = \{(a_1, \ldots, a_n) : a_i \in \mathbb{R}\}$ forms a vector space over $\mathbb{R}$ with component-wise addition and scalar multiplication. This is the foundational setting for analytic geometry, physics (mechanics, electromagnetism), engineering, computer graphics, machine learning, and economics.

Example 1.2 ($\mathbb{C}^n$)

The set of all n-tuples of complex numbers $\mathbb{C}^n = \{(a_1, \ldots, a_n) : a_i \in \mathbb{C}\}$ forms a vector space over $\mathbb{C}$ (and also over $\mathbb{R}$, with dimension $2n$). Essential in quantum mechanics, signal processing (Fourier analysis), electrical engineering, and solving differential equations.

Example 1.3 ($\mathbb{Z}_p^n$)

The set of all n-tuples of elements in the finite field $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$ (with p prime) forms a vector space over $\mathbb{Z}_p$ with component-wise operations. Central to cryptography (elliptic curve cryptography, Reed-Solomon codes), coding theory, and computer science.

Example 1.4 (Function Space)

Let X be any set. The set of all functions $f : X \to \mathbb{F}$ from X to a field $\mathbb{F}$ forms a vector space over $\mathbb{F}$ with pointwise addition $(f + g)(x) = f(x) + g(x)$ and scalar multiplication $(c \cdot f)(x) = c \cdot f(x)$. The foundational object of functional analysis; appears in probability theory (random variables), signal processing, and differential equations.

Example 1.5 (Continuous Functions)

The set $C(I)$ of all continuous real-valued functions on an interval $I \subset \mathbb{R}$ forms a vector space over $\mathbb{R}$ with pointwise operations. Fundamental in real analysis, differential equations, Fourier series, and physics (modeling continuous phenomena).

Example 1.6 (Polynomials of Finite Degree)

The set of all polynomials of degree strictly less than n with coefficients in $\mathbb{F}$ defines a vector space over $\mathbb{F}$. Widely used in numerical analysis (interpolation), approximation theory, and signal processing.

Example 1.7 (Sequence Space ℓ^p)

The set of all sequences $(a_1, a_2, \ldots)$ of elements in $\mathbb{F}$ such that $\sum_{i=1}^{\infty} |a_i|^p < \infty$ forms a vector space over $\mathbb{F}$ for $1 \leq p < \infty$. Core example in functional analysis; arises in statistics (finite variance sequences), signal processing, and probability theory.

The examples above should convince you that linear algebra is fundamental in almost every quantitative field, to the point where it should be considered universal knowledge.

Theorem 1.1 (No Zero Divisors)

There are no zero divisors of vector space V. That is,

$$\lambda v = 0 \implies \lambda = 0 \text{ or } v = 0 \tag{1}$$

Proof. $\lambda v = 0 \implies \lambda v + \lambda v = 0 + \lambda v \implies 2\lambda v = \lambda v \implies (2\lambda - \lambda)v = 0$. But $\lambda \neq 0$, so v must equal 0. This leads to a contradiction.

Now we move onto the other primary object of study: the linear map.

Definition 1.2 (Linear Map)

Given vector spaces V, W over the same field $\mathbb{F}$, a function $T : V \longrightarrow W$ is called a **linear map** if it satisfies the following linearity properties:

$$T(v_1 + v_2) = T(v_1) + T(v_2), \quad T(cv) = cT(v) \tag{2}$$

for all $v_1, v_2 \in V, c \in \mathbb{F}$.

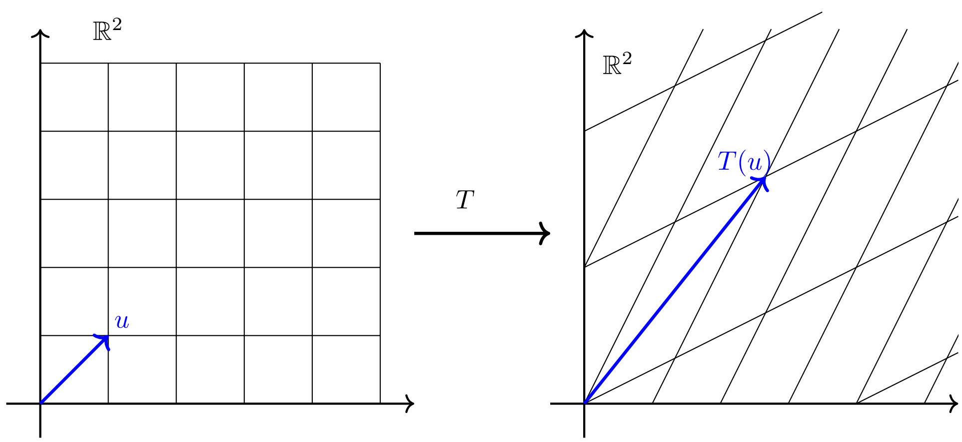


Figure 3: A great way to visualize this is that linear maps transform *lines to lines*. That is, we can think of the grid moving in a uniform way so that all squares become parallelograms. 3Blue1Brown has a really nice animation in his Linear Algebra Youtube Series.

Definition 1.3 (Isomorphism)

If a linear map $T : V \to W$ is called a **(vector space) isomorphism** if T is bijective. In this case, T^{-1} is also a linear map. Then, we say V is **isomorphic** to W, denoted $V \simeq W$.

Proof. It suffices to show that bijection $\iff$ the inverse is linear.

Following the terminology in algebra, we sometimes call a linear map $T : V \to V$ an *endomorphism* and an isomorphism $T : V \to V$ an *automorphism.*

1.1 Basis and Dimension

While there are a lot of things we can do with the axioms themselves, it is often convenient to add in a bit of structure to our vector space. First, note that we can write out vectors as sums of other vectors.

Definition 1.4 (Linear Combination)

Given vector space V, a **linear combination** of a finite[a] number of vectors $v_1, v_2, \ldots, v_n \in V$ is a vector of the form

$$v = c_1 v_1 + c_2 v_2 + \ldots + c_n v_n \tag{3}$$

where $c_1, \ldots, c_n \in \mathbb{F}$.

[a] We restrict ourselves to the finite case since we do not know how to evaluate series yet.

Definition 1.5 (Span)

The **span** of a collection of vectors $v_1, v_2, \ldots, v_n \in V$ is defined in the following equivalent ways:

1. *Smallest Subspace.* It is the intersection of all subspaces $W \subset V$ that each contains all $v_1, \ldots, v_n$.[a]
2. *Set of Linear Combinations.* It is the subspace of vectors defined

$$\text{span}\{v_1, v_2, \ldots, v_n\} \coloneqq \{c_1 v_1 + c_2 v_2 + \ldots + c_n v_n \mid c_1, \ldots, c_n \in \mathbb{F}\} \tag{4}$$

[a] This is the more general definition and extends to the infinite-dimensional case.

Proof.

It clearly follows that $v_1, \ldots, v_n$ span the whole space V if every vector in V can be expressed as a linear combination of the v_i's.

Definition 1.6 (Linear Independence)

Vectors $v_1, \ldots, v_n$ are

1. **linearly independent** if

$$c_1 v_1 + c_2 v_2 + c_3 v_3 + \ldots + c_n v_n = 0 \implies c_1, \ldots, c_n = 0 \tag{5}$$

2. **linearly dependent** if it is not linearly independent, i.e. there exists some $1 \leq i \leq n$ such that v_i is in the span of vectors $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$.

Definition 1.7 (Basis)

A set of linearly independent vectors $v_1, \ldots, v_n$ that span vector space V is called a **basis** of V. These vectors v_i are called **basis vectors**. Note that this basis is not unique; it is actually highly un-unique.

Example 1.8 (Standard Basis)

The basis e_i of $\mathbb{F}^n$ are the vectors with every element equal to 0 except for the ith element, which is equal to 1.

Theorem 1.2 (Maximally Independent Set is a Basis)

Let V be a vector space and S be a subset of V (not necessarily a subspace). Then, any maximal linearly independent subset $\{e_1, e_2, \ldots, e_k\}$ of a set S is a basis of $\operatorname{span} S$.

Definition 1.8 (Dimension)

The number of vectors in any basis of vector space V is called the **dimension** of V, denoted $\dim V$.

Proof. We must show that this is well defined by proving that every possible basis of a vector space V has the same number of vectors.

Theorem 1.3 (Isomorphism to $\mathbb{F}^n$)

Every n-dimensional vector space V over $\mathbb{F}$ is isomorphic to $\mathbb{F}^n$, the set of n-tuples of elements in $\mathbb{F}$.

Corollary 1.4 (Isomorphism Between n-Dimensional Vector Spaces)

Finite-dimensional vector spaces of the same field are isomorphic if and only if their dimensions are the same.

Example 1.9

The field of complex numbers $\mathbb{C}$, regarded as a vector space over $\mathbb{R}$, has dimension 2.

1.2 Subspaces

Definition 1.9 (Subspace)

Given vector space V, A subset $W \subset V$ is a **subspace** if sums and scalar multiples of elements of W belong to W.

Note that $\{0\}, V$ are subspaces of V. Regarding notation, when we write $W \subset V$ from now on, this is in the subspace sense, and not the set-theoretic subset.

Definition 1.10 (Hyperplane)

A $(n-1)$-dimensional subspace of an n-dimensional space is called a **hyperplane**.

Definition 1.11 (Sum of Subspaces)

The **sum of subspaces** $U_1, U_2, \ldots, U_n \subset V$ is defined

$$\sum_{i=1}^n U_i := \Big\{ \sum_{i=1}^n u_i \mid u_i \in U_i \Big\} \tag{6}$$

That is, it is the set of all vectors that can be expressed as the sum of vectors in each of its respective space.

Definition 1.12 (Direct Sum of Spaces)

Given subspaces $U_1, \ldots, U_n \subset V$, if V_i's are pairwise disjoint—except at the origin—then we call their sum the **direct sum**.

$$\bigoplus_{i=1}^{n} V_i \equiv \Big\{ \sum_{i=1}^{n} v_i \mid v_i \in V_i \Big\} \tag{7}$$

It is a vector space where each vector can be expressed uniquely as the sum of vectors in each respective subspace.

Proof.

The crucial difference between the sum and the direct sum is that the direct sum requires the subspaces to be disjoint except for at the origin, which allows the expression of each vector in $V_1 \oplus \ldots \oplus V_n$ to be unique. It is also worth noting that the Cartesian product of vector spaces is merely just the set of tuples of vectors that are in each respective space and is *not* a vector space (since addition and multiplication is not defined on that new set). If we define the operations component-wise, then

$$\prod_{i=1}^{n} V_i = \sum_{i=1}^{n} V_i \tag{8}$$

Example 1.10 (Vector Space as Direct Sum of Basis)

Note that we can also define the direct sum of spaces U and V by their basis. That is, given that the basis for U is $\{e_i\}_{i=1}^n$ and the basis for V is $\{f_j\}_{j=1}^n$, the basis for $U \oplus V$ is

$$\{(e_1, 0), (e_2, 0), \ldots, (e_n, 0), (0, f_1), \ldots, (0, f_m)\} \tag{9}$$

Example 1.11 (Complex Vector Spaces as Real Vector Spaces)

Vector spaces over one field can be interpreted as a vector space over another field. This is most common when interpreting complex vector spaces as real ones. For example, given a complex vector space Z with basis $\{z_1, z_2, \ldots, z_n\}$, every vector can be expressed as

$$z = \sum_{j=1}^{n} c_j z_j, \; c_j \in \mathbb{C} \tag{10}$$

We can set $c_j = a_j + b_j i$ uniquely, with $a, b \in \mathbb{R}$, and rewrite

$$z = \sum_{j=1}^{n} a_j z_j + b_j(i z_j) \tag{11}$$

$\implies \{z_j\} \cup \{i z_j\}$ forms a basis of Z as a *real vector space.*

Theorem 1.5 (Dimension of Direct Sum)

The dimension of the direct sum of vector spaces is

$$\dim \bigoplus_{i=1}^{n} V_i = \sum_{i=1}^{n} \dim V_i \tag{12}$$

Proof. This follows from the basis construction of the direct sum of V_i's.

Definition 1.13 (Congruence Relations on Vector Spaces)

Given vector space V and subspace $W \subset V$ we say that two vectors $v_1, v_2 \in V$ are **congruent modulo** W, denoted $v_1 \equiv v_2 \pmod{W}$, if $v_2 - v_1 \in W$. This is a congruence relation.

Proof. It remains to prove that $\equiv$ is a congruence relation. The congruence classes $\{y\}$ is the set of all vectors that are congruent modulo W to y.

Definition 1.14 (Quotient Vector Space)

The **quotient (vector) space** V modulo W, denoted V/W, is the set of all congruence classes modulo W. We can define addition and scalar multiplication on this set as such

$$[x] + [y] = [x + y] \tag{13}$$

Theorem 1.6 (Decomposition into Subspace and Quotient Space)

Given vector space V, W a subspace of V. Then,

$$V \simeq W \oplus \frac{V}{W} \tag{14}$$

Proof.

1.3 Dual Spaces

Now we talk about dual spaces, which are nice since their construction and existence doesn't require us to have any structure on the vector space.

Definition 1.15 (Dual Space)

Given a vector space V over $\mathbb{F}$, the **dual vector space** $V^\dagger$ is defined as the vector space of linear maps $\ell : V \rightarrow \mathbb{F}$. Given $\ell_1, \ell_2 \in V^\dagger$, $v \in V$, and $c \in \mathbb{F}$, we define

1. *Addition.* $(\ell_1 + \ell_2)(v) = \ell_1(v) + \ell_2(v)$.
2. *Scalar Multiplication.* $(c\ell_1)(v) = c\ell_1(v)$.

Proof. We must prove that this is indeed a vector space.

As the name suggests, we call the vectors of the dual space as *dual vectors*. Often, mathematicians will call them *functionals*, since usually the base space V is a vector space of functions, and functionals are mappings that act on functions. Analysis of this in the infinite-dimensional case requires much more machinery, so we will focus on the finite-dimensional case.

Theorem 1.7 (Dimensions of Dual of Finite-Dimensional Vector Spaces)

Let V be a finite-dimensional vector space and V^* its dual. Then,

$$\dim V = \dim V^\dagger \tag{15}$$

Proof.

Definition 1.16 (Dual Basis)

Given a basis $\{e_1, e_2, \ldots, e_n\}$ of V, the **dual basis** $\{f_1, f_2, \ldots, f_n\}$ of $V^\dagger$ has vectors satisfying

$$f_j(e_i) = \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{16}$$

where δ_{ij} is called the **Kronecker delta function**.

While we initially view elements of V as "things" and elements of $V^\dagger$ as linear functions, this thought is actually erroneous. Given $l \in V^\dagger$, we see that

$$l : V \longrightarrow \mathbb{F} \tag{17}$$

But since both V and $V^\dagger$ are vector spaces, we can also see that given $x \in V$, x is also a linear function

$$x : V^\dagger \longrightarrow \mathbb{F}, \text{ where } x(l) \equiv l(x) \tag{18}$$

But this means that x is an element of $V^{\dagger\dagger}$ the dual of V! This statement is elaborated with the following theorem.

Theorem 1.8 (Canonical Isomorphisms of Double Duals)

$V^{\dagger\dagger}$ is **naturally isomorphic**[a] to vector space V.

[a]What we mean by natural is that we do not need to select a basis in either vector space to define the isomorphism.

Proof. We fix a vector $l \in V^\dagger$, and given $x \in V, \phi \in V^{\dagger\dagger}$, we define

$$\phi(l) \coloneqq l(x) \tag{19}$$

This defines a one-to-one correspondence between V and $V^{\dagger\dagger}$.

On the contrary, there is no way to define an isomorphism between V and $V^\dagger$ without further structure on V. In conclusion, this theorem tells us that it is important to be aware of this *duality* between elements $x \in V$ and $l \in V^\dagger$, and thus we should interpret $x \in V$ as a linear function of $V^\dagger$ and $l \in V^\dagger$ as a linear function of V.

Definition 1.17 (Annihilator)

Let W be a subspace of V. Then the set of functions in $V^\dagger$ that vanish on W, that is, satisfy

$$l(w) = 0 \text{ for all } w \in W \tag{20}$$

is called the **annihilator** of W, denoted W^0. If $W = V$, then it is easy to see that W^0 is trivial.

Theorem 1.9 (Annihilator and Dual Quotient Space)

Let V be a vector space and $W \subset V$. Then,

1. $W^0 \simeq (V/W)^\dagger$
2. $\dim W^0 + \dim W = \dim V$
3. $W^{00} = W$.

Proof. Listed.

1. The isomorphism is defined as such. Given $l \in W^0$, we define $L \in (V/W)^\dagger$ as

$$L\{x\} \equiv l(x) \tag{21}$$

Theorem 1.10 (Quadrature Formula)

Let l be an interval on $\mathbb{R}$ containing $t_1, t_2, \ldots, t_n$ n distinct points. Then, given any polynomial p with degree $< n$, there exist n real numbers $c_1, c_2, \ldots, c_n$ such that

$$\int_l p(t)dt = c_1 p(t_1) + c_2 p(t_2) + \ldots + c_n p(t_n) \tag{22}$$

called the *quadrature formula* suffices. That is, the integral of any polynomial over l can be expressed as a linear combination of the polynomials evaluated at n distinct points in l.

Proof. The space of all polynomials with degree $< n$ is an n-dimensional vector space, denote it V. We define the basis of the dual space $V^\dagger$ as

$$\phi_i(p) \equiv p(t_i),\ i = 1, 2, \ldots, n \tag{23}$$

with addition and scalar multiplication defined

$$(\phi + \gamma)(p) \equiv \phi(p) + \gamma(p)$$
$$(c\phi)(p) \equiv c\phi(p)$$

We can see that the ϕ's are indeed linear since, given $p, q \in \mathbb{R}[t]$

$$\phi_i(p+q) = (p+q)(t_i) = p(t_i) + q(t_i) = \phi_i(p) + \phi_i(q)$$
$$\phi_i(cp) = (cp)(t_i) = cp(t_i) = c\phi_i(p)$$

We claim that all the phi_i's are linearly independent. Assume that

$$\sum_{i=1}^n c_i \phi_i(p) = \sum_{i=1}^n c_i p(t_i) = 0 \tag{24}$$

Since the ϕ's must be linearly independent for every polynomial p, set it equal to

$$q_k(t) \equiv \prod_{j \neq k} (t - t_j),\ k = 1, 2, \ldots, n \tag{25}$$

$p = q_k$ must imply that all $\phi_i(p) = 0$ for all $i \neq k$, which implies that $c_k = 0$ in the linear combination. Repeating this for $k = 1, 2, \ldots, n$ results in all $c_i = 0$, implying that the ϕ_i's form a basis of $V^\dagger$. Clearly, the function of definite integration over l is a linear mapping from $V \longrightarrow \mathbb{R}$, meaning that it is in $V^\dagger$. Therefore, it can be expressed as a linear combination of ϕ_i's.

1.4 Exercises

Exercise 1.1 (Lax 1.1)

Show that the zero of vector addition is unique.

Exercise 1.2 (Lax 1.2)

Show that the vector with all components zero serves as the zero element of classical vector addition.

Exercise 1.3 (Lax 1.3)

Show that (i) and (iv) are isomorphic.

Exercise 1.4 (Lax 1.4)

Show that if S has n elements, (i) and (iii) are isomorphic.

Exercise 1.5 (Lax 1.5)

Show that when $K = \mathbb{R}$, (iv) is isomorphic with (iii) when S consists of n distinct points of $\mathbb{R}$.

Exercise 1.6 (Lax 1.6)

Denote by X the linear space of all polynomials $p(t)$ of degree $< n$, and denote by Y the set of polynomials that are zero at $t_1, \ldots, t_j$, $j < n$. (i) Show that Y is a subspace of X. (ii) Determine dim Y. (iii) Determine dim X/Y.

Exercise 1.7 (Lax 1.7)

Prove (i)-(iii) above. Show furthermore that if $x_1 \equiv x_2$, then $kx_1 \equiv kx_2$ for every scalar k.

Exercise 1.8 (Lax 1.9)

Show that the set of all linear combinations of $x_1, \ldots, x_j$ is a subspace of X, and that it is the smallest subspace of X containing $x_1, \ldots, x_j$. This is called the subspace spanned by $x_1, \ldots, x_j$.

Exercise 1.9 (Lax 1.10)

Show that if the vectors $x_1, \ldots, x_j$ are linearly independent, then none of the x_i is the zero vector.

Exercise 1.10 (Lax 1.15)

Show that the above definition of addition and multiplication by scalars is independent of the choice of representatives in the congruence class.

Exercise 1.11 (Lax 1.11)

Prove that if X is finite dimensional and the direct sum of $Y_1, \dots, Y_m$, then

$$\dim X = \sum \dim Y_j.$$

Exercise 1.12 (Lax 1.12)

Show that every finite-dimensional space X over K is isomorphic to K^n, $n = \dim X$. Show that this isomorphism is not unique when n is > 1.

Exercise 1.13 (Lax 1.14)

Show that two congruence classes are either identical or disjoint.

Exercise 1.14 (Lax 1.18)

Show that

$$\dim X_1 \oplus X_2 = \dim X_1 + \dim X_2.$$

Exercise 1.15 (Lax 1.19)

X a linear space, Y a subspace. Show that $Y \oplus X/Y$ is isomorphic to X.

Exercise 1.16 (Lax 1.17)

Prove Corollary 6': A subspace Y of a finite-dimensional linear space X whose dimension is the same as the dimension of X is all of X.

Exercise 1.17 (Lax 2.1)

Given a nonzero vector x_1 in X, show that there is a linear function l such that

Exercise 1.18 (Lax 2.2)

Verify that $Y^\perp$ is a subspace of X'.

Exercise 1.19 (Lax 2.3)

Prove Theorem 6: Denote by Y the smallest subspace containing S:

$$S^\perp = Y^\perp.$$

Exercise 1.20 (Lax 2.4)

In Theorem 6 take the interval I to be $[-1, 1]$, and take n to be 3. Choose the three points to be $t_1 = -a$, $t_2 = 0$, and $t_3 = a$.

(i) Determine the weights m_1, m_2, m_3 so that (9) holds for all polynomials of degree < 3.

(ii) Show that for $a > \sqrt{1/3}$, all three weights are positive.

(iii) Show that for $a = \sqrt{3/5}$, (9) holds for all polynomials of degree < 6.

Exercise 1.21 (Lax 2.5)

In Theorem 6 take the interval I to be $[-1, 1]$, and take n to be 4. Choose the four points to be $-a, -b, b, a$.

(i) Determine the weights m_1, m_2, m_3, and m_4 so that (9) holds for all polynomials of degree < 4.

(ii) For what values of a and b are the weights positive?

Exercise 1.22 (Lax 2.6)

Let $\mathcal{P}_2$ be the linear space of all polynomials

$$p(x) = a_0 + a_1 x + a_2 x^2$$

with real coefficients and degree ≤ 2. Let ξ_1, ξ_2, ξ_3 be three distinct real numbers, and then define

$$\ell_j = p(\xi_j) \quad \text{for} \quad j = 1, 2, 3.$$

(a) Show that ℓ_1, ℓ_2, ℓ_3 are linearly independent linear functions on $\mathcal{P}_2$.

(b) Show that ℓ_1, ℓ_2, ℓ_3 is a basis for the dual space $\mathcal{P}_2'$.

(c) (1) Suppose $\{e_1, \ldots, e_n\}$ is a basis for the vector space V. Show there exist linear functions $\{\ell_1, \ldots \ell_n\}$ in the dual space V' defined by

$$\ell_i(e_j) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

Show that $\{\ell_1, \ldots, \ell_n\}$ is a basis of V', called the *dual basis*.

(2) Find the polynomials $p_1(x), p_2(x), p_3(x)$ in $\mathcal{P}_2$ for which ℓ_1, ℓ_2, ℓ_3 is the dual basis in $\mathcal{P}_2'$.

Exercise 1.23 (Lax 2.7)

Let W be the subspace of $\mathbb{R}^4$ spanned by $(1, 0, -1, 2)$ and $(2, 3, 1, 1)$. Which linear functions $\ell(x) = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4$ are in the annihilator of W?

Exercise 1.24 (Math 403 Spring 2020, Week 1.2)

If $e_1, e_2, \ldots, e_n$ is a basis for V and we define linear maps e_i' to K by

$$e_i'(e_j) = \delta_{ij},$$

then prove $e_1', e_2', \ldots, e_n'$ form a basis for V'.

Exercise 1.25 (Math 403 Spring 2020, Week 4.2)

Find the Jordan Normal Form, J, of

$$A = \begin{pmatrix} 4 & -2 & -6 \\ -1 & 2 & 2 \\ 1 & -1 & -1 \end{pmatrix}$$

and a matrix S such that $J = S^{-1}AS$.

Exercise 1.26 (Math 403 Spring 2020, Week 4.3)

Let $d^n(\alpha)$ denote the dimension of the nullspace of $(\alpha I - A)^n$ for a matrix A and a scalar α. In the following examples we list all the possible nonzero values of $d^n(\alpha)$. In each case find the Jordan Normal Form of A or prove A cannot exist.

1. $d^1(1) = 2$, and $d^n(1) = 3$ if $n \geq 2$.
2. $d^n(2) = 1$ if $n \geq 1$; $d^1(1) = 2$, $d^n(1) = 4$ if $n \geq 2$.
3. $d^1(1) = 1$, $d^n(1) = 3$ if $n \geq 2$.

Exercise 1.27 (Math 403 Spring 2020, Week 7.1)

Let

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Compute $\Delta(\sigma_x)\Delta(\sigma_y)$ if the state vector is $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Is this consistent with the Heisenberg Uncertainty Principle?

Exercise 1.28 (Math 403 Spring 2020, Week 8.1)

Compute the singular value decompositions of the following matrices

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

Exercise 1.29 (Math 403 Spring 2020, Week 8.2)

Let M be any matrix over $\mathbb{R}$ and let $M^{(k)}$ be the rank k approximation of M we did in class. That is $M^{(k)}$ is obtained from an SVD of M where we set all but the k largest singular values to 0.

1. Show that $B = M^{(k)}$ is not the *unique* matrix that minimizes the norm

$$||M - B||_2$$

   if $\sigma_k = \sigma_{k+1}$ where $\sigma_1 \geq \sigma_2 \geq \cdots$ are the singular values of M.
2. Discuss if $B = M^{(k)}$ is the unique matrix that minimizes this norm if $\sigma_k > \sigma_{k+1}$.

Exercise 1.30 (Math 403 Spring 2020, Week 9.1)

Let

$$A = \begin{pmatrix} 0 & 0 & 0 & 1/3 \\ 1/2 & 0 & 1 & 1/3 \\ 0 & 1 & 0 & 1/3 \\ 1/2 & 0 & 0 & 0 \end{pmatrix}.$$

Compute the asymptotic behaviour of $A^N v$ where v has coordinates $(1/4, 1/4, 1/4, 1/4)$ and N is large.

Exercise 1.31 (Math 403 Spring 2020, Week 10.1)

Prove $v \otimes 0 = 0$ in $V \otimes V$ for any vector $v \in V$.

Exercise 1.32 (Math 403 Spring 2020, Week 10.2)

Prove there is a natural isomorphism between the vector spaces $\mathrm{Hom}(V, V)$ and $V \otimes V^*$.

Exercise 1.33 (Math 403 Spring 2020, Week 10.3)

If $\{e1, e2, \cdots\}$ is a basis for V, which vector in $V \otimes V^*$ is identified with the identity in $\mathrm{Hom}(V, V)$ under this isomorphism.

Exercise 1.34 (Math 403 Spring 2020, Week 10.4)

Given a map $f : Y \otimes X \to Z$ what is the naturally associated map $Y \to X^* \otimes Z$? Give your answer in terms of components assuming bases are given for all the spaces involved.

Exercise 1.35 (Math 403 Spring 2020, Week 10.5)

Let $V = \mathbb{R}^3$ with a basis $\{e_1, e_2, e_3\}$ and with the standard inner product. Consider the vector $v \in V \otimes V$ which if we write in terms of components

$$v = \sum_{ij} m_{ij} e_i \otimes e_j,$$

then the matrix M with entries m_{ij} is given by

$$M = \begin{pmatrix} 1 & 2 & -1 \\ 1 & -2 & -1 \\ -1 & 0 & 1 \end{pmatrix}.$$

Find vectors $x_1, x_2, y_1, y_2 \in V$ such that

$$v = x_1 \otimes y_1 + x_2 \otimes y_2,$$

with x_1 perpendicular to x_2, and y_1 perpendicular to y_2.

Exercise 1.36 (Math 403 Spring 2020, Week 10.6)

What does the exterior algebra of $\mathbb{R}^3$ look like? What is the wedge product of two vectors?

Exercise 1.37 (Math 403 Spring 2020, Week 11.1)

The Baker-Campbell-Hausdorff formula says $e^A e^B = e^{A*B}$, where

$$A * B = \sum_{k=1}^{\infty} F_k$$

and F_k is of total order k in A and B. We have

$$F_1 = A + B$$

$$F_2 = \frac{1}{2}[A, B].$$

Compute F_3 expressing it in terms of commutators.

Exercise 1.38 (Math 403 Spring 2020, Week 11.2)

Let $H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $S = \exp(i\theta H) \in \mathrm{SU}(2)$. Let M be a traceless self-adjoint 2×2 matrix. Then let M correspond to a three dimensional vector with components (x, y, z) by decomposing using the Pauli matrices as

$$M = x\sigma_x + y\sigma_y + z\sigma_z.$$

Suppose we define

$$M' = S^{-1} M S.$$

Show M' is also traceless and self-adjoint. If M' is similarly associated to a three dimensional vector, show S thereby induces a linear transformation on $\mathbb{R}^3$. What exactly is this transformation?

Exercise 1.39 (Math 403 Spring 2020, Week 12.1)

Recall that the Baker-Campbell-Hausdorff formula says the group law is given by $e^A e^B = e^{A*B}$, where

$$A * B = \sum_{k=1}^{\infty} F_k$$

and F_k is of total order k in A and B. We have

$$F_1 = A + B$$

$$F_2 = \frac{1}{2}[A, B].$$

You computed F_3 last week and expressed it purely in terms of the bracket. Let A and B be elements of the vector space V and let the bracket be a bilinear form $V \otimes V \to V$. Now think of this as an arbitrary map not necessarily given by commutators.

1. Prove $-(-B) * (-A) = A * B$ and thus the bracket obeys $[A, B] = -[B, A]$.
2. Using your knowledge of F_3, prove that associativity of the group law implies the Jacobi identity

$$[A, [B, C]] + [B, [C, A]] + [C, [A, B]] = 0.$$

Exercise 1.40 (Math 403 Spring 2020, Week 12.2)

Let n denote the irreducible representation of $\mathfrak{sl}_2\mathbb{C}$ of dimension n. In class we showed $2 \otimes 2 = 3 \oplus 1$. Perform similar decompositions for:

1. $3 \otimes 2$
2. $\mathrm{Sym}^2 3$
3. $\Lambda^3 4$.

2 Linear Maps

Let's first recall some of the basic properties an notation for functions.

Definition 2.1 (Image)

The **image** of linear map $T : U \to V$, denoted $\text{Im}(T)$, is the range of T.

$$\text{Im}(T) \coloneqq \{T(u) \in V \mid u \in U\} \tag{26}$$

We claim $\text{Im}(T)$ is a subspace of V.

Proof. It suffices to show that the image is a subspace.

Definition 2.2 (Rank)

The **rank** of a linear map A is the dimension of its image.

Definition 2.3 (Nullspace)

The **nullspace** of $T : U \to V$, denoted $\ker(T)$, is the is the kernel of T.

$$\ker(T) \coloneqq \{u \in U \mid T(u) = 0\} \tag{27}$$

We claim $\ker(T)$ is a subspace of U.

Proof. It suffices to show that the image is a subspace.

For simplicity, we usually just denote composition simply by concatenation: $S \circ T = ST$.

Definition 2.4 (Restriction)

Let $\varphi : U \longrightarrow V$ be a linear mapping and let $U_1 \subset U, V_1 \subset V$ be subspaces. Such that

$$\varphi u \in V \text{ for all } u \in U \tag{28}$$

Then, the linear mapping

$$\varphi_1 : U_1 \longrightarrow V_1,\ \varphi_1 u = \varphi_u, u \in U_1 \tag{29}$$

is called the **restriction** of φ to U_1, V_1. It suffices the identity

$$\varphi \circ i_U = i_V \circ \varphi_1 \tag{30}$$

where $i_U : U_1 \longrightarrow U, i_V : V_1 \longrightarrow V$ are canonical injections. Equivalently, we say that the diagram below is **commutative**.

$$\begin{array}{ccc} U & \xrightarrow{\varphi} & V \\ {\scriptstyle i_U}\uparrow & & {\scriptstyle i_V}\uparrow \\ U_1 & \xrightarrow{\varphi_1} & V_1 \end{array}$$

We can also define

$$\varphi_1 \equiv i_v^{-1} \varphi i_U \tag{31}$$

Example 2.1

Let U_1 be a subspace of U and given the quotient map

$$\pi : U \longrightarrow U/U_1 \tag{32}$$

Then,

$$\ker \pi = U_1, \ \operatorname{Im} \pi = U/U_1 \tag{33}$$

Note that a quotient map is always surjective.

Theorem 2.1 (Rank Nullity Theorem)

Let $T : U \longrightarrow V$ be linear. Then,

$$\dim \ker T + \dim \operatorname{Im} T = \dim U \tag{34}$$

This theorem is quite intuitive, if we visualize the map.

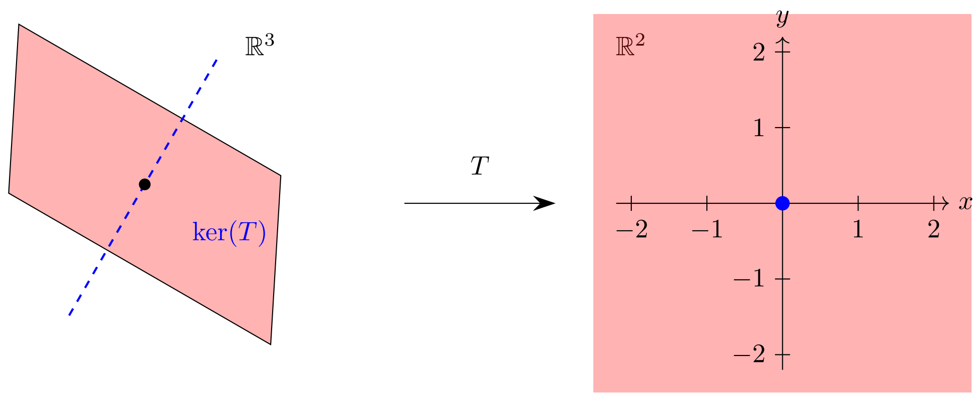


Figure 4: We just have to realize that given a linear transformation mapping from a n-dimensional V to a m-dimensional U, every vector in V will either get mapped to $0 \in U$ or will get mapped to a nonzero vector in U. In this case, the kernel will get mapped to 0 and everything else is (not always, but in this case) $\mathbb{R}^2$.

Proof.

Theorem 2.2

Given $A : V \longrightarrow U$ a linear mapping between vector spaces and $b \in U$, all solutions to the equation $Ax = b$ is in $a + \ker A$, that is, of the form

$$x = a + y, \ y \in \ker A \tag{35}$$

where $x = a$ is one solution.

Corollary 2.3

A linear map A is injective if and only if $\ker A = 0$.

2.1 Invertibility

We now introduce the concepts of left and right invertibility of linear mappings.

Theorem 2.4 (Left/Right-Invertibility)

A linear mapping $T : U \longrightarrow V$, with $\dim U = n, \dim V = m$, is **left-invertible**. That is, there exists linear S such that

$$ST = I \tag{36}$$

if and only if T is injective $\iff$ $\text{rank}(T) = n$. Linear T is **right-invertible**, that is, there exists linear S such that

$$TS = I \tag{37}$$

if and only if T is surjective $\iff$ $\text{rank}(T) = m$.

Proof. We will only prove the case for left-invertibility. Right invertibility follows analogously.

1. $(\leftarrow)$ T is injective $\implies$ $\text{rank}(T) = \dim U = \dim \text{Im}\, T$. Let $(\text{Im}\, T)'$ exist such that

$$\text{Im}\, T \oplus (\text{Im}\, T)' = V \tag{38}$$

   We define the isomorphism

$$\tilde{T} : V \longrightarrow \text{Im}\, T \tag{39}$$

   and then define S. Given that $v = w + w' \in V$, with $w \in \text{Im}\, T, w' \in (\text{Im}\, T)'$,

$$S : V \longrightarrow U,\ S(v) \equiv \tilde{T}^{-1}(v) \tag{40}$$

   $\implies ST(u) = \tilde{T}^{-1}T(u) = u \iff ST = I$.

2. $(\rightarrow)$ We prove the contrapositive. T is not injective $\implies$ $\dim \ker T > 0$ $\implies$ there exists 2 linearly independent vectors $x, y \in U$ such that

$$Tx = Ty \tag{41}$$

   Assume that a left inverse S exists. Then

$$x = STx = STy = y \implies x = y \tag{42}$$

   leading to a contradiction $\implies$ the left-inverse does not exist.

Definition 2.5 (Inverse)

The inverse of a linear map A, denoted A^{-1} is a unique linear map satisfying

$$AA^{-1} = A^{-1}A = I \tag{43}$$

where I is the identity map.

Corollary 2.5

A linear map is invertible if and only if it is an isomorphism.

Definition 2.6 (General Linear Group)

$\text{Aut}(V)$ of vector space V also forms a group under composition. We denote it $\text{GL}(V)$. The group of automorphisms of $\mathbb{R}^n$ and $\mathbb{C}^n$ is denoted $\text{GL}(\mathbb{R}^n)$ and $\text{GL}(\mathbb{C}^n)$, respectively. The group of all invertible $n \times n$ matrices over $\mathbb{R}$ and $\mathbb{C}$ is denoted $\text{GL}_n(\mathbb{R})$ and $\text{GL}_n(\mathbb{C})$. $\text{GL}_n(\mathbb{R})$ is also denoted $\text{GL}(n, \mathbb{R})$, and similarly for $\text{GL}(n, \mathbb{C})$.

Theorem 2.6

Given that V is a real vector space,

$$GL(V) \simeq GL(\mathbb{R}^n) \simeq GL_n(\mathbb{R}) \tag{44}$$

since $\text{GL}_n(\mathbb{R})$ are representations of linear operators. Similarly, if V is a complex vector space,

$$GL(V) \simeq GL(\mathbb{C}^n) \simeq GL_n(\mathbb{C}) \tag{45}$$

Definition 2.7 (Special Linear Group)

The group of all real $n \times n$ matrices that have determinant 1 is called the **special linear group**, denoted $\text{SL}_n(\mathbb{R})$. It is a subgroup of $\text{GL}_n(\mathbb{R})$. The group of all complex $n \times n$ matrices with determinant 1 is denoted $\text{SL}_n(\mathbb{C})$. It is a subgroup of $\text{GL}_n(\mathbb{C})$.

2.2 Transpose

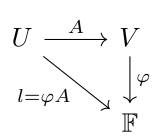
We finally end this section by defining the transpose of a linear mapping.

Definition 2.8 (Transpose)

Given a linear mapping $A : U \longrightarrow V$, let there exist a certain $\varphi \in V^*$. Then, there exists a corresponding $l \in U^*$ such that

$$l \equiv \varphi A \tag{46}$$

This mapping $A^T : V^* \longrightarrow U^*$ that assigns every φ to a corresponding l is called the **transpose** of A. Note that the transpose is canonically formed when defining any linear map. We do not need any additional structure on U or V to define A^T.



It is worth mentioning that A^T maps every element in the annihilator V^0 to an element in U^0, but not necessarily the other way around.

Theorem 2.7

For a linear map A, it is true that

$$(\text{Im}\, A)^0 = \ker A^T, \quad \text{Im}\, A = (\ker A^T)^0 \tag{47}$$

2.3 Algebra of Linear Maps

Now, we are going to be a bit meta and talk about the set of linear maps. This itself turns out to be a vector space!

Definition 2.9 (Vector Space of Linear Maps)

$\text{Hom}(U, V)$ is the vector space of linear mappings, with addition and scalar multiplication defined

$$(S + T)(x) \equiv S(x) + T(x)$$
$$(cT)(x) \equiv cT(x)$$

Proof.

Theorem 2.8 (Composition is Distributive)

Composition is (right and left) distributive with respect to the addition of linear maps. That is,

$$(R + S) \circ T = R \circ T + S \circ T \tag{48}$$
$$T \circ (R + S) = T \circ R + T \circ S \tag{49}$$

In the theorem above, it seems like composition acts sort of like multiplication. Recall that such an algebraic structure that has both the properties of a vector space and multiplication is called an algebra.

Definition 2.10 (Algebra of Endomorphisms)

$\text{End}(V)$ is an associative non-commutative algebra.

Proof.

Example 2.2

A rotation around any axis or a flip across any hyperplane is an element of $\text{End}(\mathbb{R}^n)$.

Definition 2.11 (Projection)

A **projection mapping** is a linear mapping P where

$$P = P^2 \tag{50}$$

Example 2.3

Let P be an orthogonal projection mapping onto a subspace Y of X. $\text{Im}\, P = Y$, and $\ker P = Y^\perp$ or the span of vectors in X that are "orthogonal" to Y. Note that we haven't actually endowed a structure onto X to even define orthogonality yet, so this definition is purely visual and not mathematically rigorous.

Example 2.4

Reflections, projections, shears, and rotations are all linear maps. Differentiation and integration are also examples of linear mappings.

Linear maps over vector spaces over different fields are generally not well defined since the definition of homomorphisms do not cover the fields in which vector spaces are associated with.

Definition 2.12 (Operator Norm)

Definition 2.13 (Frobenius Norm)

2.4 Factorization of Linear Maps

The construction of the restriction is an enormously helpful tool for many proofs and very useful for factoring linear mappings.

Theorem 2.9 (Universal Property of Factorization)

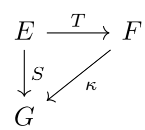
Given linear mappings $T : E \longrightarrow F$, $S : E \longrightarrow G$ such that

$$\ker T \subseteq \ker S \tag{51}$$

Then there exists a map κ such that

$$S = \kappa T \tag{52}$$

or equivalently, such that the diagram below commutes.



Definition 2.14 (Induced Mapping of Quotients)

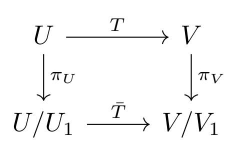
Given $T : U \longrightarrow V$ with quotient maps

$$\pi_U : U \longrightarrow U/U_1,\ \pi_V : V \longrightarrow V/V_1 \tag{53}$$

the **induced mapping of the quotient spaces** is the unique mapping $\bar{T} : U/U_1 \longrightarrow V/V_1$ such that

$$\bar{T} \circ \pi_U = \pi_V \circ T \tag{54}$$

or equivalently, the following diagram commutes.



Theorem 2.10 (First Isomorphism Theorem)

Every linear mapping can be written as the composition of a surjective mapping followed by an injective mapping. That is, every T can be factored into

$$T = T_{inj} \circ T_{surj} \tag{55}$$

Proof. We can induce a quotient mapping to construct a factoring of a linear mapping. We can define the unique mapping

$$\bar{T} : U/\ker T \longrightarrow V \tag{56}$$

such that, $T = \bar{T} \circ \pi_U$, or that

$$\begin{array}{ccc} U & \xrightarrow{T} & V \\ \downarrow{\scriptstyle \pi_U} & \nearrow{\scriptstyle \bar{T}} & \\ U/\ker T & & \end{array}$$

commutes. Clearly, $\bar{T}$ is injective since if it were not, $\bar{T}\pi_U x = 0 \implies Tx = 0 \implies x \in \ker T \implies \pi_U x = 0$. This also means that the restriction of $\bar{T}$ to $U/\ker T$

$$\bar{\bar{T}} : U/\ker T \longrightarrow \operatorname{Im} T \tag{57}$$

is a linear isomorphism. Thus, for any T, it can be written as $\bar{T} \circ \pi_U$, with $\bar{T}$ injective and π_U surjective.

Theorem 2.11 (Second Isomorphism Theorem)

Given E_1, E_2 subspaces of E. Then,

$$\frac{E_1}{E_1 \cap E_2} \simeq \frac{E_1 + E_2}{E_2} \tag{58}$$

In fact, they are naturally isomorphic.

Proof. $E_1 + E_2$ can be decomposed to $E_1' \oplus (E_1 \cap E_2) \oplus E_2'$, where E_1' consists of the subspace of vectors x that can only be expressed as $x = x_1, x_1 \in E_1$ and E_2' are vectors that can only be expressed as $x = x_2, x_2 \in E_2$. Define the projection mapping

$$\text{proj} : E_1 + E_2 \longrightarrow E_1' \tag{59}$$

Since $E_1 = E_1' \oplus (E_1 \cap E_2)$, we can define the natural isomorphism

$$\kappa : E_1' \longrightarrow \frac{E_1}{E_1 \cap E_2}, \; \kappa x = \{x\} \tag{60}$$

We now define the mapping $\varphi : E_1' \longrightarrow (E_1 + E_2)/E_2$ such that

$$\pi = \varphi \text{proj} \tag{61}$$

given by the diagram

$$\begin{array}{ccc} E_1 + E_2 & \xrightarrow{\pi} & \frac{E_1+E_2}{E_2} \\ \downarrow{\scriptstyle proj} & \nearrow{\scriptstyle \varphi} & \\ E_1' & \xrightarrow{\kappa} & \frac{E_1}{E_1 \cap E_2} \end{array}$$

Such a φ exists because proj is surjective and can thus be inverted. We now claim that φ is an isomorphism. $\ker \text{proj} = \ker \pi = E_2 \implies \kappa$ is injective. Given $x = x_1 + y + x_2 \in E_1 + E_2$ such that

$x_1 \in E_1', y \in E_1 \cap E_2, x_2 \in E_2'$,

$$\pi(x) = \pi(x_1 + y + x_2) = \pi(x_1) = \varphi \text{proj}(x_1) = \varphi(x_1) \tag{62}$$

meaning that for every vector $v \in (E_1 + E_2)/E_2$, it can be expressed as $v = \pi(x) = \varphi(x_1)$, meaning that there exists a $x_1 \in E_1'$ mapping to v under $\varphi \iff \varphi$ is surjective. So, φ is an isomorphism $\implies \varphi\kappa^{-1}$ is an isomorphism.

Corollary 2.12

In the special case when $E_1 \oplus E_2 = E$, then the theorem states that

$$E_1 \simeq \frac{E}{E_2} \tag{63}$$

Theorem 2.13 (Third Isomorphism Theorem)

Let V be a vector space, and let W and U be subspaces of V such that $U \subset W$. Then W/U is a subspace of V/U, and we have the natural isomorphism

$$\frac{V/U}{W/U} \simeq \frac{V}{W} \tag{64}$$

Proof. Define a linear map $T : V/U \longrightarrow V/W$ by $T([v]_U) = [v]_W$. This map is well-defined and surjective. The kernel of T is the set of all $[v]_U \in V/U$ such that $[v]_W = [0]_W$, which implies $v \in W$. Thus $\ker(T) = W/U$. By the First Isomorphism Theorem, $(V/U)/\ker(T) \simeq \text{Im}(T)$, which yields $(V/U)/(W/U) \simeq V/W$.

Let $f_1, f_2, \cdots, f_n$ be any n linear functionals of U. Define the subspace $F \subset E$ as

$$F \equiv \bigcap_{i=1}^{n} \ker f_i \tag{65}$$

and define linear map

$$T : U \longrightarrow \mathbb{F}^n,\ T(x) \equiv (f_1(x), f_2(x), \cdots, f_n(x)) \tag{66}$$

$\implies \ker T = F$. So, $T : U \longrightarrow \mathbb{F}^n$ defines the isomorphism

$$\bar{T} : U/F \longrightarrow \text{Im}\, T \tag{67}$$

2.5 Exact Sequences

Now, we introduce the concept of exact sequences which is useful in the factoring of linear maps. Note that exact sequences are used in group theory to factor transformation groups.

Definition 2.15 (Exact Sequences)

A sequence of linear mappings

$$F \xrightarrow{T} E \xrightarrow{S} G \tag{68}$$

is **exact at E** if $\text{Im}\, T = \ker S$.

Notice that if we have an exact sequence $0 \xrightarrow{T} E \xrightarrow{S} G$, then $0 = \text{Im}\, T = \ker S \implies S$ is injective. If we have exact sequence $F \xrightarrow{T} E \xrightarrow{S} 0$, then $\text{Im}\, T = \ker S = E \implies T$ is surjective.

Definition 2.16 (Short Exact Sequences)

A **short exact sequence** is a sequence of the form

$$0 \to F \xrightarrow{T} E \xrightarrow{S} G \to 0 \tag{69}$$

such that it is exact at F, E, and G. It is clear that the first and last maps are the zero maps. With this definition, we can easily prove that

1. T is injective
2. S is surjective
3. $E/\operatorname{Im} T \simeq G$

Example 2.5

The sequence

$$0 \to E_1 \xrightarrow{i} E \xrightarrow{\pi} E/E_1 \to 0 \tag{70}$$

is exact, where i denotes the canonical injection and π the canonical projection. This example is the only example of an exact sequence between vector spaces up to isomorphism.

Definition 2.17

A commutative diagram of the form

$$\begin{array}{ccccccccc} 0 & \longrightarrow & F_1 & \xrightarrow{T_1} & E_1 & \xrightarrow{S_1} & G_1 & \longrightarrow & 0 \\ & & \downarrow{\alpha} & & \downarrow{\beta} & & \downarrow{\gamma} & & \\ 0 & \longrightarrow & F_2 & \xrightarrow{T_2} & E_2 & \xrightarrow{S_2} & G_2 & \longrightarrow & 0 \end{array}$$

where both horizontal sequences are short exact sequences and α, β, γ are homomorphisms between linear spaces is a **homomorphism of exact sequences**. If α, β, γ are linear isomorphisms, then this is an **isomorphism of exact sequences**.

Theorem 2.14

A short exact sequence of vector spaces

$$0 \to F \xrightarrow{T} E \xrightarrow{S} G \to 0 \tag{71}$$

is split if it essentially presents E as the direct sum of groups F and G. That is, there exists an isomorphism of exact sequences.

$$\begin{array}{ccccccccc} 0 & \longrightarrow & F & \xrightarrow{T_1} & E & \xrightarrow{S_1} & G & \longrightarrow & 0 \\ & & \downarrow{\alpha} & & \downarrow{\beta} & & \downarrow{\gamma} & & \\ 0 & \longrightarrow & F & \xrightarrow{T_2} & F \oplus G & \xrightarrow{S_2} & G & \longrightarrow & 0 \end{array}$$

or equivalently, there exists an isomorphism between E and $F \oplus G$.

Definition 2.18

Given a short exact sequence

$$0 \to F \xrightarrow{T} E \xrightarrow{S} G \to 0 \tag{72}$$

if there exists a map $\kappa : G \longrightarrow E$, such that $S \circ \kappa = I$, then the sequence is said to be a **split short exact sequence**, written

$$0 \to F \xrightarrow{T} E \xleftrightarrow{S,\kappa} G \to 0 \tag{73}$$

Theorem 2.15

Every short exact sequence can be split.

Proof. It will be proved later that S is surjective $\implies$ S is left invertible.

Definition 2.19 (Stable Subspaces)

Given $T : E \longrightarrow E$, a subspace $E_1 \subset E$ is called **stable**

$$x \in E_1 \implies Tx \in E_1 \tag{74}$$

That is, the restriction of T to E_1, denoted

$$T_1 : E_1 \longrightarrow E_1 \tag{75}$$

is well-defined. Clearly, $\operatorname{Im} T$ and $\ker T$ is stable, and the induced map

$$\bar{T} : E/E_1 \longrightarrow E/E_1 \tag{76}$$

is a linear endomorphism of E/E_1.

We end this subsection by defining the induced linear map from the direct sum of spaces.

Definition 2.20

Given linear maps $T_i \in \operatorname{End}(V_i)$ for $i = 1, 2, \cdots, n$, the induced linear map

$$\bigoplus_{i=1}^n T_i : \bigoplus_{i=1}^n V_i \longrightarrow \bigoplus_{i=1}^n V_i \tag{77}$$

is defined

$$(\bigoplus_{i=1}^n T_i)(\bigoplus_{i=1}^n x_i) \equiv \bigoplus_{i=1}^n T_i x_i \tag{78}$$

2.6 Exercise

3 Inner Products

Definition 3.1 (Inner Product)

An **inner product** on a vector space V over field $\mathbb{C}$ is a mapping

$$(\cdot, \cdot) : V \times V \longrightarrow \mathbb{R} \tag{79}$$

satisfying three properties

1. First Argument Linearity: $(\lambda x + \mu y, z) = \lambda(x, z) + \mu(y, z)$
2. Conjugate symmetry: $(x, y) = (y, x)$
3. $(x, y) \geq 0$, with $(x, y) = 0 \iff x = y$

An inner product allows us to define some notion of an angle between two vectors in V. A vector space V with an inner product is called an **inner product space**. Note that when the field is $\mathbb{C}$, the inner product is **sesqui-linear**, that is, linear with respect to the first argument and **skew linear** with respect to the second. When $\mathbb{R}$, it is bilinear.

The inner product of a vector space V over $\mathbb{R}$ is an element of $V^* \otimes V^*$. This concept of the metric tensor occurs when studying Riemannian manifolds in general relativity.

Definition 3.2 (Inner Product Induces Norm)

An inner product induces a norm in the following way

$$||x|| \equiv \sqrt{(x, x)} \tag{80}$$

Theorem 3.1 (Schwarz Inequality)

For all $x, y \in V$,

$$|(x, y)| \leq ||x||||y|| \tag{81}$$

Example 3.1 (Dot Product)

Given vectors $x, y \in \mathbb{R}^n$,

$$x \cdot y \equiv \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \equiv \sum_{i=1}^{n} x_i y_i \tag{82}$$

Example 3.2 (Integral Product)

Let $C^0[a, b]$ be the space of all continuous real-valued functions defined over the interval $[a, b] \subset \mathbb{R}$. Given $f, g \in C^0[a, b]$,

$$(f, g) \equiv \int_a^b f(x) g(x) dx \tag{83}$$

is an inner product on $C^0[a, b]$.

Theorem 3.2 (Pythagorean Theorem)

$$||x||^2 + ||y||^2 = ||x+y||^2 \tag{84}$$

Theorem 3.3

$$||x|| = \max_{||y||=1} (x, y) \tag{85}$$

3.1 Orthogonal Vectors

Definition 3.3 (Orthogonal Vectors)

Two vectors x, y of an inner product space are said to be **orthogonal** if

$$(x, y) = 0 \tag{86}$$

Note that the definition of orthogonality is dependent on the definition of the inner product. If the inner product is defined differently, then orthogonality will be defined differently. In the case when the inner product is defined to be the dot product, orthogonality is defined to be the "normal" perpendicularity between vectors. We can further define subspaces to be orthogonal.

Definition 3.4 (Orthogonal Subspaces)

Two subspaces Y, Z of inner product space Z are said to be orthogonal to each other if

$$(y, z) = 0 \text{ for every } y \in Y, z \in Z \tag{87}$$

Definition 3.5 (Orthogonal Complement)

Given a subspace Y of inner product space X, the **orthogonal complement** of Y, denoted $Y^\perp$, is defined

$$\{x \in X \mid (x, y) = 0 \quad \forall y \in Y\} \tag{88}$$

which is the set of all vectors in X orthogonal to every vector in Y. Clearly, $Y \oplus Y^\perp = X$.

Algorithm 3.1 (Grahm-Schmidt)

Theorem 3.4 (Orthonormal Basis in Hilbert Space)

Every inner product space has a basis consisting of vectors that are pairwise orthogonal, called an **orthogonal basis**. Furthermore, each vector in the orthogonal basis can be scaled to have magnitude 1, forming an **orthonormal basis**.

Proof. The algorithm used to construct an orthonormal basis is called **Grahm-Schmidt**. We start off with any basis, not necessarily orthonormal, of X, denoted $\{x_1, x_2, \cdots, x_n\}$. We first assign

$$x_1 = l_1 \tag{89}$$

Then we take x_2 and find the orthogonal component (with respect to l_1) with the equation

$$l_2 = x_2 - \text{proj}_{l_1}(x_2) \tag{90}$$

This creates an orthogonal basis for span$\{x_1, x_2\}$. Then we take x_3 and find the orthogonal component (with respect to span$\{l_1, l_2\}$.

$$l_3 = x_3 - \text{proj}_{l_1}(x_3) - \text{proj}_{l_2}(x_3) \tag{91}$$

This creates an orthogonal basis for span$\{x_1, x_2, x_3\}$. We repeat this process until we complete the basis of X, using the general equation

$$l_k = x_k - \sum_{i=1}^{k-1} \text{proj}_{l_k}(x_k) = x_k - \sum_{i=1}^{k-1} \frac{x_k \cdot l_k}{||l_k||^2} l_k,\ k = 1, 2, \cdots, n \tag{92}$$

Finally, we take these orthogonal vectors and normalize them to magnitude 1. Note that this algorithm does not produce a unique orthonormal basis. Rather, it is highly un-unique.

3.2 Adjoint Operators

Definition 3.6 (Adjoint Operator)

Let $A : U \longrightarrow V$ be a linear mapping between inner product spaces, with the inner product in U and V denoted $(\cdot, \cdot)_U$ and $(\cdot, \cdot)_V$, respectively. We can fix any $v \in V$ and define the linear function $l \in U^*$

$$l(\cdot) = \big(A(\cdot), v\big)_V \tag{93}$$

Since U is naturally isomorphic to U^*, we can define

$$l(\cdot) \equiv (\cdot, u') \tag{94}$$

to get

$$\big(\cdot, u'\big)_U \equiv \big(A(\cdot), v\big)_V \tag{95}$$

By combining (8), which defines an isomorphism between U^* and V, and (9), the natural isomorphism between U and U^*, equation (10) takes the composition of these to define an isomorphism from V to U. This isomorphism is called the **adjoint** of A.

$$A^\dagger : V \longrightarrow U,\ \big(\ \cdot\ , A^\dagger v\big)_U = \big(A(\cdot), v\ \big)_V \tag{96}$$

By definition, given any $v \in V$, $A^\dagger v$ is defined so that the equality

$$(u, A^\dagger v) = (Au, v) \tag{97}$$

holds for all values of $u \in U$.

It is important to note that the adjoint is not the same as the transpose since the transpose is a mapping between the dual spaces. Furthermore, the transpose is canonically defined upon defining the linear transformation $A : U \longrightarrow V$, while defining the adjoint requires the additional structure of an isomorphism from U to U^* and from V to V^*. There are two ways to define these isomorphisms.

First, we can define dot products on both U and V and define the natural isomorphism

$$i : U \longrightarrow U^*,\ i(u) \equiv (u, \cdot) \in U^*$$
$$j : V \longrightarrow V^*,\ j(v) \equiv (v, \cdot) \in V^*$$

This canonically creates the mapping

$$i^{-1} A^T j : V \longrightarrow U \tag{98}$$

which we define as the adjoint $A^\dagger$. This method using natural isomorphisms is precisely how we have defined the adjoint above. There is a second way, however. We can fix **orthonomal** bases on U and V and then assign them their respective dual spaces (satisfying the Kronecker delta function). Let the basis of U be $\{u_1, \cdots, u_n\}$, U^* be $\{u_1', \cdots, u_n'\}$, V be $\{v_1, \cdots, v_m\}$, and V^* be $\{v_1', \cdots, v_m'\}$. Now we can define the isomorphisms

$$\begin{aligned} i' &: U \longrightarrow U^*,\ i'(u) \equiv c_1 u_1' + \cdots + c_n u_n' \\ j' &: V \longrightarrow V^*,\ j'(v) \equiv k_1 v_1' + \cdots + k_m v_m' \end{aligned}$$

and then define the adjoint as

$$A^\dagger \equiv i'^{-1} A^T j' \tag{99}$$

Let us compare these two definitions. Given a vector $u = a_1 u_1 + \cdots + a_n u_n, \tilde{u} = b_1 u_1 + \cdots b_n u_n \in U$,

$$\begin{aligned} i(u)(\tilde{u}) \equiv (u, \tilde{u}) &= \Big(\sum_{\alpha=1}^n a_\alpha u_\alpha, \sum_{\beta=1}^n b_\beta u_\beta \Big) = \sum_{\alpha,\beta} a_\alpha b_\beta \delta_\beta^\alpha = \sum_{\gamma=1}^n a_\gamma b_\gamma \\ i'(u)(\tilde{u}) &\equiv \Big(\sum_{i=1}^n a_i u_i' \Big) \Big(\sum_{j=1}^n b_j u_j \Big) = \sum_{i,j} a_i b_j u_i'(u_j) = \sum_{i,j} a_i b_j \delta_j^i = \sum_{k=1}^n a_k b_k \end{aligned}$$

Similarly for vector $v = g_1 v_1 + \cdots g_n v_n, \tilde{v} = h_1 v_1 + \cdots + h_n v_n \in V$,

$$\begin{aligned} i(v)(\tilde{v}) \equiv (v, \tilde{v}) &= \Big(\sum_{\alpha=1}^n g_\alpha v_\alpha, \sum_{\beta=1}^n h_\beta v_\beta \Big) = \sum_{\alpha,\beta} g_\alpha h_\beta \delta_\beta^\alpha = \sum_{\gamma=1}^n g_\gamma h_\gamma \\ i'(v)(\tilde{v}) &\equiv \Big(\sum_{i=1}^n g_i u_i' \Big) \Big(\sum_{j=1}^n h_j v_j \Big) = \sum_{i,j} g_i h_j v_i'(v_j) = \sum_{i,j} g_i h_j \delta_j^i = \sum_{k=1}^n g_k h_k \end{aligned}$$

Therefore, $i = i'$ and $j = j'$, meaning that the two derivations of the adjoint $A = i^{-1} A^T j = i^{-1\prime} A^T j'$ are exactly the same! We must note that the basis endowed on both U and V must be orthonormal for it to "mimic" the inner product. The derivation of the adjoint in these two equivalent methods may help the reader further understand that the adjoint $A^\dagger$ is really just a composition of fundamental linear functions $j : V \longrightarrow V^*$, $A^T : V^* \longrightarrow U^*$, and $i^{-1} : U^* \longrightarrow U$ that are all canonically created as soon as $A : U \longrightarrow V$ is created, along with the inner product spaces U and V.

$$\begin{array}{ccc} U & \xrightarrow{A} & V \\ \downarrow i & & \downarrow j \\ U^* & \xleftarrow[A^T]{} & V^* \end{array}$$

However, it is hard to grasp a visual intuition of adjoint operators in general. Note that the properties of the transpose indicate that given $A : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ with the standard orthonormal basis and dot product, the matrix representation of $A^\dagger$ is just A^T. If A is a matrix over $\mathbb{C}$, then $A^\dagger$ is $A^H \equiv \bar{A}^T$, the **Hermitian transpose**, or **conjugate transpose**, of A.

Note that this definition of the adjoint of linear operators is completely unrelated to the definition of an adjoint of a matrix!

We now describe one common application of adjoints.

Theorem 3.5

Let $A \in \text{Mat}(m \times n, \mathbb{R})$ with $m > n$. This means that the system of equations $Ax = p$ is an overdetermined system and will have no solutions with probability 1. However, we can find the **best-fit**

solution of the system. That is, the vector x that minimizes $\|Ax - p\|^2$ is the solution z of

$$A^\dagger A z = A^\dagger p \tag{100}$$

z is therefore, the "closest approximation" of the solution of $Ax = p$ that lives in $\mathbb{R}^n$.

Theorem 3.6

Let P_Y be the orthogonal projection onto Y. Then,

1. $P_Y = P_Y^2$.
2. $P_Y = P_Y^\dagger$.

Theorem 3.7 (Properties of the Adjoint)

Let $A, B : X \longrightarrow U$, $C : U \longrightarrow V$ be linear mappings. Then,

1. $(A + B)^\dagger = A^\dagger + B^\dagger$
2. $(CA)^\dagger = A^\dagger C^\dagger$
3. $(A^{-1})^\dagger = (A^\dagger)^{-1}$ if A is bijective
4. $(A^\dagger)^\dagger = A$

Definition 3.7

Linear mapping A is **self adjoint** if and only if $A = A^\dagger$. If M is any linear mapping, then its self-adjoint part is

$$M_\delta = \frac{M + M^\dagger}{2} \tag{101}$$

Theorem 3.8 (Spectral Theorem)

A n-dimensional self-adjoint map H over $\mathbb{C}$ has real eigenvalues and an orthonormal basis of genuine eigenvectors. That is, its eigendecomposition consists of n pairwise orthogonal eigenspaces.

Corollary 3.9

Given a real self-adjoint matrix H, there exists a real invertible matrix M such that $M^\dagger H M = D$, with D diagonal and the column vectors form an orthonormal basis.
So, given self-adjoint $H : X \longrightarrow X$, the whole space can be written as the direct sum of pairwise orthogonal eigenspaces.

$$X = \bigoplus_{i=1}^{n} E(\lambda_i) \tag{102}$$

which implies that every $x \in X$ can be written uniquely as

$$x = x_1 + x_2 + \cdots + x_n,\ x_i \in E(\lambda_i) \tag{103}$$

Definition 3.8

Given that P_j is the orthogonal projection onto the jth eigenspace $E(\lambda_j)$, that is

$$P_j(x) = x_j \in E(\lambda_j), \;\; (P_j \text{ also self adjoint}) \tag{104}$$

the **spectral resolution** of self-adjoint mapping H is the decomposition into the form

$$H = \sum_j \lambda_j P_j \implies Hx = \bigg(\sum_j \lambda_j P_j\bigg) x = \sum_j \lambda_j x_j \tag{105}$$

The resolution of the identity is

$$I = \sum_j P_j \tag{106}$$

Theorem 3.10

Given the spectral resolution of self-adjoint H,

$$H = \sum_j \lambda_j P_j \implies H^2 = \sum_j \lambda_j^2 P_j \tag{107}$$

Note that the spectral resolution of a self adjoint mapping is precisely the eigendecomposition of the mapping into its 1-dimensional eigenspaces. It is merely a simpler form of the eigendecomposition in the specific case when the linear mapping is self-adjoint.

Theorem 3.11

Let H, K be self-adjoint mappings such that $HK = KH$. Then H and K have the same spectral resolution, i.e. they have the same eigendecomposition.

$$H = \sum_j a_j P_j, \;\; K = \sum_j b_j P_j \tag{108}$$

Proof. $x \in E(a) Hx = ax \implies KHx = aKx \implies HKx = aKx \implies Kx \in E(a)$. Similarly, we can do this with K to find $x \in E(a) \implies Hx \in E(a)$, meaning that K and H have the same eigendecompositions (though their eigenvalues are not necessarily equal).

Definition 3.9 (Anti-Self-Adjoint)

Map A is **anti-self adjoint** if $A^\dagger = -A$. Conjugate symmetry implies that

$$A^\dagger = A \iff (iA)^\dagger = -(iA) \tag{109}$$

So, given an anti-self adjoint map A, we can apply the spectral resolution to iA.

Theorem 3.12

Given anti-self adjoint $A : \mathbb{C}^n \longrightarrow \mathbb{C}^n$,

1. eigenvalues of A are purely imaginary
2. we can choose an orthonormal basis of eigenvectors of A

Proof. This easily follows from the Spectral Theorem.

Definition 3.10 (Normal Maps)

$N : X \longrightarrow X$ is a **normal mapping** if $N^\dagger N = NN^\dagger$. Self-adjoint, anti-self adjoint, and unitary matrices are all normal. Surprisingly, the set of normal matrices are not closed under addition nor multiplication, so they do not form a group.

Theorem 3.13

A map N is normal if and only if it has an orthonormal basis of eigenvectors, i.e. it is unitarily diagonalizable. That is,

$$N = U^\dagger D U \tag{110}$$

Proof. ($\rightarrow$) Let

$$H = \frac{1}{2}(N + N^\dagger),\ A = \frac{1}{2}(N - N^\dagger) \tag{111}$$

$N^\dagger N = NN^\dagger \implies AH = HA$, where H is self adjoint, A is anti-self adjoint, and $N = H + A, N^\dagger = H - A$. Since $AH = HA$, they have the same spectral resolution of orthonormal eigenspaces, which also forms the same spectral resolution for $N = H + A$.
($\leftarrow$) $A = U^\dagger D U \implies A^\dagger A = (U^\dagger D U)(U^\dagger \bar{D} U) = U^\dagger D \bar{D} U = AA^\dagger$.

3.3 Orthogonal Projections

The concept of orthogonality also allows us to define orthogonal projections onto a vector or subspace.

Definition 3.11 (Orthogonal Projectjion)

Let $x \in X$ and let Y be a subspace of X. Then x can be decomposed into the form $x = y + z,\ y \in Y, z \in Y^\perp$. The **orthogonal projection** of x onto Y is then defined as

$$\text{proj}_Y(x) = y \tag{112}$$

Orthogonal projections are linear transformations.

Theorem 3.14

Given that $x \in \mathbb{R}^n$ is projected onto a 1-dimensional subspace Y. The orthogonal projection of x onto Y can be computed with the formula

$$\text{proj}_Y(x) = \frac{x \cdot y}{||y||^2} y \tag{113}$$

where y is an arbitrary vector in Y and $\cdot$ is the dot product in $\mathbb{R}^n$. Furthermore, for a k-dimensional subspace Y, we can calculate the projection by first adding up the projections of x onto a set of basis vectors of Y and then adding them up. That is, given basis $r_1, r_2, \cdots, r_k$ of Y,

$$\text{proj}_Y(x) = \sum_{i=1}^k \text{proj}_{r_i}(x) = \sum_{i=1}^k \frac{x \cdot r_i}{||r_i||^2} r_i \tag{114}$$

Given that we have an orthonormal basis $\{r_i\}_{i=1}^k$ of subspace Y in $\mathbb{R}^n$, we can more simply define

$$\text{proj}_Y(x) = \sum_{i=1}^k (x \cdot r_i) r_i \tag{115}$$

3.4 Isometries

Definition 3.12 (Isometry)

An **isometry** M of metric space (X, d) is a mapping that preserves all distances. That is, for all $x, y \in X$,

$$d(x, y) = d(Mx, My) \tag{116}$$

The set of all isometries, denoted $\text{Isom}(X)$, is a group that is generated by all translations, rotations, and reflections.

Since linear maps always preserve the origin, we will focus on origin-preserving isometries, which is a subgroup called the orthogonal group.

Definition 3.13 (Orthogonal Group)

The **orthogonal group** of a real Euclidean space of dimension n, denoted $O(n)$, is the group of all origin-preserving isometries of the space consisting of rotations and reflections. The matrix representation of this group is the set of real $n \times n$ matrices where the column vectors form an orthonormal basis. Note that the determinant of every element of $O(n)$ is ± 1.

Definition 3.14 (Orthogonal Matrix)

An **orthogonal matrix** is the matrix representation of an element in $O(n)$. It is the real $n \times n$ matrix where all the column vectors are pairwise orthogonal and all have magnitude 1.

Theorem 3.15

The rows of an orthogonal matrix are also pairwise orthonormal.

Theorem 3.16

Given an orthogonal matrix M,

$$M^T = M^{-1} \tag{117}$$

Definition 3.15 (Special Orthogonal Group)

The **special orthogonal group** of a real Euclidean space of dimension n, denoted $SO(n)$, is the group of all isometries that preserve the handedness of the space consisting only of rotations. It is a subgroup of $O(n)$. The matrix representation of this group is the set of real $n \times n$ matrices where the column vectors are pairwise orthonormal and the determinant $= 1$.

We extend this concept to complex Euclidean spaces.

Definition 3.16 (Unitary Group)

The **unitary group of degree** n is the group of all complex $n \times n$ matrices where the columns are pairwise orthogonal. It is denoted $U(n)$.

Example 3.3

$U(1)$ is the set of complex numbers with norm 1.

Definition 3.17 (Special Unitary Group)

The **special unitary group of degree** n is the group of all complex $n \times n$ matrices where the columns are pairwise orthogonal and determinant $= 1$. It is denoted $SU(n)$.

The groups mentioned in this section are examples of *Lie Groups*. Lie groups in general will not be defined in here, since they require knowledge of smooth manifolds and differential geometry. In order to analyze these abstract groups, we use the exponential map $e \in \text{End}(\text{ Mat}(n, \mathbb{F})$ to reduce these Lie groups to Lie algebras.

3.5 Dual Inner Product Spaces

Theorem 3.17

The inner product endowed on V induces a natural isomorphism between V and V^*.

Proof. We fix $y \in V$ and simply define the isomorphism to be.

$$l(y) \equiv (x, y) \tag{118}$$

which defines a bijection between $x \in V$ and $l \in V^*$.

Note that given vector spaces U, V, the set of all linear mappings $A : U \longrightarrow V$ also forms a vector space. More specifically, it is a rank (1,1) tensor product space. This means that we can define similar Euclidean structures on them. The norm of a matrix is worth mentioning. Note that the structures and concepts of metrics, norms, inner products, distances, magnitudes, orthogonality, and basis are not intrinsic properties of the vector space. So, we will not assume the existence of these structures unless otherwise stated or explicitly implied.

4 Matrices

We now describe the construction of the matrix realization of a linear map from $V \longrightarrow U$. In order to do this, we *must* define a basis for each V and U. If $V = U$, then we usually define the same basis for both the domain and codomain.

Let the basis for U be $\{u_1, u_2, \cdots, u_n\}$ and the basis of V be $\{v_1, v_2, \cdots, v_m\}$. In fact, the assignment of this specific basis is a linear map in of itself. That is,

$$\begin{aligned} i : U \longrightarrow \mathbb{F}^n, i(u_\alpha) = e_\alpha \\ j : V \longrightarrow \mathbb{F}^m, j(v_\beta) = e_\beta \end{aligned}$$

However, we do not usually include this transformation in the notation. We just denote $i(u)$ as u and $j(v)$ as v. Every vector $u \in U$ can then be represented as a linear combination

$$u = \sum_{j=1}^{n} c_j u_j \tag{119}$$

By linearity of the mapping $A : U \longrightarrow V$,

$$Au = A\bigg(\sum_{j=1}^{n} c_j u_j\bigg) = \sum_{j=1}^{n} c_j A u_j \tag{120}$$

This means that A can be completely, uniquely determined by defining how it maps the n basis vectors $u_j \in U$, that is, by defining the values

$$Au_1, Au_2, \cdots, Au_{n-1}, Au_n \tag{121}$$

Each Au_j will be an element of V, which means that Au_j can be decomposed into the linear combination of v_i's. That is,

$$Au_j = \sum_{i=1}^{m} a_{ij} v_i, \ j = 1, 2, \cdots, n \tag{122}$$

We are done. Given the basis of the domain and codomain, the elements a_{ij} are precisely the entries of the $m \times n$ matrix $(1 \leq i \leq m, 1 \leq j \leq n)$.

$$v = Au \iff \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_n \end{pmatrix} \tag{123}$$

It is important to note that the matrix is *not* A in of itself. In the most rigorous sense, the matrix A is really just equal to the composition of mappings $j^{-1}Ai$, but for simplicity it is just written as A. It is just one representation of a linear map given the two bases of the domain and codomain. Furthermore, as soon as one writes down a matrix to represent a linear map, they are automatically assuming some choice of basis given by i and j.

Definition 4.1

The **algebra** of $n \times n$ matrices over field $\mathbb{F}$, denoted $\text{Mat}(n, \mathbb{F})$, is defined with regular matrix addition and multiplication.

Furthermore, we can define the mapping between linear operators $T : \mathbb{F}^n \longrightarrow \mathbb{F}^m$ and $m \times n$ matrices (given that there is a basis for both $\mathbb{F}^n, \mathbb{F}^m$.

Definition 4.2

The linear mapping between the algebras

$$\rho : \text{Hom}(\mathbb{F}^n, \mathbb{F}^m) \longrightarrow \text{Mat}(m \times n, \mathbb{F}) \tag{124}$$

is a multiplicative group homomorphism. This mapping that assigns abstract group elements of linear mappings to matrices is called a **representation**.

Theorem 4.1

$\text{Mat}(n, \mathbb{F}) \simeq \text{End}(\mathbb{F}^n)$

Proof. A matrix is completely determined by the basis mapping i. By definition, a linear mapping over $\mathbb{F}$ is a basis mapping if and only if it is an element of $\text{End}(\mathbb{F}^n)$.

Note that the composition operation in the algebra of linear operators is realized as the operation of matrix multiplication. These are two distinct operations that are related only through the basis mappings i and j.

Example 4.1

Let $\alpha : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be the linear transformation of the counterclockwise rotation by θ and $\beta : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ be the counterclockwise rotation of ϕ. Then the matrix representation of $\alpha \circ \beta$ is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{125}$$

$$= \begin{pmatrix} \cos\theta\cos\phi - \sin\theta\sin\phi & -\sin\phi\cos\theta - \cos\phi\sin\theta \\ \sin\theta\cos\phi + \cos\theta\sin\phi & -\sin\theta\sin\phi + \cos\theta\cos\phi \end{pmatrix} \tag{126}$$

But the counterclockwise rotation by θ and then ϕ is really just a counterclockwise rotation by $\theta + \phi$, which has the matrix representation

$$\begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix} \tag{127}$$

Since both matrices must be equivalent, this produces the trigonometric identities for angle addition.

$$\sin(\theta+\phi) = \sin\theta\cos\phi + \cos\theta\sin\phi$$
$$\cos(\theta+\phi) = \cos\theta\cos\phi - \sin\theta\sin\phi$$

Theorem 4.2

Given mappings $A_i \in \text{End}(V_i)$ for $i = 1, 2, \cdots, n$, the matrix representation of the induced linear mapping $A_1 \oplus A_2 \oplus \cdots \oplus A_n$ is the block matrix

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_n \end{pmatrix} : \bigoplus_{i=1}^{n} V_i \longrightarrow \bigoplus_{i=1}^{n} V_i \tag{128}$$

4.1 Change of Basis

Definition 4.3 (Active, Passive Transformation)

A linear transformation A that maps every vector from U to a vector in V is called an **active transformation**. However, a **passive transformation**, or a **change of basis transformation**, linearly transforms the set of basis vectors to another set of basis vectors within the same space. That is, a passive transformation takes the components of a vector v with respect to basis $\{e_1, e_2, \cdots, e_n\}$ and merely represents v with respect to another set of basis $\{f_1, f_2, \cdots, f_n\}$.

It is obvious that a passive transformation in V is an element of $\text{End}(V)$. But note that an element of $\text{End}(V)$ could be interpreted *both* as a passive and active transformation. Usually, the context will make it clear whether we are interpreting a transformation as passive or active. We now provide the construction of the change of basis.

Suppose $e_1, e_2, \cdots, e_n$ is a basis for vector space V and $f_1, f_2, \cdots, f_n$ is another basis for V. So, every basis vector f_i can be presented as a linear combination of the old basis vectors.

$$f_j = \sum_{i=1}^{n} s_{ij} e_i \quad \text{for all i, j} \tag{129}$$

A general vector $x \in V$ will transform as such

$$\begin{aligned} x &= \sum_j y_j f_j \quad \text{for } y_1, y_2, \cdots \in \mathbb{F} \\ &= \sum_{i,j} y_j s_{ij} e_i \\ &= \sum_i \Big(\sum_j s_{ij} y_j \Big) e_i \\ &= \sum_i x_i e_i \implies x_i = \sum_j s_{ij} y_j \end{aligned} \tag{130}$$

Similarly to the process of how we constructed matrix representations of linear operators, this process makes it clear that s_{ij} are the entries of the $n \times n$ matrix representation of the passive mapping S. The final line of the equation above can be expressed, in terms of matrices, as

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ \cdots \\ x_n \end{pmatrix} = \begin{pmatrix} & & \\ & & \\ & S & \\ & & \\ & & \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ \cdots \\ y_n \end{pmatrix} \tag{131}$$

This is a change of basis, since both the coefficients x_i and y_i represent the same vector x in V, but through a different basis determined by S. Note that S must be an invertible matrix since we are mapping bases to bases. So, given that $x = Sy$, if $Ax = b$ is a matrix equation, then

$$Ax = b \implies ASy = Sb' \implies S^{-1}ASy = b' \tag{132}$$

where b' is the set of new coefficients for the vector with respect to the basis induced by S. This leads to the concept of matrix similarities. We once again note that whenever we create a matrix as an $m \times n$ entry of numbers, we are intuitively fixing a basis (not necessarily orthonormal, even) for the vectors that the matrix is transforming on. For example, the matrix A in $y' = Ax'$ transforms the vector x' with respect to the basis which x' is in, i.e. the basis $e_1', e_2', \cdots, e_n'$. This transformation is not the same if it were to act on the vector x, which is determined by the basis $e_1, e_2, \cdots, e_n$. Therefore, we must also "change" the matrix A acting on x' in order to account for the change in basis from x' to x. This change is

$$A \rightarrow B = SAS^{-1} \tag{133}$$

where matrix A represents the transformation with respect to basis formed by the column vectors of S, and B represents the same transformation with respect to the basis formed by the column vectors of S^{-1}.

Definition 4.4 (Similar Matrices)

Two matrices A and B are **similar** if and only if there exists an invertible matrix S such that $B = SAS^{-1}$. A and B both represent the same transformation T but merely in different bases. Matrix similarity is a relation that partitions the n^2-dimensional matrix algebra $\text{Mat}(n, \mathbb{R})$ into similarity classes.

4.2 Solving Systems of Equations

Definition 4.5 (Linear System of Equations)

Fix a field $\mathbb{F}$. A **linear equation** with variables $x_1, x_2, \cdots, x_n$ is in the form

$$a_1 x_1 + a_2 x_2 + a_3 x_3 + \cdots + a_n x_n = b \tag{134}$$

where the **coefficients** a_i and the **free term** b belong to $\mathbb{F}$. If $b = 0$, then (3) is called a **homogeneous equation** and if $b \neq 0$, then it is called a **inhomogeneous equation**.

A system of m linear equations with n variables has the following general form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots &= \cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

By matrix multiplication, this system is equal to the matrix equation $Ax = b$.

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{135}$$

That is, given a linear transformation $A : \mathbb{F}^n \longrightarrow \mathbb{F}^m$ and a vector $b \in \mathbb{F}^m$, we must find the preimage of b under A. Clearly, x is a solution of this matrix equation if and only if it is a solution of the system of equations.

We can interpret this matrix equation in two ways. First, we introduce the *hyperplane interpertation.* The solution to each linear equation of n variables represents an affine hyperplane in $\mathbb{F}^n$. Therefore, the solutions to the system of m linear equations is simply the intersection of the m affine hyperplanes of dimension $n-1$ within $\mathbb{R}^n$. That is, x is a solution of $Ax = b$ if and only if

$$x \in \bigcap_{i=1}^{m} \Big\{ (x_1, x_2, \cdots, x_n) \mid \sum_{j=1}^{n} a_{ij} x_j = b_i \Big\} \tag{136}$$

The *column space interpretation* presents $Ax = b$ in this equivalent form.

$$x_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix} + x_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix} + \ldots + x_n \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \tag{137}$$

That is, the solutions $x_1, x_2, \cdots, x_n$ are precisely the coefficients of the linear combination of the column vectors of A that add up to vector b. Equivalently, it is the realization of vector b with respect to the coordinate system of the column vectors of A. Note that the column space need not be a basis of $\mathbb{F}^m$. It does not need to be linearly independent nor does it need to span $\mathbb{F}^n$.

Definition 4.6 (Coefficient Matrix)

The matrix A under the system is called the **coefficient matrix** and the matrix

$$\tilde{A} \equiv \begin{pmatrix} | & | & \cdots & | & | \\ a_1 & a_2 & \cdots & a_n & b \\ | & | & \cdots & | & | \end{pmatrix} \equiv \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{pmatrix} \tag{138}$$

is called the **extended matrix**.

Definition 4.7

A system of equations is called **compatible** if it has at least one solution and **incompatible** otherwise.

Definition 4.8 (Elementary Transformation)

An **elementary transformation** of a system of linear equation is one of the following three types of transformations

1. adding an equation multiplied by a number to another *later* equation
2. interchanging two equations
3. multiplying an equation by a nonzero number

Definition 4.9 (Elementary Row Transformation)

An **elementary row transformation** of a matrix is one of the following three types of transformations

1. adding a row multiplied by a number to another *later* row
2. interchanging two rows
3. multiplying a row by a nonzero number

Clearly, these two definitions are equivalent since every elementary transformation of a system has a corresponding row transformation in its extended matrix. Given the ith row of a matrix, a "later" row means the jth row, where $j > i$. Defining property (i) to add to a later row does not actually restrict where we can add rows to, since property (ii) allows us to add any scalar multiple of any row to any other row. We define it this way for future convenience in defining the LUP Decomposition.

Definition 4.10

The elementary transformations on a $m \times n$ matrix A is equivalent to left matrix multiplication by the following $m \times m$ matrices. Due to the following difficulty in presenting these matrices in a general form, we present them in the specific 4×4 case and hope that the reader can extrapolate this process to general matrices.

1. Adding row i multiplied by scalar α to row j (where $j > i$) is denoted $E^1_{\alpha \times i + j}$. The matrix is the

identity matrix with α in the (j, i) position.

$$E^1_{2\times 1+2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E^1_{-3\times 2+4} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -3 & 0 & 1 \end{pmatrix} \tag{139}$$

2. Interchanging the ith and jth row is denoted by matrix E^2_{ij}. Note that these are permutation matrices, or more speficially, transpositions.

$$E^2_{23} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E^2_{24} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \tag{140}$$

3. Multiplying the ith row by a scalar α is denoted by matrix $E^3_{\alpha\times i}$.

$$E^3_{3\times 3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad E^3_{7\times 1} = \begin{pmatrix} 7 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{141}$$

Theorem 4.3

Each elementary matrix is invertible and their inverses are also elementary matrices. More specifically,

1. $(E^1_{\alpha\times i+j})^{-1} = E^1_{-\alpha\times i+j}$ (same matrix but α changed to -α)
2. $(E^2_{ij})^{-1} = E^2_{ij}$ (same matrix)
3. $(E^3_{\alpha\times i})^{-1} = E^3_{(1/\alpha)\times i}$ (same matrix but α changed to $1/\alpha$)

Elementary column operations of are equivalent to right multiplication of matrices.

Definition 4.11 (Pivot)

The **pivot** of a row $(a_1, a_2, \cdots, a_n)$ is its first nonzero element. If this element is a_k, then k is the **index** of the pivot.

Definition 4.12 (Echelon Form)

A matrix is in **Echelon form**, or **row Echelon form**, if

1. the indices of the pivots of its nonzero rows form a strictly increasing sequence, like steps
2. zero rows, if they exist, are at the bottom

Thus, a matrix in Echelon form is in form

$$\begin{pmatrix} a_{1j_1} & * & \dots & \dots & * \\ 0 & a_{2j_2} & * & \dots & * \\ 0 & 0 & \ddots & \dots & * \\ 0 & 0 & 0 & a_{rj_r} & \vdots \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \tag{142}$$

where $*$'s represent arbitrary numbers, a_{ij_i}'s are nonzero (with indices j_i, and the entries to the left of below them are 0. Property (i) also states that $j_1 < j_2 < \cdots < j_r$. Let us denote the Echelon form of matrix A as ref(A).

Theorem 4.4

Every matrix can be reduced to step form by elementary row transformations.

Proof. The relevant algorithm used will not be shown here, but we will mention that this procedure is called **Gauss Elimination**, or **row reduction**.

The computational efficiency of Gauss Elimination is well known. Solving a system of n equations with n variables with this algorithm requires approximately $2n^3/3$ operations, meaning that it has arithmetic complexity of $O(n^3)$. However, for matrices of large order, multiple problems can occur.

The algorithm generally does not have memory problems if the field is finite or if the coefficients are floating-point numbers. However, if the coefficients are integers or rational numbers, the intermediate entries of the algorithm can grow exponentially large, so bit complexity is exponential. However, there is a variant of Gaussian elimination, called the Bareiss algorithm, that avoids this problem, but has bit complexity of $O(n^5)$. Another problem is numerical instability, caused by the possibility of dividing by numbers very close to 0. Any such number would have its existing error amplified. Gaussian elimination algorithm is generally known to be stable for positive-definite matrices.

Under the column space interpretation, Gaussian Elimination is really just an algorithm that performs a change of basis in steps. Each elementary operation simultaneously changes all of the vectors of the column space in such a way that eventually, this set of vectors will be "nice-looking" with a lot of zero entries. Under the hyperplane interpretation, it is a bit harder to visualize, but it is sufficient to say that each elementary operation either "stretches/compresses" (iii) a hyperplane or it "rotates" (i) the hyperplane around the axis where the solution exists. Either way, the intersection between the hyperplane and the set of solutions do not change.

Definition 4.13 (Step Form)

A system of linear equations is said to be in **step form** if its extended matrix is in Echelon form.

Definition 4.14

A matrix is in **reduced row echelon form**, denoted rref(A), if

1. it is in row echelon form
2. the pivots are all equal to 1
3. each column containing a pivot has zeros in all other entries

Theorem 4.5

Every matrix can be reduced to reduced row echelon form by elementary row operations.

Proof. We will briefly describe the method to do this. We first reduce matrix A to step form. Then, we perform the algorithm known as **back substitution**, where we start with the bottom row and use elementary operations to cancel out terms in upper rows.

Definition 4.15

A system of linear equations is said to be **solved** if its extended matrix is in reduced row echelon form.

Definition 4.16

A matrix is called **lower triangular** if $a_{ij} = 0$ for $i < j$. It is called **upper triangular** if $a_{ij} = 0$ for $i > j$. A square matrix is **diagonal** if $a_{ij} = 0$ for $i \neq j$.

Theorem 4.6

Elementary operations on either a system of linear equations or its extended matrix does not change its solutions.

Proof. It is easy to see this is true when performing the computations with the three transformations. We can prove this more abstractly (tbd): Given the system $Ax = b$ with $x \in \mathbb{F}^n, b \in \mathbb{F}^m$. We see that $A \in \text{Mat}(m \times n, \mathbb{F}) \implies \tilde{A} \in \text{Mat}(m \times (n+1), \mathbb{F})$. Each elementary row transformation on $\tilde{A}$, denote it E, is a bijective mapping. Let us define the mapping

$$\text{sol} : \text{Mat}\big(m \times (n+1), \mathbb{F}\big) \longrightarrow 2^{\mathbb{F}^n},\ \text{sol}\begin{pmatrix} A & b \end{pmatrix} \equiv \{x \in \mathbb{F}^n \mid Ax = b\} \tag{143}$$

where $2^{\mathbb{F}^n}$ is the set of all subsets of $\mathbb{F}^n$. By matrix multiplication, we see that

$$E\begin{pmatrix} A & b \end{pmatrix} = \begin{pmatrix} EA & Eb \end{pmatrix} \tag{144}$$

Since E is bijective, it is invertible. So,

$$\begin{aligned} \text{sol}\Big(E\begin{pmatrix} A & b \end{pmatrix}\Big) &= \text{sol}\begin{pmatrix} EA & Eb \end{pmatrix} && (145) \\ &= \{x \mid EAx = Eb\} && (146) \\ &= \{x \mid Ax = b\} && (147) \\ &= \text{sol}\begin{pmatrix} A & b \end{pmatrix} && (148) \end{aligned}$$

Note the importance of this theorem. This result is the foundation behind the applications of Jordan Elimination.

Definition 4.17

A linear system can have either have no possible solutions (**overdetermined**), one unique solution, or multiple solutions (**underdetermined**) (infinite solutions if $\text{char}\,\mathbb{F} = 0$). We can say with probability 1 that given a random $m \times n$ matrix A with random m-dimensional vector b, the system $Ax = b$ has

1. 0 solutions if $m > n$, since there are more equations than variables
2. 1 solution if $m = n$ with the same number of equations and variables
3. Infinite solutions if $m < n$ since there are more variables than equations

Definition 4.18

The variables corresponding to the indices of the pivots are called the **pivot variables**. The rest of the variables are called **free variables**

Because of theorem 3.3, we can determine whether a system has 0, 1, or multiple solutions by looking at the extended matrix's Echelon form. The case for 0 solutions is easy.

Theorem 4.7

The system $Ax = b$ has 0 solutions if and only if $\text{ref}(\tilde{A})$ contains a row in the form

$$\begin{pmatrix} 0 & 0 & \cdots & 0 & c \end{pmatrix}, \; c \neq 0 \tag{149}$$

Proof. The existence of this row is equivalent to the linear equation

$$0x_1 + 0x_2 + \cdots + 0x_n = c, \; c \neq 0 \tag{150}$$

which is absurd and cannot have any solution. Under the hyperplane interpretation, we can visualize all the hyperplanes failing to have a common point.

Corollary 4.8

Given $m \times n$ matrix A, if $m > n$ and the row vectors of A are all linearly independent, then the system $Ax = b$ has 0 solutions.

Theorem 4.9

The system $Ax = b$ has 1 solution if and only if $\text{ref}(A)$ is diagonal.

Proof. $\text{ref}(A)$ being diagonal implies that there exists at least one solution and also implies the absence of any free variables.

Theorem 4.10

The system $Ax = b$ has multiple solutions if and only if $\text{ref}(A)$ has free variables.

Proof. Clear.

Definition 4.19 (Rank)

The number of pivots in $\text{ref}(A)$ is called the **rank** of A, denoted $\text{rk}(A)$.

Theorem 4.11

Let A be a $m \times n$ matrix. Then $\text{rk}(A) \leq \min\{m, n\}$.

Proof. By definition, the number of pivots cannot exceed the number of variables nor can it exceed the number of equations.

Definition 4.20

A $n \times n$ matrix A is called **nonsingular** if and only if $\text{rk}(A) = n$. It is **singular** if and only if $\text{rk}(A) < n$. Clearly, $\text{rk}(A) \not> n$.

4.3 Four Fundamental Spaces

We will begin to bring over the general concepts of linear transformations and state them within the realm of matrices. We will start with the concept of dual vectors.

It is customary to interpret vectors in the abstract sense as a column of n numbers. Given that vectors are column vectors, it is sometimes useful (but not entirely comprehensive) to interpret covectors as row vectors. That is, given a vector v and covector l, l linearly maps v to a field element by left matrix multiplication.

$$l(v) = \begin{pmatrix} l_1 & l_2 & \cdots & l_n \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{pmatrix} = \sum_{i=1}^{n} l_i v_i \tag{151}$$

Definition 4.21 (Transpose of a Matrix)

The **transpose** of matrix A, denoted A^T, is the matrix with entries $(a^T)_{ij} = a_{ij}$. That is, it is A, "flipped over."

We illustrate why this definition of a transpose is equivalent to the abstract definition to the transpose of a linear map. Given a linear map $A : U \longrightarrow V$ with $\dim U = n, \dim V = m$, we can fix a basis on both U and V to define its matrix A. The abstract definition states that

$$A^T : V^* \longrightarrow U^*,\ l \equiv \varphi A \tag{152}$$

Treating l and φ as row vectors, we can see that the $m \times n$ matrix A maps the $1 \times m$ covector φ to the $1 \times n$ covector l. Note that this linear mapping is realized through *right multiplication* of A on φ. It is customary to present linear maps as *left* multiplication, so by "flipping" (i.e. taking the matrix transpose) of all the elements in the equation, we get

$$l^T \equiv A^T \varphi^T \tag{153}$$

which presents the mapping in the more usual way of left matrix multiplication. Note that l^T and φ^T are still covectors. Just because they are now represented as column vectors, it does not mean that they are not covectors, which is why we shouldn't be too dependent on the row vector interpretation of dual vectors mentioned above.

Continuing the previous point, note that the way we represent vectors and linear transformation has all been arbitrarily chosen. There is nothing innate about the way we express these transformation as matrix multiplication. This last example especially shows us that the entire definition of the matrix transpose (rooting from the abstract definition) is dependent on our *initial choice* to represent linear mappings as *left* matrix multiplication and to represent all vectors as column vectors.

Theorem 4.12 (Properties of the Transpose)

Given that $A, B : U \longrightarrow V$ is linear, c a constant

1. $(A^T)^T = A$.
2. $(A + B)^T = A^T + B^T,\ (cA)^T = cA^T$.
3. $(AB)^T = B^T A^T$.
4. If A is invertible, $(A^{-1})^T = (A^T)^{-1}$ and A invertible $\implies A^T$ invertible.
5. $x \cdot y = x^T y$. Furthermore,

$$Ax \cdot y = (Ax)^T y = x^T A^T y = x \cdot A^T y \tag{154}$$

Definition 4.22

Matrix A is a **symmetric matrix** if $A = A^T$. A is **skew-symmetric**, or **anti-symmetric**, if $A^T = -A$.

Now we are ready to describe the four fundamental spaces of a matrix A: the column space, row space, nullspace, and left nullspace. All four of these spaces are subspaces, but we will not check them here.

Definition 4.23 (Column Space)

The **column space** of matrix A, denoted $C(A)$, is the span of its column vectors. That is,

$$C(A) = \text{span}\{a_1, a_2, \cdots, a_n\} \tag{155}$$

We will denote the column vectors with lowercase a_i's.

Definition 4.24 (Row Space)

The **row space** of matrix A, denoted $R(A)$, is the span of its row vectors. That is,

$$R(A) = \text{span}\{A_1, A_2, \cdots, A_m\} \tag{156}$$

We will denote the row vectors with uppercase A_i's.

Definition 4.25 (Null Space)

The kernel of linear transformation is called the **nullspace** of its associated matrix, denoted $\text{Null}(A)$.

Definition 4.26

The **left nullspace** of matrix A is the nullspace of A^T. It is denoted $\text{Null}(A^T)$.

Theorem 4.13

By the column space interpretation, it is clear that

$$C(A) = \text{Im } A \tag{157}$$

We state the matrix analogue of Theorem 2.5.

Theorem 4.14

A vector is a solution to the system of equation $Ax = b$ if and only if it is of the form

$$a + \text{Null } (A) \tag{158}$$

where a is one solution.

Theorem 4.15

Let $A : \mathbb{F}^n \longrightarrow \mathbb{F}^m$ be a $m \times n$ matrix with rank k. Assuming that $\mathbb{F}^n$ and $\mathbb{F}^m$ are inner product spaces,

$$Null(A) = R(A)^\perp \iff Null(A)^\perp = R(A) \tag{159}$$
$$Null(A^T) = C(A)^\perp \iff Null(A^T)^\perp = C(A) \tag{160}$$

That is, $\text{Null}(A)$ and $R(A)$ are orthogonal complements in $\mathbb{F}^n$, with $\dim R(A) = k$ and $\dim \text{Null}(A) = n-k$. $\text{Null}(A^T)$ and $C(A)$ are orthogonal complements in $\mathbb{F}^m$, with $\dim C(A) = k$ and $\dim \text{Null}(A^T) = m-k$.

Corollary 4.16

The solution to the homogeneous system $Ax = 0$ is precisely $\text{Null}(A)$.

Definition 4.27

The homogeneous system $Ax = 0$ always has a *trivial solution* $x = 0$.

Example 4.2

Given a system of linear equations

$$\begin{aligned} x + 3y - 2z &= 5 \\ 3x + 5y + 6z &= 7 \\ 2x + 4y + 3z &= 8 \end{aligned}$$

We put it into extended matrix form A and perform Gauss Elimination to get $\text{rref}(A)$.

$$\begin{pmatrix} 1 & 3 & -2 & 5 \\ 3 & 5 & 6 & 7 \\ 2 & 4 & 3 & 8 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 3 & -2 & 5 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & 1 & 2 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 & 0 & -15 \\ 0 & 1 & 0 & 8 \\ 0 & 0 & 1 & 2 \end{pmatrix} \tag{161}$$

So, $\text{rref}(A)$ has the solution $(-15, 8, 2)$ and it is unique because there are no free variables.

This leads to the following theorem.

Theorem 4.17

The set of n linear equations with n variables can be expressed in the form of $Ax = b$, where A is an $n \times n$ matrix.

$$Ax = b \text{ has a unique solution} \iff A \text{ is nonsingular} \iff \text{rk}\,(A) = n \tag{162}$$

Proof. A is nonsingular is equivalent to saying that $\text{rref}(A) = I_n$, where I_n is the $n \times n$ identity matrix. This clearly means that $\text{rref}(\tilde{A})$ will always reveal unique solutions.

Theorem 4.18

$n \times n$ matrix A is invertible if and only if it is nonsingular.

Proof. A is nonsingular $\iff$ $Ax = b$ will always have a unique solution $\iff$ A is an isomorphism from $\mathbb{F}^n$ to itself $\iff$ by definition, A is invertible.

The realization of an endomorphism of $\mathbb{F}^n$ in matrix form is a $n \times n$ matrix. The realization of an automorphism of $\mathbb{F}^n$ in matrix form is an $n \times n$ nonsingular matrix. This set is actually a multiplicative, nonabelian group denoted $\mathrm{GL}_n(\mathbb{F})$ and is one example of a Lie Group.

Theorem 4.19

There are k free variables in A if and only if $\dim \mathrm{Null}(A) = k$.

Proof. We do not give a rigorous proof but we outline one. Each free variable corresponds to a free vector in the row echelon form of A that are all linearly independent. Since the span of these free vectors is equal to $\mathrm{Null}(A)$, the k vectors form a basis of $A \implies$ by definition, $\dim \mathrm{Null}(A) = k$.

Theorem 4.20

$$\mathrm{rk}(A) = \dim \text{ Im } A = \dim C(A) \tag{163}$$

Proof. Let A be a $m \times n$ matrix over $\mathbb{F}$. Then, let $\mathrm{rk}(A) = k$, which implies that there are $n - k$ free variables $\implies$ dim $\mathrm{Null}(A) = n - k$. By rank nullity,

$$\dim \mathrm{Im}\, A = n - \dim \mathrm{Null}(A) = n - (n - k) = k = \mathrm{rk} A \tag{164}$$

This theorem establishes the consistency in definition between the rank of an abstract mapping mentioned in chapter 2 and the rank of its matrix representation. We can in fact establish strong claims on top of this.

Theorem 4.21

$$\dim C(A) = \dim R(A) \tag{165}$$

Proof. Let A be a $m \times n$ matrix of rank r. There are r pivots and a pivot in each nonzero row of $\mathrm{ref}(A)$, so $\dim R(A) = r$. The previous theorem says $r = \dim C(A)$.

Corollary 4.22

$$C(A) \simeq R(A) \tag{166}$$

Proof. While this is a direct result of the dimensions of the two subspaces being equal, it is worthwhile to mention this alternative proof. We will prove that the linear mapping A is the isomorphism itself. Let $\mathrm{rk}(A) = r$ and let $\{v_1, v_2, \cdots, v_r\}$ be a basis for $R(A)$. Then, the set $\{Av_1, Av_2, \cdots, Av_r\}$ are r vectors in $C(A)$. They are linearly independent because

$$\begin{aligned} \sum_{i=1}^r c_i A v_i = A \sum_{i=1}^r c_i v_i = 0 &\implies \sum_{i=1}^r c_i v_i \in \mathrm{Null}(A), \text{ but } \sum_{i=1}^r c_i v_i \in R(A) \\ &\implies \sum_{i=1}^r \in \mathrm{Null}(A) \cap R(A) = \{0\} \end{aligned}$$

Since $\dim C(A) = r$, $\{Av_i\}$ must form a basis of $C(A)$. Therefore, A is a bijection between vector spaces and is thus an isomorphism.

Corollary 4.23

$$\text{rk}(A) = \text{rk}(A^T) \tag{167}$$

Theorem 4.24

The product of square lower triangular matrices is a lower triangular matrix. The product of square upper triangular matrices is an upper triangular matrix.

4.4 Decomposition of Matrices

4.4.1 LUP Decomposition

Theorem 4.25 (LU Decompositions)

If a $m \times n$ matrix A can be reduced to row echelon form using only elementary row operations E^1, it can be decomposed into the product of a lower triangular $m \times m$ matrix L with diagonal entries equal to 1 and an upper triangular $m \times n$ matrix U.

$$A = LU \tag{168}$$

This is called **LU decomposition**, or **LU factorization**.

Proof. We reduce A to its echelon form ref(A) by successively multiplying elementary matrices E^{γ_i} representing elementary operation (i). After a finite amount of steps r, we will reduce it to ref(A).

$$\text{ref}(A) = E^{\gamma_r} E^{\gamma_{r-1}} \cdots E^{\gamma_2} E^{\gamma_1} A = \bigg(\prod_{i=0}^{r-1} E^{\gamma_{r-i}} \bigg) A \tag{169}$$

Since each E^{γ_i} is invertible, we multiply the product of the inverses of the elementary matrices of operation (i), which are also elementary matrices of operation (i).

$$(E^{\gamma_1})^{-1} (E^{\gamma_2})^{-1} \cdots (E^{\gamma_r})^{-1} ref(A) = \bigg(\prod_{j=1}^{r} (E^{\gamma_j})^{-1} \bigg) \bigg(\prod_{i=1}^{r-1} E^{\gamma_{r-i}} \bigg) A = A \tag{170}$$

Since each $(E^{\gamma_j})^{-1}$ is an elementary row operation, it is lower diagonal, and by theorem 3.16, their product is also lower triangular. It is easy to prove that if the diagonal entries are furthermore equal to 1, then the product has diagonal entries equal to 1. Finally, it is clear that every matrix in row echelon form is upper triangular, and we are done.

$$A = \bigg(\prod_{j=1}^{r} (E^{\gamma_j})^{-1} \bigg) \text{ref}(A) = LU \tag{171}$$

Note that the existence of the LU decomposition for a general $m \times n$ matrix is not guaranteed. It will not exist if we must switch rows in matrix A in order to reduce it to its echelon form. It does not matter whether we need to use elementary operation (ii) or not. Only the necessity of elementary operation (iii) to reduce the matrix determines the existence of the LU decomposition. The decomposition is also unique.

Finding the LU decomposition of a matrix is useful for solving systems of linear equations. Given a system in the form of $Ax = b$, if we know the LU decomposition of A, we can rewrite the system as

$$LUx = b \tag{172}$$

Setting $y = Ux$, we can easily solve the system $Ly = b$ using forward substitution and then we can solve the system $Ux = y$ using back substitution. Therefore, knowing this decomposition beforehand greatly aids in computing the solutions to the linear system. But computing L and U in order to solve this system takes as much effort as solving the system using Gauss Elimination in the first place.

It is imperative to mention a similar decomposition for $n \times n$ matrices, known as **LUP decomposition**.

Definition 4.28

An $n \times n$ permutation matrix is a matrix of 0s and 1s with exactly one 1 in each row and column. The set of all $n \times n$ permutation matrices form a multiplicative matrix group of order $n!$. We can also view this group as the matrix representation of the symmetric group S_n.

Example 4.3

The set of all 2×2 permutation matrices is

$$S_2 = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\} \tag{173}$$

and the set of all 3×3 permutation matrices is

$$S_3 = \left\{ I_3, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \right\} \tag{174}$$

Theorem 4.26

Every $n \times n$ matrix A can be decomposed into the form $A = PLU$, where L is lower triangular, U is upper triangular, and P is a permutation matrix.

Proof. We can modify the Gauss Elimination algorithm to do all the row interchanges in the beginning. The permutation matrices form a group, so the product of all the initial row changes is a permutation matrix. Call it P'. The previous theorem states that we can do LU decomposition on $P'A$.

$$P'A = LU \implies A = P'^{-1}LU = PLU \tag{175}$$

Since P'^{-1} is also in the symmetric group of permutations, we can denote it as P.

Corollary 4.27

Every $n \times n$ matrix A can be decomposed into the form LUP. That is, in the form

$$A = LUP = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ * & 1 & 0 & \ldots & 0 \\ * & * & 1 & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ * & * & \cdots & * & 1 \end{pmatrix} \begin{pmatrix} u_{11} & * & * & \ldots & * \\ 0 & u_{22} & * & \ldots & * \\ 0 & 0 & u_{33} & \ldots & \vdots \\ \vdots & \vdots & \vdots & \ddots & * \\ 0 & 0 & \ldots & 0 & u_{nn} \end{pmatrix} \begin{pmatrix} & & \\ & P & \\ & & \end{pmatrix} \tag{176}$$

Proof. We decompose $A^T = P_0 L_0 U_0$, where P_0 is a permutation matrix, L_0 lower triangular, U_0 upper triangular. This implies that

$$A = A^{TT} = U_0^T L_0^T P_0^T = LUP \tag{177}$$

since U_0^T is lower triangular and L_0^T is upper triangular. Note that L is unique, but U is not unique, so this decomposition is not unique.

This decomposition can also be used to solve matrix equations

$$AX = B \tag{178}$$

Since this equation can be expressed in the form

$$A \begin{pmatrix} | & & | \\ x_1 & \cdots & x_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ Ax_1 & \cdots & Ax_n \\ | & & | \end{pmatrix} = \begin{pmatrix} | & & | \\ b_1 & \cdots & b_n \\ | & & | \end{pmatrix} \tag{179}$$

solving this matrix is equivalent to solving the system of systems of linear equations

$$Ax_1 = b_1, Ax_2 = b_2, \cdots, Ax_n = b_n \tag{180}$$

i.e. by solving one column at a time. This method can also be used to solve

$$AX = I \tag{181}$$

to find $X = A^{-1}$. Equivalently, we can left multiply elementary matrices to reduce A to rref(A).

$$E^{\gamma_r} E^{\gamma_{r-1}} \cdots E^{\gamma_2} E^{\gamma_1} AX = \text{rref}(A) X = E^{\gamma_r} E^{\gamma_{r-1}} \cdots E^{\gamma_2} E^{\gamma_1} I = \prod_{i=0}^{r-1} E^{\gamma_{r-i}} \tag{182}$$

If rref(A) $= I$, then

$$A^{-1} = \prod_{i=0}^{r-1} E^{\gamma_{r-i}} \tag{183}$$

and if rref(A) $\neq I$, then A^{-1} does not exist. This is in fact precisely the method of finding the inverse where we do Gauss Elimination on the extended matrix

$$\left(\begin{array}{c|c} A & I \end{array} \right) \longrightarrow \left(\begin{array}{c|c} I & A^{-1} \end{array} \right) \tag{184}$$

4.4.2 QR Decomposition

The QR decomposition is often used to simplify these linear least squares problems into a more manageable equation.

Definition 4.29 (QR Decomposition)

Any $m \times n$ matrix A over $\mathbb{F}$ (where $m \geq n$) may be decomposed into the product

$$A = QR \tag{185}$$

where Q is an $m \times n$ matrix with pairwise orthonormal column vectors (meaning $Q^\dagger Q = I_n$), and R is an $n \times n$ upper triangular matrix. If A has full column rank (rk(A) $= n$), the decomposition is unique provided that we require the diagonal entries of R to be strictly positive.

Proof. We use the Gram-Schmidt orthogonalization process on the column vectors of A. Let $a_1, a_2, \ldots, a_n$ be the columns of A. We can recursively construct the orthonormal columns $q_1, \ldots, q_n$ of Q.

For the first step, let $v_1 = a_1$. Then $q_1 = v_1/\|v_1\|$. We set $r_{11} = \|v_1\|$, giving $a_1 = r_{11} q_1$.

For step j (where $1 < j \leq n$), we subtract the projections of a_j onto the previously computed orthogonal vectors $q_1, \ldots, q_{j-1}$:

$$v_j = a_j - \sum_{i=1}^{j-1} (a_j, q_i) q_i \tag{186}$$

If A has full column rank, then a_j is not in the span of the previous vectors, so $v_j \neq 0$. We define $q_j = v_j/\|v_j\|$.

Let $r_{ij} = (a_j, q_i)$ for $i < j$ and $r_{jj} = \|v_j\|$. This allows us to write a_j as:

$$a_j = \sum_{i=1}^{j} r_{ij} q_i \tag{187}$$

This is precisely the matrix multiplication $A = QR$, where R is upper triangular and Q has orthonormal columns q_i. The condition $r_{jj} = \|v_j\| > 0$ makes the decomposition unique for full-rank matrices.

Because $Q^\dagger Q = I$, the QR decomposition can greatly simplify normal equations in linear least squares problems:

$$A^T A x = A^T b \implies (QR)^T (QR) x = R^T Q^T Q R x = R^T R x = R^T Q^T b \tag{188}$$

$$\implies Rx = Q^T b \tag{189}$$

$$\implies x = R^{-1} Q^T b \tag{190}$$

The classical Gram-Schmidt process is often numerically unstable.

Algorithm 4.1 (Modified Gram-Schmidt for QR Decomposition)

The Modified Gram-Schmidt algorithm provides better numerical stability by orthogonalizing the remaining column vectors against the newly computed orthonormal vector immediately at each step.

for $j = 1, \ldots, n$ **do**
    $v_j = a_j$
end for
for $i = 1, \ldots, n$ **do**
    $r_{ii} = \|v_i\|$
    $q_i = v_i / r_{ii}$
    for $j = i+1, \ldots, n$ **do**
        $r_{ij} = (v_j, q_i)$
        $v_j = v_j - r_{ij} q_i$
    end for
end for

4.4.3 Cholesky Decomposition

Definition 4.30 (Cholesky Decomposition)

Every Hermitian (or real symmetric), positive-definite matrix A can be decomposed into the product

$$A = LL^\dagger \tag{191}$$

where L is a lower triangular matrix with strictly positive real diagonal entries. This decomposition is unique.

Proof. We proceed by induction on the dimension n of the matrix A.
Base case: For $n = 1$, $A = (a_{11})$. Since A is positive-definite, $a_{11} > 0$. We simply choose $L = (\sqrt{a_{11}})$, which is unique and positive.
Inductive step: Assume the theorem holds for $(n-1) \times (n-1)$ matrices. We partition the $n \times n$ matrix A as:

$$A = \begin{pmatrix} a_{11} & w^\dagger \\ w & A_{22} \end{pmatrix} \tag{192}$$

where a_{11} is a positive scalar (since A is positive-definite, its diagonal entries are positive), w is an $(n-1) \times 1$ column vector, and A_{22} is an $(n-1) \times (n-1)$ Hermitian matrix.
We seek a lower triangular matrix L of the form:

$$L = \begin{pmatrix} l_{11} & 0 \\ l_{21} & L_{22} \end{pmatrix} \tag{193}$$

such that $A = LL^\dagger$. Multiplying out $LL^\dagger$, we get:

$$\begin{pmatrix} l_{11}^2 & l_{11} l_{21}^\dagger \\ l_{11} l_{21} & l_{21} l_{21}^\dagger + L_{22} L_{22}^\dagger \end{pmatrix} = \begin{pmatrix} a_{11} & w^\dagger \\ w & A_{22} \end{pmatrix} \tag{194}$$

Matching the blocks, we uniquely determine $l_{11} = \sqrt{a_{11}}$ (since $l_{11} > 0$). Then we can uniquely determine the column vector $l_{21} = w / l_{11}$.
Finally, we must satisfy $L_{22} L_{22}^\dagger = A_{22} - l_{21} l_{21}^\dagger$. The matrix $S = A_{22} - l_{21} l_{21}^\dagger$ is the Schur complement of a_{11} in A. Because A is positive-definite, its Schur complement S is also positive-definite. By our inductive hypothesis, the $(n-1) \times (n-1)$ positive-definite matrix S has a unique Cholesky decomposition $S = L_{22} L_{22}^\dagger$. Thus, L_{22} exists and is unique, completing the proof.

Algorithm 4.2 (Cholesky-Banachiewicz Algorithm)

Intuitively, this algorithm computes the lower triangular matrix L starting from the upper left corner and proceeding row by row. It requires about half the operations of the LU decomposition.

for $i = 1, \ldots, n$ **do**
  for $j = 1, \ldots, i$ **do**
    if $j = i$ **then**
      $l_{jj} = \sqrt{a_{jj} - \sum_{k=1}^{j-1} |l_{jk}|^2}$
    else
      $l_{ij} = \frac{1}{l_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik} \bar{l}_{jk} \right)$
    end if
  end for
end for

4.5 Matrices in Block Form

Theorem 4.28

Given 2×2 block matrices

$$X = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix}, \quad Y = \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} \tag{195}$$

We can compute XY similarly to regular matrix multiplication, treating the blocks as entries.

$$XY = \begin{pmatrix} A_1 & B_1 \\ C_1 & D_1 \end{pmatrix} \begin{pmatrix} A_2 & B_2 \\ C_2 & D_2 \end{pmatrix} = \begin{pmatrix} A_1A_2 + B_1C_2 & A_1B_2 + B_1D_2 \\ C_1A_2 + D_1C_2 & C_1B_2 + D_1D_2 \end{pmatrix} \tag{196}$$

Furthermore, this process can be done in general for any $m \times n$ block matrix X and $n \times p$ block matrix Y.

Theorem 4.29

Given that I_N, A, B are $n \times n$ matrices, define the $(2n) \times (2n)$ matrix

$$X = \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} \tag{197}$$

Then

$$\det X = \det B \tag{198}$$

Proof. We can perform Gauss elimination to reduce X without affecting the determinant.

$$\det \begin{pmatrix} I & 0 \\ A & B \end{pmatrix} = \det \begin{pmatrix} I & 0 \\ 0 & B \end{pmatrix} = \det B \tag{199}$$

since it satisfies the correct properties for $\det B$.

Corollary 4.30

$$\det \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} = \det A \det D \tag{200}$$

Proof.

$$\det \begin{pmatrix} A & 0 \\ C & D \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ C & I \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & D \end{pmatrix} = \det \begin{pmatrix} A & 0 \\ C & I \end{pmatrix} \det \begin{pmatrix} I & 0 \\ 0 & D \end{pmatrix} \tag{201}$$

However,

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} \neq \det A \det D - \det B \det C \tag{202}$$

Rather, we introduce the following theorem

Theorem 4.31

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A) \det(D - CA^{-1}B) \tag{203}$$

$$= \det(D) \det(A - BD^{-1}C) \tag{204}$$

Proof.

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A & 0 \\ C & I \end{pmatrix} \begin{pmatrix} I & A^{-1}B \\ 0 & D - CA^{-1}B \end{pmatrix} \tag{205}$$

by similarity, equation (6) is equal to equation (7).

Definition 4.31 (Block Diagonal Matrix)

A **block diagonal matrix** is a square matrix in block form such that the diagonal blocks are square matrices and all off-diagonal blocks are zero matrices.

$$A = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k \end{pmatrix} \tag{206}$$

Theorem 4.32

Given a matrix A in block diagonal form, with diagonal blocks $A_1, A_2, \cdots, A_k$,

$$\det A = \prod_{i=1}^k A_i, \quad \operatorname{Tr} A = \sum_{i=1}^k \operatorname{Tr} A_i \tag{207}$$

Furthermore, A is invertible if and only if all the A_i's are invertible, and

$$A^{-1} = \begin{pmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k \end{pmatrix} = \begin{pmatrix} A_1^{-1} & 0 & \dots & 0 \\ 0 & A_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_k^{-1} \end{pmatrix} \tag{208}$$

Proof. The results are obvious when performing block multiplication or Gauss Elimination.

4.6 Duality Theorem

The duality theorem allows us to define linear programming and establish the equivalence of primal/dual optimization.

In this section we will denote vector inequalities as entry-wise inequalities.Recall that elements of a vector space X can be interpreted as column vectors, and elements of the dual of the vector space X^* can be interpreted as row vectors. Therefore, value of ϕ at x is denoted

$$\phi(x) = \phi_1 x_1 + \phi_2 x_2 + \cdots + \phi_n x_n \tag{209}$$

Furthermore, the dual of X^* is X itself, and given that Y is a linear subspace of X, the annihilator of $Y^\perp$ is Y.

$$X = X^{**}, \qquad Y = Y^{\perp\perp} \tag{210}$$

Suppose Y is defined as the linear space spanned by the m vectors $y_1, y_2, \cdots, y_m$ in X. That is, Y consists of all vectors y of the form

$$y = \sum_{j=1}^m a_j y_j \tag{211}$$

It is clear by linearity that ϕ belongs to $Y^\perp$ if and only if

$$\phi(y_j) = 0, \qquad j = 1, 2, \cdots, m \tag{212}$$

That is, a vector y can be written as a linear combination of m given vectors y_j if and only if every ϕ that annihilates the m vectors a also annihilates 0. Now, we state a theorem that allows us to check if a vector y can be written as a *nonnegative* linear combinations of the y_js.

Theorem 4.33

A vector y can be written as a linear combination of given vectors y_j with nonnegative coefficients if and only if every $\zeta \in X^*$ that satisfies

$$\zeta(y_j) \geq 0, \;\; j = 1, 2, \cdots, m \tag{213}$$

also satisfies

$$\zeta(y) \geq 0 \tag{214}$$

Proof. The proof is not the easiest to construct rigorously, but it can be visualized easily.

Corollary 4.34

Given a $n \times m$ matrix Y, a vector y with n components can be written in the form

$$y = Yp, \;\; p \geq 0 \tag{215}$$

if and only if every row vector ζ that satisfies

$$\zeta Y \geq 0 \tag{216}$$

also satisfies

$$\zeta y \geq 0 \tag{217}$$

Theorem 4.35

Given an $n \times m$ matrix Y and a column vector y with n components, the inequality

$$y \geq Yp, \;\; p \geq 0 \tag{218}$$

is satisfied if and only if every ζ that satisfies

$$\zeta Y \geq 0, \;\; \zeta \geq 0 \tag{219}$$

also satisfies

$$\zeta y \geq 0 \tag{220}$$

Theorem 4.36 (Duality Theorem)

Let Y be a given $n \times m$ matrix, y a given column vector with n components, and γ a given row vector with m components. Let

$$S = \sup_p \{\gamma p\} \tag{221}$$

for all column vectors p with m components satisfying $y \geq Yp, \; p \geq 0$. A well-defined such p is called

supremum admissible. Additionally, let

$$s = \inf_{\zeta}\{\zeta y\} \tag{222}$$

for all row vectors ζ with n components satisfying $\gamma \leq \zeta Y,\ \zeta \geq 0$. A well-defined such ζ is called **infimum admissible**. Given that admissible vectors p and ζ exist, then S and s are finite and

$$S = s \tag{223}$$

5 Tensors

There are multiple ways to construct tensor product spaces, but intuitively, the best explanation of it was from Aspinwall: a glorified comma. Note that all the constructions are equivalent and will lead to the exact same properties of tensors. The first method defines tensors outright as multilinear maps, without the need for a basis.

5.1 Tensor Product of Two Spaces

Definition 5.1

The tensor product of two vector spaces V and W is a vector space, denoted $V \otimes W$, created by the bilinear map

$$\otimes : V \times W \longrightarrow V \otimes W,\ (x, y) \mapsto x \otimes y$$

That is,

$$V \otimes W \equiv \{x \otimes y \mid x \in V, y \in W\}$$

where the elements of $V \otimes W$ are called **tensors**. Note that since we have defined the operation $\otimes$ to be bilinear, it satisfies the properties

1. $(u_1 + u_2) \otimes v = u_1 \otimes v + u_2 \otimes v$
2. $v \otimes (u_1 + u_2) = v \otimes u_1 + v \otimes u_2$
3. $(\lambda u) \otimes v = u \otimes (\lambda v) = \lambda(u \otimes v)$

Moreover, each tensor $x \otimes y$ is itself a bilinear operator

$$x \otimes y : V^* \otimes W^* \longrightarrow \mathbb{F}$$

Using these properties we will deduce further qualities of tensor product spaces. First, given a basis $\{e_i\}$ of n-dimensional space V and $\{f_j\}$ of m-dimensional space W, we can construct a basis

$$\{e_i \otimes f_j \mid 1 \leq i \leq n, 1 \leq j \leq m\}$$

of $V \otimes W$ using only the bilinearity properties of $\otimes$.

Example 5.1

Let V^* be a 4-dimensional vector space with basis $\{e^0, e^1, e^2, e^3\}$. Then the basis of $V^* \otimes V^*$ is

$$\begin{gathered}\{e^0 \otimes e^0,\ e^0 \otimes e^1,\ e^0 \otimes e^1,\ e^0 \otimes e^1, \\ e^1 \otimes e^0,\ e^1 \otimes e^1,\ e^1 \otimes e^2,\ e^1 \otimes e^3, \\ e^2 \otimes e^0,\ e^2 \otimes e^1,\ e^2 \otimes e^2,\ e^2 \otimes e^3, \\ e^3 \otimes e^0,\ e^3 \otimes e^1,\ e^3 \otimes e^2,\ e^3 \otimes e^3\}\end{gathered}$$

That is, every tensor can be expressed as a linear combination of these vectors, which implies

$$\dim V \otimes W = (\dim V)(\dim W)$$

By equality of dimensionality and bilinearity, it is obvious that

$$V \otimes W \simeq \text{Hom}(V \times W, \mathbb{F})$$

In fact, they are naturally isomorphic.

Notice that we still haven't actually defined how to "calculate" using the operator $x \otimes y$. It turns out that defining a tensor product is unique up to isomorphism. That is, if $(V \otimes W, \otimes_1)$ and $(V \otimes W, \otimes_2)$ are two tensor product spaces sufficing bilinearity, then $V \otimes_1 W \simeq V \otimes_2 W$. This result is formally stated in the theorem below, which establishes a one-to-one correspondence of linear maps on the tensor product space with bilinear maps on the Cartesian product space.

Theorem 5.1 (Universal Property of Tensor Products)

Let $S : V \times W \to V \otimes W$ be the canonical map $S(v, w) = v \otimes w$. For every **bilinear** map $\varphi : V \times W \longrightarrow \mathbb{F}$, there exists a unique **linear** map $\psi : V \otimes W \longrightarrow \mathbb{F}$ such that

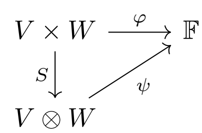
$$\varphi(x, y) = \psi(x \otimes y) \; \forall x \in V, y \in W$$

Proof. Since $\{e_i \otimes f_j\}$ is a basis for $V \otimes W$, any linear map ψ is fully determined by its action on these basis vectors. We define ψ by:

$$\psi(e_i \otimes f_j) = \varphi(e_i, f_j)$$

Extending this linearly to the whole space $V \otimes W$ gives the unique map required.

This property can be visualized by the following commutative diagram. The map S is not an isomorphism, but it is the universal bilinear map.[3]



That is, $\psi \circ S = \varphi$. This establishes a canonical isomorphism between the space of bilinear forms on $V \times W$ and the dual space of the tensor product:

$$\text{Bilinear}(V \times W, \mathbb{F}) \cong (V \otimes W)^*$$

However, a more common perspective in linear algebra is to view tensors as linear maps between vector spaces. By fixing one component of the bilinear map, we obtain the canonical isomorphism:

$$V^* \otimes W \simeq \text{Hom}(V, W) \tag{224}$$

That is, an element $T = \sum \alpha_i \otimes w_i \in V^* \otimes W$ can be interpreted as a linear map $T : V \to W$. Explicitly, for any $v \in V$:

$$T(v) = \sum_i \alpha_i(v) w_i$$

We will focus on elements of $V^* \otimes W$. Given bases $\{e_i\}$ for V and $\{f_j\}$ for W, let $\{\epsilon_i\}$ be the dual basis for V^*. A tensor $T \in V^* \otimes W$ can be written as:

$$T = \sum_{i,j} A_{ji} \, \epsilon_j \otimes f_i$$

The coefficients A_{ji} correspond exactly to the matrix entries of the linear map. This realization of a tensor product between a covector and a vector is often called the **outer product**.

Definition 5.2

Given vector spaces U, V with coordinate vectors $u \in \mathbb{F}^m$ and $v \in \mathbb{F}^n$, the **outer product** is defined

[3] Thanks to Thomas Kidane for pointing out a previous error.

as the matrix multiplication uv^T:

$$u \otimes_{\text{outer}} v \equiv uv^T = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 & \ldots & u_1 v_n \\ \vdots & \ddots & \vdots \\ u_m v_1 & \ldots & u_m v_n \end{pmatrix}$$

Definition 5.3

Given vector spaces U, V with defined bases in each of them, the **outer product** of two vectors $u \in U$ and $v \in V$ is defined

$$u \otimes v \equiv uv^T \equiv \begin{pmatrix} u_1 \\ u_2 \\ \cdots \\ u_m \end{pmatrix} \otimes \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 & \cdots & u_1 v_n \\ u_2 v_1 & \cdots & u_2 v_n \\ \cdots & \cdots & \cdots \\ u_m v_1 & \cdots & u_m v_n \end{pmatrix}$$

Note that the $\otimes$ symbol in here represents the outer product, not the tensor product. Note that the tensor rank of the outer product of two vectors is $(2, 0)$.

Abstractly speaking, the outer product of $u \in U$ and $v \in V$ is the element $u \otimes v \in U \otimes V$, which is a rank-(2,0) tensor, not a rank-(1,1) tensor! Just because it "looks" like a matrix, $u \otimes v$ should not be interpreted as a linear map. Such a $m \times n$ matrix could really be the realization of either a (2,0) tensor, (1,1) tensor, or a (0,2) tensor.

However, if U is an inner product space, then it is possible to define $u \times v$ as a linear map from $U \longrightarrow W$. The structure of the inner product on U allows us to define the canonical isomorphism ϕ between U and U^*. Then, we can define the canonical injections $i : U \longrightarrow U \otimes V$ and $j : U^* \longrightarrow U^* \otimes V$ to get the commutative diagram

$$\begin{array}{ccc} U \otimes V & \xrightarrow{\gamma} & U^* \otimes V \\ {\scriptstyle i}\uparrow & & {\scriptstyle j}\uparrow \\ U & \xrightarrow{\phi} & U^* \end{array}$$

Given that

$$\phi(u) \equiv l \text{ such that } (u, x) = l(x) \forall x \in U$$

we can define the mapping $\gamma : j\phi i^{-1}$ such that

$$\gamma(u \otimes v) \equiv \phi(u) \otimes v \equiv l \otimes v \in U^* \otimes V$$

which is ultimately a linear mapping from $U \longrightarrow V$ since

$$l \otimes v(u_0, \cdot) \equiv l(u_0) v(\cdot)$$

with $l(u_0) \in \mathbb{F}$ and $v(\cdot)$ a vector. This proves the following theorem.

Theorem 5.2

The matrix rank of the outer product of any 2 vectors is 1.

Proof. Trivial.

We can extrapolate and see that for higher order tensor products, we would get an n-dimensional array of scalars. A matrix is a 2-dimensional array of numbers since it is the tensor product of two vectors.

Definition 5.4

Given vector spaces U, V with defined bases in each of them, the **Kronecker product** of two vectors $u \in U$ and $v \in V$ is defined

$$u \otimes_{Kron} v \equiv \begin{pmatrix} u_1 \\ u_2 \\ \cdots \\ u_m \end{pmatrix} \otimes \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 \\ u_1 v_2 \\ \cdots \\ u_m v_{n-1} \\ u_m v_n \end{pmatrix}$$

Clearly, the outer product and Kronecker product are closely related, and we can interpret the Kronecker product as a form of "vectorization" or "flattening out" of the outer product.

5.2 Higher Order Tensor Product Spaces

Since $U \otimes W$ is a vector space itself, we can multiply it further to create higher order tensor product spaces.

$$U \otimes W \otimes U \otimes \cdots$$

Note that by construction, the operation of tensor product on vector spaces is commutative and associative in the sense that

$$V \otimes W \simeq W \otimes V$$

and

$$(U \otimes V) \otimes W \simeq U \otimes (V \otimes W) \simeq U \otimes V \otimes W$$

which allows us to write tensor products of any finite number of vector spaces $V_1, V_2, \cdots, V_n$ without parantheses. By induction, we can keep constructing higher order tensor products as such

$$V_1 \otimes V_2 \to (V_1 \otimes V_2) \otimes V_3 \to \big((V_1 \otimes V_2) \otimes V_3\big) \otimes V_4 \to \cdots$$

to get the tensor product space

$$\bigotimes_{i=1}^n V_i \equiv V_1 \otimes V_2 \otimes \cdots \otimes V_n$$

with tensors in the form

$$\bigotimes_{i=1}^n v_i \equiv v_1 \otimes v_2 \otimes v_3 \otimes \cdots \otimes v_n; \; v_i \in V_i$$

defined as the following multilinear map

$$\bigotimes_{i=1}^n v_i : \prod_{i=1}^n V_i^* \longrightarrow \mathbb{F}, \quad \bigg(\bigotimes_{i=1}^n v_i\bigg)\big(l_1, l_2, \cdots, l_n\big) \equiv \prod_{i=1}^n v_i(l_i), \; l_i \in V_i^*$$

This map can then be used to easily see the following statement

$$\bigotimes_{i=1}^n V_i \simeq \text{Hom}\Big(\prod_{i=1}^n V_i^*, \mathbb{F}\Big)$$

Similarly to the section about the tensor product of two spaces, we can "partially" fill in the inputs of a general tensor $v_1 \otimes v_2 \otimes \cdots \otimes v_n$ to interpret them as multilinear operators that can take in k vectors and output $n-k$ vectors. That is, tensors (written as τ below) are multilinar maps from a cartesian product of vector spaces to a tensor product of vector spaces.

$$\tau : V_1 \times \cdots \times V_n \longrightarrow W_1 \otimes \cdots \otimes W_m$$

For example,

$$\text{Hom}\Big(\prod_{i=1}^n V_i^*, \mathbb{F}\Big) \simeq \text{Hom}\Big(\prod_{i=2}^n V_i^*, V_1\Big) \simeq \text{Hom}\Big(\prod_{i=3}^n V_i^*, V_1 \otimes V_2\Big) \simeq \cdots$$

Furthermore, we can generalize the universal property of two tensors to the following theorem, which is also called the **fundamental principle of tensor algebra**.

Theorem 5.3 (Universal Property of n-tensors)

Given a **multilinear** mapping $\varphi : V_1 \times \cdots \times V_n \longrightarrow \mathbb{F}$, there exists a unique **linear** map $\psi : V_1 \otimes \cdots \otimes V_n \longrightarrow \mathbb{F}$ such that

$$\varphi(v_1, \ldots, v_n) = \psi(v_1 \otimes \cdots \otimes v_n)$$

This correspondence establishes the canonical isomorphism:

$$\left(\bigotimes_{i=1}^n V_i\right)^* \simeq \mathcal{L}(V_1, \ldots, V_n; \mathbb{F})$$

where $\mathcal{L}(V_1, \ldots, V_n; \mathbb{F})$ denotes the vector space of all n-linear forms on the Cartesian product $V_1 \times \cdots \times V_n$.

Definition 5.5

Given that

$$\{e_{i_j}\}_{i_j=1}^{k_j} \text{ of } V_j,\ j = 1, 2, \cdots, n$$

are n sets of bases for each V_j,

$$\{\bigotimes_{j=1}^n e_{i_j}\}_{i_1, \cdots, i_n} \text{ is a basis of } \bigotimes_{j=1}^n V_j$$

Theorem 5.4

Given vector spaces $V_1, V_2, \cdots, V_n$,

$$\dim \bigotimes_{i=1}^n V_i = \prod_{i=1}^n \dim V_i$$

Proof. This follows naturally from the construction of the basis.

We move on to talk about something quite enlightening: the tensor product of linear operators, which are themselves tensors.

Definition 5.6

Given linear operators $A \in \text{End}(V)$, $B \in \text{End}(W)$, we can construct the linear operator

$$A \otimes B \in \text{End}(V \otimes W)$$

such that

$$(A \otimes B)(x \otimes y) \equiv Ax \otimes By \in V \otimes W$$

Notice that since A, B are linear operators, they are tensors. More specifically, $A \equiv \alpha \otimes u$ and $B \equiv \beta \otimes v$, so

$$\begin{aligned}
(A \otimes B)(x \otimes y) &\equiv Ax \otimes By \\
&= (\alpha \otimes u)x \otimes (\beta \otimes v)y \\
&= \alpha(x)\beta(y)\, u \otimes v \\
&= \big((\alpha \otimes \beta)(x \otimes y)\big)(u \otimes v)(\cdot, \cdot) \\
&= \big((\alpha \otimes \beta) \otimes (u \otimes v)\big)\big((x \otimes y), (\cdot \otimes \cdot)\big) \\
&= \big((\alpha \otimes \beta) \otimes (u \otimes v)\big)(x \otimes y)
\end{aligned}$$

$\implies A \otimes B \equiv \alpha \otimes \beta \otimes u \otimes v.$

We will work through an example that gives the matrix representation of the tensor product of linear mappings. For simplicity, let us work with the example when

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \; B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

Given that U has basis $\{u_1, u_2\}$ and V has basis $\{v_1, v_2\}$, $U \otimes V$ will have basis

$$\{u_1 \otimes v_1, u_1 \otimes v_2, u_2 \otimes v_1, u_2 \otimes v_2\}$$

We then define the induced linear mapping $A \otimes B : U \otimes V \longrightarrow U \otimes V$ by defining it on its basis vectors. Note that the linear mapping $(A \otimes B)(u \otimes v)$ must be an element of $U \otimes V$, implying that it is defined

$$(A \otimes B)(u \otimes v) \equiv Au \otimes Bv$$

This is called the **tensor product** of operators A and B. So, the tensor product of matrices A and B can be calculated

$$\begin{aligned}
(A \otimes B)(u_1 \otimes v_1) &= (a_{11}u_1 + a_{21}u_2) \otimes (b_{11}v_1 + b_{21}v_2) \\
&= a_{11}b_{11}(u_1 \otimes v_1) + a_{11}b_{21}(u_1 \otimes v_2) \\
&\quad + a_{21}b_{11}(u_2 \otimes v_1) + a_{21}b_{21}(u_2 \otimes v_2) \\
\cdots &= \cdots \\
(A \otimes B)(u_2 \otimes v_2) &= (a_{12}u_1 + a_{22}u_2) \otimes (b_{12}v_1 + b_{22}v_2) \\
&= a_{12}b_{12}(u_1 \otimes v_1) + a_{12}b_{22}(u_1 \otimes v_2) \\
&\quad + a_{22}b_{12}(u_2 \otimes v_1) + a_{22}b_{22}(u_2 \otimes v_2)
\end{aligned}$$

In matrix form, this results in the 4×4 matrix (also in block form)

$$A \otimes B \equiv \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{pmatrix}$$

representing the linear transformation from $U \otimes V$ to itself under the basis $\{u_i \otimes v_j\}$.

Theorem 5.5

In general,the tensor product of matrices $A \in \text{End}(V)$ and $B \in \text{End}(W)$ (with basis of V, W defined) is the $(mn) \times (mn)$ matrix

$$A \otimes B \equiv \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1}B & a_{n2}B & \cdots & a_{nn}B \end{pmatrix}$$

represented in block form.

Theorem 5.6

$$\begin{aligned} &\text{Tr}\, A \otimes B = \text{Tr}\, A \cdot \text{Tr}\, B \\ &\det A \otimes B = (\det A)^n (\det B)^m \end{aligned}$$

Theorem 5.7

For finite dimensional space V and W,

$$\text{End}(V \otimes W) = \text{End}(V) \otimes \text{End}(W)$$

5.3 Contractions, Tensor Algebras

Definition 5.7

Given vector space V, a **rank** (k,l)**-tensor product space** of V, denoted $\mathbb{T}^k_l V$, is defined

$$\mathbb{T}^k_l V \equiv \bigg(\bigotimes_{i=1}^k V \bigg) \otimes \bigg(\bigotimes_{j=1}^l V^* \bigg) \equiv V^{\otimes k} \otimes V^{*\otimes l}$$

That is, $\mathbb{T}^k_l$ is the space of all (k,l)-tensors. A **rank** (k,l)**-tensor** is an element of a rank (k,l) tensor product space. Note that all tensors are vectors and all tensor product spaces are vector spaces, too. The order in which we multiply V's and V^*'s matter, but in most cases, and from now, we will work with tensor product spaces strictly in the form

$$V^{\otimes k} \otimes V^{*\otimes l}$$

where the V's are multiplied first and V^*'s last. So, $\mathbb{T}^1_1 \equiv V \otimes V^*$, but $\mathbb{T}^1_1 \not\equiv V^* \otimes V$. We can do this because the tensor product of spaces are commutative in the sense that we can always find an isomorphism

$$V \otimes W \simeq W \simeq V$$

Example 5.2

$\mathbb{F}$ is a rank (0,0)-tensor space. V is a rank (1,0)-tensor space, and V^* is a rank (0,1)-tensor space.

We can now think of the tensor product now as a bilinear operator

$$\otimes : \mathbb{T}^p_q V \times \mathbb{T}^r_s V \longrightarrow \mathbb{T}^{p+r}_{q+s} V$$

such that

$$\Big(\bigotimes_{i=1}^{p} v_i \otimes \bigotimes_{j=1}^{q} w_j\Big) \otimes \Big(\bigotimes_{i=p+1}^{p+r} v_i \otimes \bigotimes_{j=q+1}^{q+s} w_j\Big) = \bigotimes_{i=1}^{p+r} v_i \otimes \bigotimes_{j=1}^{q+s} w_j \in \mathbb{T}^{p+r}_{q+s} V$$

Theorem 5.8

$$\mathbb{T}^2_2 V \simeq \text{End}(V) \otimes \text{End}(V)$$

That is, the tensor multiplication $\mathbb{T}^1_1 \times \mathbb{T}^1_1 \longrightarrow \mathbb{T}^2_2$ is precisely the multiplication of the linear operators.

Proof. Letting $A = u \otimes \alpha, B = v \otimes \beta$ with $u, v \in V$ and $\alpha, \beta \in V^*$, we know that

$$A \otimes B = u \otimes v \otimes \alpha \otimes \beta$$

$\implies \text{End}(V) \otimes \text{End}(V) \simeq V \otimes V \otimes V^* \otimes V^* \simeq \mathbb{T}^2_2 V.$

When working with tensors in general, we use the Einstein Summation Notation to write vectors in shorthand form

$$A^\mu e_\mu \equiv \sum_{i=1}^{n} A^i e_i$$

The indices in this context are not important here (but they are significant in physics). For example, the Einstein notation for rank (2,0) tensors is written

$$T_{\mu\nu} e^\mu \otimes e^\nu \equiv \sum_{\mu,\nu} T_{\mu\nu} e^\mu \otimes e^\nu$$

and for an n vectors,

$$\begin{aligned} T_{\mu_1,\cdots,\mu_n} \bigotimes_{i=1}^{n} e^{\mu_i} &\equiv T_{\mu_1,\cdots,\mu_n} e^{\mu_1} \otimes e^{\mu_2} \otimes \cdots \otimes e^{\mu_n} \\ &\equiv \sum_{\mu_1,\cdots,\mu_n} T_{\mu_1,\cdots,\mu_n} e^{\mu_1} \otimes e^{\mu_2} \otimes \cdots \otimes e^{\mu_n} \\ &\equiv \sum_{\mu_1,\cdots,\mu_n} T_{\mu_1,\cdots,\mu_n} \bigotimes_{i=1} e^{\mu_i} \end{aligned}$$

Since the coefficients of the shorthand tensor notation implies the tensors themselves, we can simply write

$$T_{\mu_1,\cdots,\mu_n} \equiv T_{\mu_1,\cdots,\mu_n} \bigotimes_{i=1}^{n} e^{\mu_i}$$

Clearly, this notation is not restricted to the tensor product of contravariant vectors. For example,

$$T_\mu{}^{\alpha\beta}{}_\nu \, e^\mu \otimes e_\alpha \otimes e_\beta \otimes e^\nu \equiv \sum_{\mu,\alpha,\beta,\nu} T_\mu{}^{\alpha\beta}{}_\nu \, e^\mu \otimes e_\alpha \otimes e_\beta \otimes e^\nu$$

is the form of a general tensor in the tensor space $V^* \otimes V \otimes V \otimes V^*$. Note that the order of the subscripts/superscripts in the coefficients of T matters, but again, we usually work with $\mathbb{T}^p_q$ where vector spaces V's come first and then the dual spaces V^*'s come later.

Example 5.3

Let $e_\mu \otimes e^\nu \otimes e^\lambda \in \mathbb{T}^1_2$. Then

$$\begin{aligned}(e_\mu \otimes e^\nu \otimes e^\lambda)\big(B_\epsilon e^\epsilon, A^\delta e_\delta, C^\sigma e_\sigma\big) &= e_\mu(B_\epsilon e^\epsilon) \cdot e^\nu(A^\delta e_\delta) \cdot e^\lambda(C^\sigma e_\sigma) \\ &= B_\epsilon A^\delta C^\sigma \; \delta^\epsilon_\mu \, \delta^\nu_\delta \, \delta^\lambda_\sigma \\ &= B_\mu A^\nu C^\lambda \in \mathbb{R}\end{aligned}$$

We now define the contraction of a tensor.

Definition 5.8

A **contraction** is a linear map

$$C^m_n : \mathbb{T}^p_q \longrightarrow \mathbb{T}^{p-1}_{q-1}, \; 1 \leq m \leq p, 1 \leq n \leq q$$

defined as follows. Let us define the map

$$\tilde{C}^m_n : \prod_p V \times \prod_q V^* \longrightarrow \mathbb{T}^{p-1}_{q-1} V$$

such that (where the hatted elements are taken out)

$$(x_1, \cdots, x_p, \alpha_1, \cdots, \alpha_q) \mapsto \alpha_n(x_m) \; x_1 \otimes \cdots \hat{x_m} \cdots \otimes x_p \otimes \alpha_1 \otimes \cdots \hat{\alpha_n} \cdots \otimes \alpha_q$$

This is clearly a multilinear map, so by the universal property, there exists a unique linear map $C^m_n : \mathbb{T}^p_q V \longrightarrow \mathbb{T}^{p-1}_{q-1} V$ such that

$$\bigotimes_{i=1}^p x_i \otimes \bigotimes_{j=1}^q \alpha_j \mapsto \alpha_n(x_m) \bigotimes_{i \neq m} x_i \otimes \bigotimes_{j \neq n} \alpha_j$$

This mapping C^m_n is called the *mn*th contraction of a tensor in $\mathbb{T}^p_q V$.

Note that there are multiple mappings from $\mathbb{T}^p_q \longrightarrow \mathbb{T}^{p-1}_{q-1}$, depending on the choice of m, n. This contraction function is also canonical, i.e. we did not have to endow any structures to V to define C^n_m.

We could also contract multiple steps at once with the map $\mathbb{T}^p_q \longrightarrow \mathbb{T}^{p-k}_{q-k}$, but this is really just a composition of single contractions

$$\mathbb{T}^p_q \longrightarrow \mathbb{T}^{p-1}_{q-1} \longrightarrow \mathbb{T}^{p-2}_{q-2} \longrightarrow \cdots \longrightarrow \mathbb{T}^{p-k}_{q-k}$$

Definition 5.9

Given a $(0, 2)$-tensor $F_{\alpha\beta}$, we can find its *symmetric component*

$$F_{\{\alpha\beta\}} = \frac{1}{2}\big(F_{\alpha\beta} + F_{\beta\alpha}\big)$$

and its *anti-symmetric component*

$$F_{[\alpha\beta]} = \frac{1}{2}\big(F_{\alpha\beta} - F_{\beta\alpha}\big)$$

such that

$$F_{\alpha\beta} = F_{\{\alpha\beta\}} + F_{[\alpha\beta]}$$

In shorthand form, to form a contraction, we can just write the indices that are being contracted as the same letter.

Example 5.4

When performing a contraction, it is common to make the indices that are being contracted the same. For example, ${X^{abc}}_d \in V^{\otimes 3} \otimes V^*$ can be contracted, so if we can choose the a and d indices to contract, we get

$$ {X^{abc}}_a \in V \otimes V $$

Theorem 5.9

The contraction of a linear operator $A = u \otimes \alpha$ is its trace. Notice how that the vector u comes first and the covector α comes second, since we're working in $\mathbb{T}^1_1 V$.

Proof. Given that $\{e_i\}$ is the basis for n-dimensional space V and $\{f_i\}$ is the dual basis of V^*.

$$ C^1_1(x \otimes \alpha) = \alpha(u) = \Big(\sum_{i=1}^n \alpha_i f_i\Big)\Big(\sum_{j=1}^n x_j e_j\Big) = \sum_{i,j} \alpha_i x_j \delta_i^j = \sum_{i=1}^n \alpha_i x_i $$

which is clearly the definition of the trace.

In addition to contracting a tensor with itself, we can contract a tensor X with another tensor Y.

Example 5.5

$X^{abc} Y_d \in V^{\otimes 3} \otimes V^*$

Theorem 5.10

The contraction of a linear operator $A = u \otimes \alpha$ and a vector x is precisely Ax, the image of x under the linear operator A.

$$ Ax = (u \otimes \alpha)x = \alpha(x) u \in V $$

Calculating this after defining coordinates aligns with matrix multiplication of form

$$ \begin{pmatrix} — & A_1 & — \\ — & A_2 & — \\ \dots & \dots & \dots \\ — & A_n & — \end{pmatrix} \begin{pmatrix} \dots \\ x \\ \dots \end{pmatrix} = \begin{pmatrix} A_1 \cdot x \\ A_2 \cdot x \\ \dots \\ A_n \cdot x \end{pmatrix} $$

Theorem 5.11

The contraction of the tensor product of linear operators A, B is just the regular composition AB. Note that this contraction contracts the second index of A with the first index of B. That is,

$$ C(A \otimes B) = C\big((u \otimes \alpha) \otimes (v \otimes \beta)\big) = \alpha(v)\, u \otimes \beta \in \mathbb{T}^1_1 $$

Clearly, $\alpha(v)\, u \otimes \beta$ is a really another linear map. We can evaluate ABx by performing the contraction on AB first and then contracting it with x.

$$ ABx = \alpha(v)(u \otimes \beta)(x) = \alpha(v)\beta(x)u $$

Alternatively, we can evaluate ABx equivalently by performing the contraction on Bx first and then A

$$ABx = \beta(x)\, Av = \alpha(v)\beta(x)u$$

Either way, it results in the same vector $\alpha(v)\beta(x)u$. This is expected because tensor products are associative.

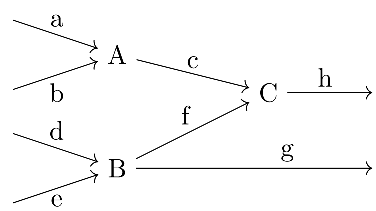
Similarly, we can contract the tensor products of general tensors T and R, which is called a **contraction of** T **with** R. Furthermore, just like linear mappings or vectors, we can factorize arbitrary tensors in their own way. The field of math dealing with this is called *Tensor Network Theory*, which has multiple applications in computer science, chemistry, and physics.

Definition 5.10

We can factorize a complex tensor X into a product of tensors that can be contracted to result in X. We can think of factoring tensors as analogous to anti-contraction. This process is best illustrated with the following example. Let us factor the tensor into three different tensors: a rank (1,2) tensor A, rank (2,2) tensor B, and rank (1,2) tensor C.

$$X_{abde}{}^{hg} = A_{ab}{}^{c} \otimes B_{de}{}^{fg} \otimes C_{cf}{}^{h}$$

We can visually represent factorization with the tensor network diagram



where the "inputs" at each node are covectors and the "outputs" are vectors. Therefore, the entire diagram, which represents the tensor X has a total of 4 inputs (indices a, b, d, e) and two outputs (indices h, g). We can see from the diagram that the indices c and f, which travels "between" the factors are the ones that are being contracted. Therefore, the contraction of c and f contracts the rank (4,6) tensor $A \otimes B \otimes C$ to a rank (2,4) tensor.

Definition 5.11

The **tensor algebra** of vector space V over field $\mathbb{F}$ is an associative, noncommutative algebra defined

$$\begin{aligned} T(V) \equiv \bigoplus_{n=0}^{\infty} V^{\otimes n} &= V^{\otimes 0} \oplus V^{\otimes 1} \oplus V^{\otimes 2} \oplus V^{\otimes 3} \oplus \cdots \\ &= \mathbb{F} \oplus V \oplus V^{\otimes 2} \oplus V^{\otimes 3} \oplus V^{\otimes 4} \oplus \cdots \end{aligned}$$

with elements being infinite-tuples

$$(a, B^{\mu}, C^{\nu\gamma}, D^{\alpha\beta\epsilon}, \cdots)$$

The addition operation is defined component-wise, and the multiplication operation is the tensor product

$$\otimes : T(V) \times T(V) \longrightarrow T(V)$$

and the identity element is

$$I = (1, 0, 0, \cdots)$$

Linearity is easily proved.

The tensor algebra is used to "add" differently ranked tensors together. In order to do this rigorously, we must define the map (which is also an isomorphism)

$$i_j : V^{\otimes j} \longrightarrow T(V),\ i_j(T^{\kappa_1, \cdots, \kappa j}) = (0, \cdots, 0, T^{\kappa_1, \cdots, \kappa j}, 0, \cdots, 0)$$

So, we can implicitly define the addition of arbitrary tensors $A \in V^{\otimes n}$ and $B \in V^{\otimes m}$ as

$$A + B \equiv i_n(A) + i_m(B) \in T(V)$$

along with the tensor multiplication of the form

$$A \otimes B \equiv i_n(A) \otimes i_m(B) \equiv i_{n+m}(A \otimes B)$$

This allows us to alternatively define the tensor product operation as

$$i_i(V^{\otimes i}) \otimes i_j(V^{\otimes j}) \equiv i_{i+j}(V^{\otimes (i+j)})$$

5.4 Exterior Algebras and Symmetric Algebras

We can define the symmetric and exterior algebras multiple ways. In here, we will construct their powers separately as quotient spaces and direct sum them to create their respective algebras. But first, we must introduce the Schmidt decomposition, which is the foundation of all the results of this section.

Theorem 5.12 (Schmidt Decomposition)

For any $w \in U \otimes V$, where U, V $(\dim U = n, \dim V = m)$ are inner product spaces over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, there exists an orthonormal basis $\{u_i\}$ of U and $\{v_j\}$ of V such that

$$w = \sum_{i=1}^{\min\{n,m\}} \alpha_i u_i \otimes v_i,\ \alpha_i \in \mathbb{F}$$

Proof. Since $U \otimes V \simeq \text{Hom}(V^*, U)$, we can interpret w as a matrix $\tilde{w}$. Using singular value decomposition, there exists unitary matrices A, B and diagonal matrix Σ such that

$$\tilde{w} = A \Sigma B^\dagger$$

$C(A)$ and $R(B^\dagger)$ determine the orthonormal basis of $U \otimes V$, and we can thus see that the minimum number of required $u \otimes v$'s is precisely the number of nonzero singular values, which is the rank of $\tilde{w}$.

Definition 5.12

Let I be a subspace of $V \otimes V$ generated by elements of the form $x \otimes x \in V \otimes V$. That is, given a basis $\{e_i\}$ of n-dimensional space V, all tensors of the form $x \otimes x \in V \otimes V$ can be written

$$x \otimes x = \sum_{i=1}^n a_i (e_i \otimes e_i) + \sum_{i \neq j} b_{ij} (e_i \otimes e_j + e_j \otimes e_i)$$

which implies that the components of $e_i \otimes e_j$ and $e_j \otimes e_i$ must be the same for every element in I.

Example 5.6

Given that V is 2-dimensional, a vector $x \in V$ can be written $x = ae_1 + be_2$, which implies

$$\begin{aligned} x \otimes x &= (ae_1 + be_2) \otimes (ae_1 + be_2) \\ &= a^2(e_1 \otimes e_1) + ab(e_1 \otimes e_2) + ba(e_2 \otimes e_1) + b^2(e_2 \otimes e_2) \\ &= a^2(e_1 \otimes e_1) + b^2(e_2 \otimes e_2) + ab(e_1 \otimes e_2 + e_2 \otimes e_1) \end{aligned}$$

Since we can group the components $e_i \otimes e_j$ and $e_j \otimes e_i$ together to $e_i \otimes e_j + e_j \otimes e_i$, the basis of I is

$$\{e_1 \otimes e_1, \cdots, e_n \otimes e_n, e_1 \otimes e_2 + e_2 \otimes e_1, \cdots, e_{n-1} \otimes e_n + e_n \otimes e_{n-1}\}$$

Definition 5.13

Now, we can define the **second exterior power of** V as

$$\Lambda^2 V \equiv \frac{V \otimes V}{I}$$

and it follows that

$$\dim \Lambda^2 V = n^2 - \dim I = \frac{1}{2} n(n-1)$$

We denote the elements of $\Lambda^2 V$ as $x \wedge y$, which really just represents the equivalence class of $x \otimes y$ in the quotient space. It is clear that $x \otimes x \in I \implies x \wedge x = 0$, so

$$\begin{aligned} 0 = (x+y) \wedge (x+y) &= x \wedge x + x \wedge y + y \wedge x + y \wedge y \\ &= x \wedge y + y \wedge x \\ &\implies x \wedge y = -y \wedge x \end{aligned}$$

That is, the wedge product is antisymmetric. Note also that we can assume distributivity of $\wedge$ since it is just the quotient operation of another operation $\otimes$ that satisfies distributivity. We can construct a basis on $\Lambda^2 V$, given by

$$\{e_i \wedge e_j \mid i < j\}$$

Again, we note that $i < j$ since $e_i \wedge e_i = 0$ and $e_i \wedge e_j = -e_j \wedge e_i$.

One realization of the space $\Lambda^2 \mathbb{R}^n$ is the set of antisymmetric $n \times n$ matrices. We can construct higher order exterior powers, too. For $n = 3$ (and assuming that $\dim V \geq 3$), the subspace $I \subset V \otimes V \otimes V$ is the space generated by elements of the forms

$$x \otimes x \otimes y, x \otimes y \otimes x, y \otimes x \otimes x$$

Following a similar construction, the **third exterior power of** V is

$$\Lambda^3 V \equiv \frac{V \otimes V \otimes V}{I}$$

with its elements being equivalence classes of the form

$$x \wedge y \wedge z, \; x, y, z \in V$$

such that

$$\begin{aligned} x \wedge y \wedge z &= -x \wedge z \wedge y \\ &= -y \wedge x \wedge z \\ &= -z \wedge y \wedge x \end{aligned}$$

The basis of $\Lambda^3 V$ is

$$\{e_i \wedge e_j \wedge e_k \mid i < j < k\} \implies \dim \Lambda^3 V = \frac{1}{6} n(n-1)(n-2)$$

Generally, if σ is a permutation of the ordered list $(1, 2, \cdots, n)$, and $x_1, x_2, \cdots, x_n \in V$, then

$$x_{\sigma(1)} \wedge x_{\sigma(2)} \wedge \cdots \wedge_{\sigma(n)} = sgn(\sigma)\, x_1 \wedge x_2 \wedge \cdots \wedge x_n$$

which means that if $x_i = x_j$ for some $1 \leq i \neq j \leq n$,

$$x_1 \wedge x_2 \wedge \cdots \wedge x_n = 0$$

By constructing all the exterior powers of n-dimensional space V, we can construct the algebra

$$\Lambda(V) \equiv \bigoplus_{k=0}^{n} \Lambda^k V \equiv \Lambda^0 V \oplus \Lambda^1 V \oplus \Lambda^2 V \oplus \cdots \oplus \Lambda^n V$$

Note that $\Lambda^0 V = \mathbb{F}$ and $\Lambda^1 V = V$. Unlike the tensor algebra, the exterior algebra is finite since the exterior powers vanish for finite n. In fact,

$$\dim \Lambda^k V = \begin{cases} {}_nC_k & 0 \leq k \leq n \\ 0 & n < k \end{cases}$$

which implies that

$$\dim \Lambda(V) = 2^n$$

Definition 5.14

The nth exterior power $\Lambda^n V$ is 1 dimensional, spanned by the singular basis vector

$$e_1 \wedge e_2 \wedge \cdots \wedge e_{n-1} \wedge e_n$$

This vector is the *determinant*. Note that this construction of the determinant is consistent with our previous construction of the determinant of a matrix since $e_1 \wedge \cdots \wedge e_n$ is indeed multilinear and antisymmetric. In its purest sense,

$$e_1 \wedge \cdots \wedge e_n : \prod_{i=1}^{n} V^* \longrightarrow \mathbb{F}$$

is a mapping that is multilinear and antisymmetric. But there is an inconsistency. The matrix determinant takes in *matrices* rather than taking in n-tuples of covectors. However, we can interpret the n covectors in $V^* \times \cdots \times V^*$ as the column (or row) vectors of an $n \times n$ matrix. This completes the realization, and so we can conclude that the matrix determinant is just a realization of the more abstract determinant $e_1 \wedge \cdots \wedge e_n$.
Note that any tensor in $\Lambda^n V$ satisfies multilinearity and antisymmetricity, but only the basis vector $e_1 \wedge \cdots \wedge e_n$ satisfies the normalizing condition

$$\det I = 1$$

Since, given that the dual basis of V^* is $\{f_j\}$

$$(e_1 \wedge \cdots \wedge e_n)(f_1, f_2, \cdots, f_n) = \prod_{i=1}^{n} e_i(f_i) = \prod_{i=1}^{n} \delta_i^i = 1$$

Example 5.7

Given 3 dimensional vector space V with basis $\{e_1, e_2, e_3\}$, the wedge product of two vectors $a, b \in V$ is

$$\begin{aligned} a \wedge b &= (a_1 e_1 + a_2 e_2 + a_3 e_3) \wedge (b_1 e_1 + b_2 e_2 + b_3 e_3) \\ &= (a_2 b_3 - a_3 b_2) e_2 \wedge e_3 + (a_3 b_1 - a_1 b_3) e_3 \wedge e_1 + (a_1 b_2 - a_2 b_1) e_1 \wedge e_2 \end{aligned}$$

which is essentially the formula for the cross product $\times$ in Euclidean space. We can therefore think of the realization of the wedge product in 3 dimensional space V as the cross product.

$$\wedge : V \times V \longrightarrow \Lambda^2 V$$

Note that $\Lambda^2 V \simeq V$ if $\dim V = 3$, so we can construct the more familiar $\times$ operation in $\mathbb{R}^3$.

$$\times : \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \Lambda^2 \mathbb{R}^3 \simeq \mathbb{R}^3$$

which is consistent with $\times$ taking two vectors and outputting a third vector living in $\mathbb{R}^3$ that is orthogonal to the two input vectors.

Example 5.8

The realization of the wedge product of 3 vectors in 3 dimensional space V is the *triple scalar product*, which we will denote as $\times_3$

$$\wedge : V \times V \times V \longrightarrow \Lambda^3 V$$

Note that since $\Lambda^3 V \simeq V$ when $\dim V = 3$, we can write

$$\times_3 : \mathbb{R}^3 \times \mathbb{R}^3 \times \mathbb{R}^3 \longrightarrow \Lambda^3 \mathbb{R}^3 \simeq \mathbb{R}$$

which is consistent with $\times_3$ taking three vectors and outputting the signed volume of their parallelopiped which lies in $\mathbb{R}$.

Now we introduce the symmetric algebra and its construction. Let I be the subspace of $V \otimes V$ generated by all tensors of the form

$$u \otimes v - v \otimes u, \; u, v \in V$$

For example, given $a, b \in V$ with basis $\{e_1, e_2\}$,

$$\begin{aligned} a \otimes b - b \otimes a &= (a_1 e_1 + a_2 e_2) \otimes (b_1 e_1 + b_2 e_2) - (b_1 e_1 + b_2 e_2) \otimes (a_1 e_1 + a_2 e_2) \\ &= (a_1 b_2 - b_2 a_1) e_1 \otimes e_2 + (a_2 b_1 - b_2 a_1) e_2 \otimes e_1 \end{aligned}$$

is an element of I. We can generalize this to see that

$$\{e_i \otimes e_j - e_j \otimes e_i\}, \; i \neq j$$

is the basis for I. Now, let us define the *second symmetric power of* V as

$$\text{Sym}^2 V \equiv \frac{V \otimes V}{I}$$

where, given that $\dim V = n$,

$$\dim \text{Sym}^2 V = n^2 - \frac{1}{2} n(n-1) = \frac{1}{2} n(n+1)$$

We denote the elements of $\text{Sym}^2 V$ as $x \odot y$, which are really the equivalence classes $\{x \otimes y - y \otimes x\}$. Note that

$$\begin{aligned} x \odot y - y \odot x &= \{x \otimes y - y \otimes x\} - \{y \otimes x - x \otimes y\} \\ &= \{x \otimes y - y \otimes x - y \otimes x + x \otimes y\} \\ &= \{0\} = 0 \end{aligned}$$

$\implies x \odot y = y \odot x$. That is, the $\odot$ operator is symmetric, and $\text{Sym}^2 V$ has basis

$$\{e_i \odot e_j\}_{j \geq k}$$

One realization of $\text{Sym}^2 \mathbb{R}^n$ is the set of all symmetric $n \times n$ real matrices. We can construct higher symmetric powers satisfying this property that its tensors are invariant under transpositions.

$$x_1 \odot \cdots x_i \odot \cdots x_j \odot \cdots x_n = x_1 \odot \cdots x_j \odot \cdots x_i \odot \cdots x_n$$

for all $1 \leq i \neq j \leq n$, which implies that it is invariant under any permutation $p \in S_n$ of the x_i's. Additionally,

$$\dim \text{Sym}^k V = \binom{n+k-1}{k}$$

Definition 5.15

The **symmetric algebra** of vector space V is constructed as such

$$\text{Sym}(V) \equiv \bigoplus_{k=0}^{\infty} \text{Sym}^k V$$

Note that unlike the exterior algebra, $\text{Sym}(V)$ is infinite dimensional.

Example 5.9

The inner product $(\cdot, \cdot)$ on V is an element of $\text{Sym}^2 V$, since it is a bilinear, symmetric operation on V.

$$\odot, (\cdot, \cdot) : V \times V \longrightarrow \mathbb{F}$$

There is a simple relationship between $V \otimes V$, $\Lambda^2 V$, and $\text{Sym}^2 V$.

Theorem 5.13

$$V \otimes V \simeq \text{Sym}^2 V \oplus \Lambda^2 V$$

with isomorphism defined

$$v \otimes w \mapsto \Big(\frac{1}{2}(v \odot w), \frac{1}{2}(v \wedge w)\Big)$$

This is precisely the factoring of a rank (2,0) tensor into its symmetric and antisymmetric parts.

Proof. Given $v \otimes w \in V \otimes V$,

$$v \otimes w + w \otimes v \in \text{Sym}^2 V \text{ and } v \otimes w - w \otimes v \in \Lambda^2 V$$

By defining $v \odot w$ and $v \wedge w$ as the expressions above, the isomorphism is satisfied.

Therefore, when working in $V \otimes V$, we can interpret

$$v \wedge w = \frac{1}{2}(v \otimes w - w \otimes v)$$
$$v \odot w = \frac{1}{2}(v \otimes w + w \otimes v)$$

However,

$$V \otimes V \otimes V \not\simeq \operatorname{Sym}^3 V \oplus \Lambda^3 V$$

Schur functors are used to fix this discrepancy.

Note that we have introduced these two algebras by first constructing the quotient spaces $\Lambda^n V$ and $\operatorname{Sym}^n V$ from the tensor product spaces $T^{\otimes n}$ and then direct summing these powers to construct the algebras. We will introduce another type of construction that directly takes the quotient algebra of $T(V)$ with the two-sided ideal.

5.5 Determinants and Trace

The definition of the determinant is given first and then shown that it has the corresponding properties. We will work backward and construct the determinant from its properties.

Definition 5.16 (Determinant)

The determinant of a $n \times n$ matrix A, with column vectors $a_1, a_2, \cdots, a_n$, is a function

$$\det : \operatorname{Mat}(n, \mathbb{F}) \longrightarrow \mathbb{F} \tag{225}$$

with the following three properties

1. The determinant of the identity matrix is 1.

$$\det(I) \equiv \det(e_1, e_2, \cdots, e_n) = 1 \tag{226}$$

2. Interchanging two columns a_i and a_j of A once changes the sign of $\det A$.

$$\det(a_1, \cdots, a_i, \cdots, a_j, \cdots, a_n) = -\det(a_1, \cdots, a_j, \cdots, a_i, \cdots, a_n) \tag{227}$$

3. It is a multilinear function of the n column vectors.

$$\det(a_1, \cdots, \lambda a_i + \mu a_i', \cdots a_n) = \lambda \det(a_1, \cdots, a_i, \cdots a_n) + \mu \det(a_1, \cdots, a_i', \cdots a_n) \tag{228}$$

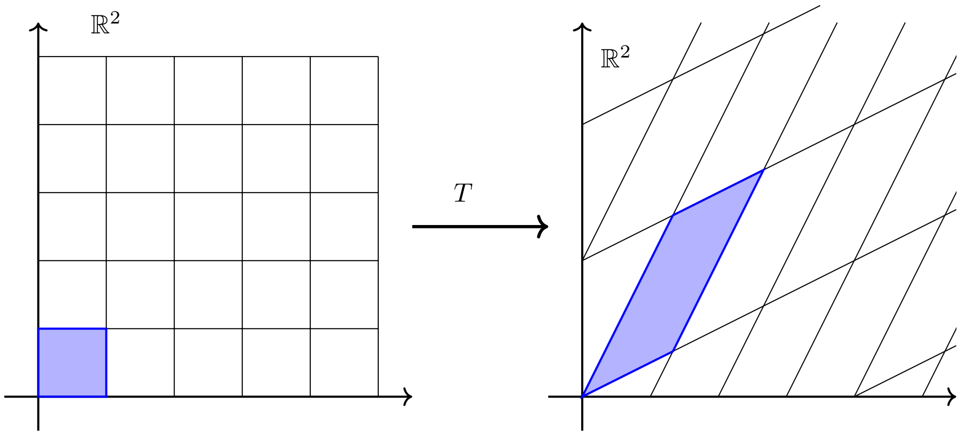


Figure 5: An important way to visualize determinants is by using the linear map visualization introduced before. That is, the determinant is the area of the transformed shaded unit square.

Theorem 5.14

The column vectors of A are linearly dependent if and only if $\det A = 0$.

Proof. By linearity, it is sufficient to prove that if two column vectors a_i and a_j of a matrix A are equal, then $\det A = 0$. This can be easily seen by property (ii) of determinants.

Theorem 5.15

$$\det\left(\prod_i A_i\right) = \prod_i \det A_i \tag{229}$$

Theorem 5.16

A matrix is invertible if and only if its determinant is nonzero.

Proof. A matrix is invertible $\iff$ it is nonsingular $\iff$ its columns are linearly independent $\iff$ its determinant is nonzero, by the previous theorem.

Corollary 5.17

Given $n \times n$ matrix A,

$$\det\left(A^{-1}\right) = \frac{1}{\det A} \tag{230}$$

Theorem 5.18

The determinants of similar matrices are equal.

Proof. Let A and B be similar matrices. Then, there exists an S such that $A = S^{-1}BS$ and

$$\det(A) = \det(S^{-1}BS = \det(S^{-1})\det(B)\det(S) = \det B \tag{231}$$

This theorem implies that the determinant is an intrinsic property of a linear transformation, so it is invariant under a change of basis. That is, choosing different matrix representations of a linear transformation does not change the determinant.

Corollary 5.19

$$\det(A) = \det(A^T) \tag{232}$$

Proof. A is similar to A^T, which will be proven in chapter 6.

Theorem 5.20

The properties of the determinant combined with the previous corollary implies that

1. Adding a scalar multiple of a row/column to another row/column doesn't affect the determinant.
2. Interchanging two rows/columns switches the sign of the determinant.
3. Multiplying a row/column by α multiplies the determinant by α.

Theorem 5.21

Let A be an $n \times n$ matrix whose first column is e_1

$$A = \begin{pmatrix} 1 & * & * & * \\ 0 & & & \\ \dots & & A_{11} & \\ 0 & & & \end{pmatrix} \tag{233}$$

where A_{11} is the $(n-1)\times(n-1)$ submatrix of A with entries $a_{ij},\ i,j > 1$. Given this,

$$\det A = \det A_{11} \tag{234}$$

Proof. Using column reduction, we can see that

$$\det A = \det \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ \dots & & A_{11} & \\ 0 & & & \end{pmatrix} \tag{235}$$

it is clear that the right hand side is equal to $\det A_{11}$ since it behaves exactly like $\det A_{11}$ with respect to the three properties.

Corollary 5.22

Let A be an upper or a lower triangular matrix. Then the determinant of A is the product of its diagonal entries. That is,

$$\det A = \prod_i a_{ii} \tag{236}$$

Proof. We apply the previous theorem recursively to satisfy when A is upper triangular. Since $\det(A) = \det(A^T)$, this fact can be applied to lower triangular matrices too.

It is once again verified that the three elementary row (and column) operations affect the determinant in the way stated in Theorem 5.5. To elaborate, since E_1, E_2, and E_3 are all lower triangular, we can compute their determinants easily

$$\det E^1_{\alpha \times i + j} = 1$$
$$\det E^2_{ij} = -1$$
$$\det E^3_{\alpha \times i} = \alpha$$

and multiplying matrix A by elementary matrices E^1, E^2, and E^3 multiplies the determinant by $1, -1$, and α, respectively.

We can describe the determinant visually. Given a linear mapping $A : V \longrightarrow V$, we can fix any basis $\{e_1, e_2, \cdots, e_n\}$ on V. Note that these basis vectors do not need to be orthogonal, nor are they restricted to any magnitude. The set of vectors

$$\Big\{ \sum_{i=1}^n c_i e_i \mid 0 \leq c_i \leq 1, i = 1, 2, \cdots, n \Big\} \tag{237}$$

forms an n-dimensional parallelepiped in V. Let the volume of this parallelepiped be U. Let W be the volume of the parallelepiped

$$\Big\{ \sum_{i=1}^n c_i A e_i \mid 0 < c_i < 1, i = 1, 2, \cdots, n \Big\} \tag{238}$$

which is formed by the transformed basis vectors $\{Ae_1, Ae_2, \cdots, Ae_n\}$. We can view this latter shape as the image of the first parallelepiped under transformation A. Then,

$$\det A = W/V \tag{239}$$

That is, the ratio of the transformed parallelepiped to the original parallelepiped is the determinant. This is consistent with the properties of the determinant. For example, if A is not isomorphic, then the parallelepiped will get "squished" into a lower-dimensional parallelepiped with volume 0. The fact that we use a ratio between the original and transformed parallelepiped allows this value to be invariant under the basis that we use.

Computationally, finding the LUP decomposition of a matrix A is the best known algorithm to compute the determinant of a general $n \times n$ matrix. That is,

$$\det A = \det L \det U \det P = \pm \det U = \pm \prod_i u_{ii} \tag{240}$$

since $\det L = 1$ and $\det P = \pm 1$.

There are other methods to compute the determinant. First, we state the simple but useful formula.

Theorem 5.23

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc \tag{241}$$

Definition 5.17

Given an $n \times n$ matrix A, the (ij)th minor of A, denoted A_{ij}, is the determinant of the $(n-1) \times (n-1)$ matrix formed by removing the ith row and j th column from A.

Theorem 5.24 (Laplace Expansion)

Let A be an $n \times n$ matrix and j any index between 1 and n. Then

$$\det A = \sum_i (-1)^{i+j} a_{ij} A_{ij} \tag{242}$$

that is, the alternating sums of the ijth minors multiplied by the ijth entries in the jth column of A. This can be done by choosing an arbitrary ith row, which leads to the alternative formula

$$\det A = \sum_j (-1)^{i+j} a_{ij} A_{ij} \tag{243}$$

Theorem 5.25 (Cramer's Rule)

Given a system of linear equations in the form $Ax = b$ where A is an $n \times n$ matrix, the solutions of this system can be expressed with the formulas

$$x_i = \frac{\det A_i}{\det A} \tag{244}$$

where $\det A_i$ is the matrix formed by replacing a_i, the ith column of A, by the column vector b.

Albeit very computationally heavy, determinants can also be used to calculate the inverse of a matrix.

Theorem 5.26

The inverse matrix A^{-1} of an invertible matrix A has the form

$$(A^{-1})_{ij} = (-1)^{i+j} \frac{\det A_{ij}}{\det A} \tag{245}$$

Definition 5.18

The trace of a square matrix A, denoted $\operatorname{Tr} A$, is the sum of its diagonal entries.

$$\operatorname{Tr}(A) = \sum_i a_{ii} \tag{246}$$

Theorem 5.27

$$\operatorname{Tr}(\lambda A + \alpha B) = \lambda \operatorname{Tr}(A) + \alpha \operatorname{Tr}(B) \tag{247}$$

Proof. Obvious if we look at the entries of A and B and see that it is bilinear.

Theorem 5.28 (Cyclic Property of the Trace)

$$\text{Tr}\left(\prod_{i=1}^{n} A_i\right) = \text{Tr}\left(A_n \prod_{i=1}^{n-1} A_i\right) \tag{248}$$

Proof. We first prove when $m = 2$. Given that the subscripts ij denote that (i,j)th element of a matrix, observe that

$$\begin{aligned} (AB)_{ij} = \sum_k A_{ik}B_{kj} \implies (AB)_{ii} &= \sum_K A_{ik}B_{ki} \\ \implies \text{Tr}(AB) &= \sum_i \sum_k A_{ik}B_{ki} \\ &= \sum_k \sum_i B_{ki}B_{ik} = Tr(BA) \end{aligned}$$

Similarly, for $m = 3$

$$\begin{aligned} (ABC)_{ij} = \sum_{k,l} A_{ik}B_{kl}C_{lj} \implies \text{Tr}(ABC) &= \sum_{i,k,l} A_{ik}B_{kl}C_{li} \\ &= \sum_{i,k,l} C_{li}A_{ik}B_{kl} = \text{Tr}(CAB) \end{aligned}$$

And so we can generalize for m.

Corollary 5.29

The trace is invariant under a change of basis. That is, the trace is an intrinsic property of a linear transformation since it does not change depending on how it is represented.

Proof. Given that A is similar to B.

$$\text{Tr}(B) = \text{Tr}(SAS^{-1}) = \text{Tr}(S^{-1}SA) = \text{Tr}(A) \tag{249}$$

Theorem 5.30

Let A be a $n \times n$ skew-symmetric matrix over $\mathbb{C}$ (or any field of characteristic $\neq 2$). If n is odd,

$$\det A = 0 \tag{250}$$

Proof.

$$\det A = \det A^T = \det -A = (-1)^n \det A \implies \det A = 0 \tag{251}$$

We can actually conclude something even futher about antisymmetric matrices.

Theorem 5.31

The determinant of an antisymmetric matrix A of even order is the square of a homogeneous polynomial of degree $n/2$ in the entries of A. That is,

$$\det A = P^2 \tag{252}$$

The polynomial P is called the **Pfaffian**.

Definition 5.19

A **Vandermonde matrix** is a square matrix whose columns form a geometric progression. That is, let $a_1, a_2, \cdots, a_n$ be n scalars. Then, $V(a_1, a_2, \cdots, a_n)$ is the $n \times n$ matrix

$$\begin{pmatrix} 1 & 1 & \ldots & 1 & 1 \\ a_1 & a_2 & \ldots & a_{n-1} & a_n \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1^{n-2} & a_2^{n-2} & \ldots & a_{n-1}^{n-2} & a_n^{n-2} \\ a_1^{n-1} & a_2^{n-1} & \ldots & a_{n-1}^{n-1} & a_n^{n-1} \end{pmatrix} \tag{253}$$

Theorem 5.32

The determinant of a Vandermonde matrix is

$$\det V(a_1, a_2, \cdots, a_n) = \prod_{j>i} (a_j - a_i) \tag{254}$$

A symmetry in the multivariable expression of a determinant can also reveal a symmetry in the matrix.

Example 5.10 (2019 Putnam A1)

The symmetric polynomial

$$f(x, y, z) = x^3 + y^3 + z^3 - 3xyz \tag{255}$$

can be expressed as the determinant of the 3×3 matrix

$$\det \begin{pmatrix} x & y & z \\ z & x & y \\ y & z & x \end{pmatrix} \tag{256}$$

We already know that the LUP decomposition is an algorithm used to compute the determinant of a general $n \times n$ matrix.

Algorithm 5.1 (Dodgson Condensation)

We will introduce another, called **Dodgson condensation**. The algorithm can be described in the following steps.

1. Let A be a given $n \times n$ matrix. Arrange A so that no zeros occur in its interior (this can be done by any combination of elementary row or column operations that would not change the determinant).
2. Create an $(n-1) \times (n-1)$ matrix B consisting of the determinants of every 2×2 submatrix of A. Explicitly,
$$B = \det \begin{pmatrix} a_{i,j} & a_{i,j+1} \\ a_{i+1,j} & a_{i+1,j+1} \end{pmatrix} \tag{257}$$
3. With this $(n-1) \times (n-1)$ matrix B, perform step 2 to obtain an $(n-2) \times (n-2)$ matrix C.

Divide each term in C by the corresponding term in the interior of A.

$$C_{i,j} = \det \begin{pmatrix} b_{i,j} & b_{i,j+1} \\ b_{i+1,j} & b_{i+1,j+1} \end{pmatrix} \Big/ a_{i+1,j+1} \tag{258}$$

4. Let $A = B$ and $B = C$. Repeat step 3 as necessary until the 1×1 matrix is found, which is the determinant.

The reason that we do not want 0s in A is because then in doing step 3 we may divide by 0.

Example 5.11

Let us find

$$\det \begin{pmatrix} -2 & -1 & -1 & -4 \\ -1 & -2 & -1 & -6 \\ -1 & -1 & 2 & 4 \\ 2 & 1 & -3 & -8 \end{pmatrix} \tag{259}$$

All of the interior elements are nonzero, so there is no need to rearrange the matrix. We calculate

$$\begin{pmatrix} -2 & -1 & -1 & -4 \\ -1 & -2 & -1 & -6 \\ -1 & -1 & 2 & 4 \\ 2 & 1 & -3 & -8 \end{pmatrix} \rightarrow \begin{pmatrix} 3 & -1 & 2 \\ -1 & -5 & 8 \\ 1 & 1 & -4 \end{pmatrix} \rightarrow \begin{pmatrix} -16 & 2 \\ 4 & 12 \end{pmatrix} \tag{260}$$

With this 2×2 matrix, we must divide each term by the interior of the original A.

$$\begin{pmatrix} -16/-2 & 2/-1 \\ 4/-1 & 12/2 \end{pmatrix} = \begin{pmatrix} 8 & -2 \\ -4 & 6 \end{pmatrix} \tag{261}$$

Calculating this determinant gives 40, and dividing by the interior of the 3×3 matrix (-5) gives $\det A = 40/-5 = -8$.

6 Spectral Theory

6.1 Spectral Theory of General Mappings

Definition 6.1 (Eigenvalue, Eigenvector)

Let $A : V \longrightarrow V$ be a linear transformation over $\mathbb{F}$. If there exists a vector $v \in V$ such that

$$Av = \lambda v, \ \lambda \in \mathbb{F}$$

then a is called an **eigenvalue** of A, and v is an **eigenvector** of A. Clearly, if a basis is realized for V and A is represented as a matrix, v would have a basis representation. However, the value of λ is invariant. The set of all eigenvalues

$$\lambda(A) \equiv \{\lambda_1, \lambda_2, \cdots, \lambda_k\}$$

is called the **spectrum** of A.

For a given eigenvalue λ and its corresponding eigenvector v, it is clear that by linearity, every vector in $\operatorname{span} v$ is an eigenvector, too.

Now that we have defined eigenvalues and eigenvectors, we first provide a visual description of these terms. Given a linear transformation $A : V \longrightarrow V$, we can visualize a certain basis of V such that all the linear transformation A does on that basis is merely extend or contract the basis vectors.

Definition 6.2

Given a $n \times n$ matrix A, the **characteristic polynomial** of A, denoted $p_A(t)$, is defined

$$p_A(t) \equiv \det(A - tI)$$

The mapping $A \mapsto p_A(t)$ can be thought of as a mapping from $\text{Mat}(n, \mathbb{F}) \longrightarrow \mathbb{F}[t]$, where $\text{Mat}(n, \mathbb{F})$ is the algebra of $n \times n$ matrices over field $\mathbb{F}$, and $\mathbb{F}[t]$ is the polynomial algebra over $\mathbb{F}$. $p_A(t)$ is invariant under matrix similarity.

The motivation for defining such a polynomial is that it allows us to compute the eigenvalues of A.

Definition 6.3

The **characteristic equation** of A is defined by equating $p_A(t) = 0$.

Theorem 6.1

The solutions of the characteristic equation of A (i.e. the roots of $p_A(t)$) is precisely the spectrum of A.

Proof. Bidirectional.

1. $(\rightarrow)$ Let there be a $t = \lambda$ such that $p_A(\lambda) = 0 \iff \det(A - \lambda I) = 0$ which is equivalent to saying that $\ker(A - \lambda I)$ is nontrivial. There must exist a $v \in \ker(A - \lambda I)$, meaning that $(A - \lambda I)v = 0 \iff Av = \lambda v$. By definition, λ is an eigenvalue of A.
2. $(\leftarrow)$ This reasoning can be extended in the opposite direction.

Theorem 6.2

Eigenvectors of a linear transformation A corresponding to different eigenvalues are linearly independent, but not necessarily orthogonal. It follows that if the characteristic polynomial of a $n \times n$ matrix A has n distinct roots, then A has n linearly independent eigenvectors.

Proof. Simple, by contradiction.

Example 6.1

It is clear that the Fibonacci sequence can be produced with matrix multiplication as such

$$\begin{pmatrix} a_{n+1} \\ a_n \end{pmatrix} = A^n \begin{pmatrix} a_1 \\ a_0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Given that

$$\lambda_1 = \frac{1+\sqrt{5}}{2}, \ \lambda_2 = \frac{1-\sqrt{5}}{2}$$

we can diagonalize A into the form

$$A = \begin{pmatrix} \frac{1}{\lambda_1 - \lambda_2} & \frac{\lambda_2}{\lambda_2 - \lambda_1} \\ \frac{1}{\lambda_2 - \lambda_1} & \frac{\lambda_1}{\lambda_1 - \lambda_2} \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & \lambda_2 \\ 1 & 1 \end{pmatrix} \implies A^n = S^{-1} \begin{pmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{pmatrix} S$$

which implies that after evaluating, we get

$$a_n = \frac{1}{\sqrt{5}} \bigg(\Big(\frac{1+\sqrt{5}}{2}\Big)^n - \Big(\frac{1-\sqrt{5}}{2}\Big)^n \bigg)$$

This is a surprising result since it also says that the expression above is always an integer for all natural number n.

Definition 6.4

Given a subspace $U_1 \subset U$ and linear transformation $T: U \longrightarrow U$. We say that U_1 is **invariant** under T if

$$u \in U_1 \implies Tu \in U_1$$

Theorem 6.3

Let $a_1, a_2, \cdots, a_n$ be the eigenvalues of A. Then

$$\sum_i a_i = \operatorname{Tr} A, \quad \prod_i a_i = \det A$$

Proof. The mapping $A \mapsto \det(A - xI)$ is a mapping from the set of $n \times n$ matrices to the polynomial algebra $\mathbb{F}[x]$. Direct application of the Viete's formulas in $\mathbb{F}[x]$ produces the statement and this result can be extended to the rest of the formulas.

Theorem 6.4 (Spectral Mapping Theorem)

Let q be any polynomial, A a square matrix with an eigenvalue a. Then:
1. $q(a)$ is an eigenvalue of $q(A)$.
2. Every eigenvalue $q(A)$ is of the form $q(a)$, where a is an eigenvalue of A.

Proof. Listed.
1. Let h be an eigenvector of A with corresponding eigenvalue a.
$$\begin{aligned} Ah = ah &\implies A^2h = Aah = aAh = a^2h \\ &\implies A^n h = a^n h \\ &\implies q(A)h = q(a)h \\ &\implies q(a) \text{ is an eigenvalue of } q(A) \end{aligned}$$
2. Let p be the eigenvalue of $q(A) \iff \det\big(q(A) - pI\big) = 0$. We expand:
$$q(s) - p = c \prod \big(s - r_i\big), r_i \in \mathbb{C} \tag{262}$$
Replacing the variable s with A, we have
$$q(A) - pI = c \prod \big(A - r_i I\big) \tag{263}$$
Since $\det\big(q(A) - pI\big) = 0$, at least one r_i, say r_k exists such that $\det\big(A - r_k I\big) = 0 \iff r_k$ is an eigenvalue of A. Since $q(r_j) - p = 0$, $p = q(r_j)$ is an eigenvalue of $q(A)$.

The following theorem is an equivalent version of the spectral mapping theorem.

Theorem 6.5

Let A be a $n \times n$ matrix and let f be a polynomial. If the chracteristic polynomial of A has factorization
$$p_A(t) = \prod_{i=1}^n (t - \lambda_i)$$
then the characteristic polynomial of the matrix $f(A)$ is given by
$$p_{f(a)}(t) = \prod_{i=1}^n (t - f(\lambda_i))$$

We can actually create a bound on the spectrum of a square matrix.

Theorem 6.6 (Gershgorin Circle Theorem)

Let $A \in \text{Mat}(n, \mathbb{C})$ with entries a_{ij}. Let $R_i = \sum_{i \neq j} |a_{ij}|$ be the sum of the absolute values of the non-diagonal entries of the ith row, and let $D_(a_{ii}, R_i) \subset \mathbb{C}$ be a closed disk with radius R_i centered at a_{ii} in the complex plane, called a **Gershgorin Disk**. Then every eigenvalue of A lies within the union of all n Gershgorin Disks. That is,
$$\lambda_j(A) \in \bigcup_{i=1}^n D_(a_{ii}, R_i) \subset \mathbb{C}, \text{ for all } j$$

Proof. Let λ be an eigenvalue of A with its eigenvector $v = (v_j)$. Scale v by multiplying it by $\pm 1/\max\{|v_j|\}_j$ to get a vector x with its maximal entry $x_i = 1$ and $|x_j| \leq 1,\ j \neq i$. Then,

$$Ax = \lambda x \implies \sum_j a_{ij} x_j = \lambda x_i = \lambda \implies \sum_{j \neq i} a_{ij} x_j + a_{ii} = \lambda$$

Applying the triangle inequality,

$$|\lambda - a_{ii}| = \left| \sum_{j \neq i} a_{ij} x_j \right| \leq \sum_{j \neq i} |a_{ij}||x_j| \leq \sum_{j \neq i} |a_{ij}| = R_i$$

Corollary 6.7

The eigenvalues of A must also lie within the Gershgorin discs C_j corresponding to the columns of A.

Proof. This is a direct result from the fact that A is similar to A^T. Alternatively, we can apply the same process in the proof above to A^T.

If one observes that the off-diagonal entries of A are small in absolute value, it can be concluded that the diagonal entries are "close" to the true eigenvalues of A. A is diagonal if and only if the Gershgorin disks are points.

Theorem 6.8 (Cayley Hamilton)

Every matrix A satisfies its own characteristic equation. That is,

$$p_A(A) = 0$$

6.2 Eigendecompositions and Jordan Normal Form

However, the entire concept of matrices are not fully grasped with just eigenvectors. If it were, then linear algebra would be a much simpler matter. To extend our toolkit, we must introduce generalized eigenvectors. From here, we will assume that our field is over $\mathbb{C}$. We use the fact that the field is over $\mathbb{C}$ because it allows us to claim that the characteristic polynomial in $\mathbb{C}[t]$ can be factored into linear components, by the fundamental theorem of algebra.

Definition 6.5

A genuine eigenvector of A satisfies $(A - aI)h = 0$. A **generalized eigenvector** f satisfies $(A - aI)^d f = 0$ for some $d \geq 1$.

To provide a visual intuition of how generalized eigenvectors transform under A, observe that

$$\begin{aligned} (A - aI)h = 0 \text{ and } (A - aI)^2 f = 0 &\implies (A - aI)^f = h && (264) \\ &\implies Af = af + h,\ Ah = ah && (265) \\ &\implies A^2 f = aAf + Ah = a^2 f + 2ah && (266) \\ &\implies A^N f = A^N f + N a^{N-1} h && (267) \end{aligned}$$

This implies that the generalized eigenvector is first scaled by a factor of a, similar to a genuine eigenvector, but then a factor of the genuine eigenvector is then added to the scaled generalized one. Note that in higher dimensions of N, a greater multiple of h must be added after scaling f.

This means that given an eigenvalue λ, there is always at least one genuine eigenvalue associated with λ. Furthermore, there may be additional generalized eigenvectors also corresponding to λ. This leads to the following definition

Definition 6.6

The subspace formed by the span of the generalized (and genuine) eigenvectors of λ form what is called the **eigenspace associated with** λ, denoted $E(\lambda)$.

We can measure the characteristics of the eigenspaces with the following definitions.

Definition 6.7

The **algebraic multiplicity** of an eigenvalue λ is the dimension of its eigenspace. It is precisely

$$\dim E(\lambda)$$

In order to compute the algebraic multiplicity of λ_i in A, we find the maximal value of d_i such that $(t-\lambda)^{d_i}$ divides $p_A(t)$. With this, we can define

$$E(\lambda) = \ker{(A - \lambda_i I)^{d_i}}$$

Theorem 6.9

Given $A : V \longrightarrow V$ with eigenspaces $E(\lambda_1), E(\lambda_2), \cdots, E(\lambda_k)$,

$$E(\lambda_1) \oplus E(\lambda_2) \oplus \cdots \oplus E(\lambda_k) = V$$

That is, every vector $v \in V$ can be uniquely expressed as the sum

$$v = h_1 + h_2 + \cdots + h_k,\ h_i \in E(\lambda_i)$$

this is called the **eigenbasis of** V.

Proof. The definition of algebraic multiplicity implies that each eigenspace is disjoint except at 0 and that their dimensions sum to $\dim V$.

Definition 6.8

The **geometric multiplicity** of an eigenvalue λ of a linear transformation A is the dimension of the span of genuine eigenvectors in its eigenspace. It is precisely

$$\dim \ker{(A - \lambda I)}$$

Note that since the span of genuine eigenvectors is a subspace of $E(\lambda)$, the geometric multiplicity is always less than or equal to the algebraic multiplicity.

Now we are ready to introduce the eigendecomposition of a linear mapping A.

Theorem 6.10

Given a linear mapping A with its eigenvalues $\lambda_1, \ldots, \lambda_k$ and associated eigenspaces $E(\lambda_1), \ldots, E(\lambda_k)$,

A maps each eigenspace to itself. That is,

$$A\big(E(\lambda_i)\big) \subset E(\lambda_i),\ i = 1, 2, \cdots, k$$

Corollary 6.11 (Jordan Normal Form)

Every linear mapping $A: V \longrightarrow V$ can be decomposed into the sum of the linear mappings of each eigenspace $E(\lambda_i)$. That is, it can be expressed in the form

$$A: \prod_i E(\lambda_i) \longrightarrow \prod_i E(\lambda_i)$$

which we can define, given $h_i \in E(\lambda_i)$,

$$A(v) = A\bigg(\sum_i h_i\bigg) = \sum_i A(h_i),\ A(h_i) \in E(\lambda_i)$$

The process of eigendecomposition for a linear mapping A is really just a clever change of basis for the $n \times n$ matrix representation of A over $\mathbb{C}$, where the new basis is now the set of genuine and generalized eigenvectors. The new matrix formed by performing the change of basis on matrix A is called the **Jordan Normal Form**, or **Jordan Canonical Form**, of A. We will now describe the construction of the JNF of an arbitrary $n \times n$ matrix.

It is actually simple. Let the eigenvalues of the matrix A be $\lambda_1, \lambda_2, \cdots, \lambda_k$, with its associated eigenspaces $E(\lambda_i)$. Let the algebraic multiplicity of eigenspace $E(\lambda_i)$ be alg_i. Then, every $n \times n$ matrix over $\mathbb{C}$ has the block form

$$J = \begin{pmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & A_k \end{pmatrix} \tag{268}$$

where each block A_i represents the transformation in $E(\lambda_i)$. This means that each A_i must be an $alg_i \times alg_i$ submatrix. The definition of the generalized eigenvectors shown in equation (11) shows that each block must be of form

$$A_i = \begin{pmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ 0 & 0 & \lambda_i & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{pmatrix} \tag{269}$$

With λ_i's in the main diagonal and 1's in the superdiagonal of A_i. The first column of A refers to the transformation of the genuine eigenvector, while the other columns refers to the transformation of the generalized eigenvectors, where λ_i refers to the scaling of the dth generalized eigenvector and the 1 refers to the adding of the $(d-1)$th generalized eigenvector to the scaled dth vector. If there are no generalized eigenvectors in an eigenspace $E(\lambda_i)$, then A_i is a 1×1 matrix (λ_i). Observe that this form is consistent with our previous theorems, especially the fact that A maps distinct eigenspaces to themselves.

Finally, the change of basis is represented through the matrix multiplication.

$$J = P^{-1}AP,\ P = \begin{pmatrix} | & | & | & | \\ f_1 & f_2 & \cdots & f_n \\ | & | & | & | \end{pmatrix}$$

where f_i is the genuine/generalized eigenvectors corresponding to the transformation represented in the ith column of J. The Jordan Normal Form of a matrix is unique up to the permutations of its diagonal blocks.

Notice that the Jordan Normal Form must be an $n \times n$ matrices over $\mathbb{C}$, not $\mathbb{R}$. However, given a matrix A over $\mathbb{R}$, we can construct a similar block diagonal form over $\mathbb{R}$. Since A is real $\implies p_A(t) \in \mathbb{R}[t]$, $\mu \in \mathbb{C}$ is a root of p_A implies that $\bar{\mu}$ is also a root. This means that in the case where $\mu = a \pm bi$ is a pair of complex eigenvectors with eigenvectors z and $\bar{z}$. The associated 2×2 Jordan block will be of form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

with the associated column vectors in P being

$$v_1 = \frac{z + \bar{z}}{2},\ v_2 = \frac{i(z - \bar{z})}{2}$$

Notice that $z \in \mathbb{C}^n$ is a complex eigenvector belonging to complex eigenvalue μ, and we make the best "approximations" of $z, \bar{z}$ and $\mu, \bar{\mu}$ with the new real vectors v_1 and v_2. Note that the Jordan block states that

$$A(v_1) = av_1 + bv_2,\ A(v_2) = -bv_1 + av_2$$

which is true since

$$\begin{aligned} A(v_1) = A\Big(\frac{z+\bar{z}}{2}\Big) &= \frac{1}{2}\big(A(z) + A(\bar{z})\big) \\ &= \frac{1}{2}\big((a+bi)z + (a-bi)\bar{z}\big) \\ &= \frac{1}{2}\big((a)(z+\bar{z}) + (bi)(z-\bar{z})\big) \\ &= a\frac{z+\bar{z}}{2} + b\frac{i(z-\bar{z}}{2} = av_1 + bv_2 \end{aligned}$$

and

$$\begin{aligned} A(v_2) = A\Big(\frac{i(z-\bar{z})}{2}\Big) &= \frac{i}{2}\big(A(z) - A(\bar{z})\big) \\ &= \frac{i}{2}\big((a+bi)z - (a-bi)\bar{z}\big) \\ &= \frac{i}{2}\big((a)(z-\bar{z}) + (bi)(z+\bar{z})\big) \\ &= a\frac{i(z-\bar{z})}{2} - b\frac{z+\bar{z}}{2} = av_2 - bv_1 \end{aligned}$$

It suffices to only modify this case for 2×2 blocks because all complex eigenvalues of real matrices must come in conjugate pairs (but this is not necessarily true for complex matrices, which have characteristic polynomials in $\mathbb{C}[t]$).

Corollary 6.12

The following 2×2 Jordan block of the form shown below can be turned into the complex Jordan block and vice versa.

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \leftrightarrow \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

However, there could be bigger Jordan blocks of generalized eigenspaces corresponding to conjugate pairs. Observe the following JNF, with columns (from left to right) corresponding to the transformations h_1 (genuine), k_1 (generalized), h_2 (genuine), and k_2 generalized).

$$\begin{pmatrix} e^{i\theta} & 1 & & \\ & e^{i\theta} & & \\ & & e^{-i\theta} & 1 \\ & & & e^{i-\theta} \end{pmatrix}$$

Using the corollary shown above, we can modify the eigenvalues and eigenvectors into real values and construct the simplest "real form" (assuming $i \neq 0, \pi$) of the matrix

$$\begin{pmatrix} \cos\theta & -\sin\theta & 1 & 0 \\ \sin\theta & \cos\theta & 0 & 1 \\ 0 & 0 & \cos\theta & -\sin\theta \\ 0 & 0 & \sin\theta & \cos\theta \end{pmatrix}$$

where the columns (from left left to right) now correspond to transformation of real eigenvectors

$$\frac{h_1 + h_2}{2}, \frac{i(h_1 - h_2)}{2}, \frac{k_1 + k_2}{2}, \frac{i(k_1 - k_2)}{2}$$

Therefore, we can state that the linear transformation represented by the two matrices in their respective bases are equivalent.

Example 6.2 (Rotation Around Vector)

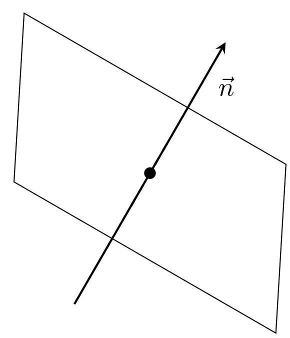


Figure 6: The linear operator that rotates around a vector v by an angle θ has an eigendecomposition of the span of v as shown (with eigenvalue 1) and the 2-dimensional plane (having two complex eigenvalues).

Definition 6.9

A matrix is **diagonalizable** if we can perform a change of basis on it to create a diagonal matrix.

Theorem 6.13

A matrix is diagonalizable if and only if its algebraic multiplicities is equal to its geometric multiplicities. That is, if the matrix only has genuine eigenvectors. This is also equivalent to saying that all of A's eigenspaces have dimension 1.

It is clear that since eigendecompositions are intrinsic to linear mappings, the JNF of similar matrices are the same. That is, the eigenvalues and the dimensions of the eigenspaces are invariant under a change of basis.

Theorem 6.14

Two matrices are similar if and only if their eigendecompositions are the same. That is, if they have the same eigenvalues and the dimensions of the corresponding eigenspaces are the same.

Proof. ($\rightarrow$) $A \sim B \implies A = S^{-1}BS = S^{-1}P^{-1}JPS = (PS)^{-1}J(PS) \implies$ JNF of A and B are the same.
($\leftarrow$) A and B have same JNF $\implies A = P^{-1}JP, B = Q^{-1}JQ \implies J = QBQ^{-1} \implies A = P^{-1}QBQ^{-1}P = (Q^{-1}P)^{-1}B(Q^{-1}P) \implies A \sim B$.

Theorem 6.15

$A \sim A^T$.

Proof. By the theorem above, it is sufficient to prove that A and A^T have the same eigendecomposition. Since $(A - \lambda I)^T = A^T - \lambda I$, $\det(A - \lambda I) = 0 \iff \det(A - \lambda I)^T = \det(A^T - \lambda I) \implies A$ and A^T have the same eigenvalues. Similarly, $\big((A - \lambda I)^d\big)^T = (A^T - \lambda I)^d \implies$ the eigenspaces of A and A^T have the same dimension.

6.3 Positive Definite Maps

Definition 6.10 (Spectral Radius)

The **spectral radius** of A is defined

$$r(A) \equiv \max_i |a_i|,\ a_i \text{ are eigenvalues} \tag{270}$$

Definition 6.11 (Positive-Semidefinite)

A self-adjoint linear mapping H from a real or complex Euclidean space onto itself is **positive definite** if

$$(x, Hx) > 0 \text{ for all } x \neq 0 \tag{271}$$

H is called **positive semidefinite** if

$$(x, Hx) \geq 0 \tag{272}$$

Theorem 6.16 (Properties of Positive Definite Matrices)

Here we state basic properties.

1. I is positive definite.
2. Positive mappings form a subspace in the space of linear mappings.

$$M, N \text{ positive} \implies M + N \text{ is positive}$$
$$M \text{ positive} \implies aM \text{ is positive for all} a \in \mathbb{F}$$

3. H positive and Q invertible $\implies Q^\dagger HQ$ positive.

Theorem 6.17

H is positive definite if and only if all of its eigenvalues are positive. Furthermore, every positive mapping is invertible.

Theorem 6.18

Every positive mapping M has a unique positive square root. That is, there exists a unique positive mapping N such that

$$N^2 = M \tag{273}$$

We denote N as $\sqrt{M}$.

Definition 6.12

Given that M, N are positive definite mappings.

$$M > N \iff M - N > 0, \text{ that is, } M \text{ is positive} \tag{274}$$

Theorem 6.19

If M, N are positive definite mappings

$$M > N \implies M^{-1} < N^{-1} \tag{275}$$

Theorem 6.20

In $\mathbb{R}^n$ endowed with the dot product, a $n \times n$ matrix A is positive definite if and only if

$$(x, Ay) = x^T Ay > 0 \tag{276}$$

for every $x, y \in \mathbb{R}^n$. A is positive semi-definite if and only if

$$(x, Ay) = x^T Ay \geq 0 \tag{277}$$

The following is a useful fact regarding inner products of $\mathbb{R}^n$.

Theorem 6.21

The set of all inner products that can be defined on $\mathbb{R}^n$ is bijective to the set of positive-definite symmetric $n \times n$ matrices A (which is itself bijective to the set of all positive-definite mappings). That is, every inner product of $\mathbb{R}^n$ can be defined

$$(x, y) \equiv x^T Ay \tag{278}$$

Note that when $A = I_n$, the inner product is the regular dot product.

6.4 Singular Value Decomposition

We now proceed to another crucial decomposition, called the singular value decomposition. While the JNF allows us to choose the most convenient choice of basis for a square matrix, the Singular Value Decomposition (SVD) allows us to decompose general $m \times n$ matrices.

Theorem 6.22 (Singular Value Decomposition)

Any linear mapping M from an n-dimensional inner product space to a m-dimensional inner product space can be decomposed into

$$M = U\Sigma V^\dagger = \begin{pmatrix} | & | & | & | \\ y_1 & y_2 & \dots & y_m \\ | & | & | & | \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & 0 \\ & \ddots & & & \vdots \\ & & \sigma_p & & 0 \\ & & & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix} \begin{pmatrix} — & x_1 & — \\ — & x_2 & — \\ — & \vdots & — \\ — & x_n & — \end{pmatrix} \tag{279}$$

where $U \in U(m), V \in U(n)$ and Σ has diagonal elements with nonnegative real entries. Also, $p = \text{rank}(M) \leq \min\{n, m\}$. This form is known as the **singular value decomposition**. The columns of U, denoted y_i, are called the **left singular vectors** and the columns of V (i.e. the rows of $V^\dagger$), denoted x_i, are called the **right singular vectors**. The diagonal entries of Σ are called the **singular values**. The SVD is unique up to the order of singular values, but it is generally constructed so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p$.

To provide a brief, yet unrigorous, justificiation of why the SVD exists, we look at the linear mapping $M : X \longrightarrow Y$, with $\dim X = n, \dim Y = m$. If M is injective ($\iff m \geq n$), given the basis $\{e_i\}$ for X, we can complete the linearly independent set $\{Me_i\}_{i=1}^n$ to a basis in Y and represent M as the mapping

$$\Sigma_{inj} = \begin{pmatrix} & & \\ & I_n & \\ & & \\ 0 & \dots & 0 \end{pmatrix} \tag{280}$$

If M is surjective ($\iff m \leq n$), then given basis $\{f_i\}_{i=1}^m$ of Y, we can choose a basis $\{e_j\}_{j=1}^n$ of X such that $M(e_i) = f_i (i = 1, 2, \ldots, m)$, and $M(e_i) = 0$ when $i > m$. This produces the matrix

$$\Sigma_{surj} = \begin{pmatrix} & & 0 \\ & I_m & \vdots \\ & & 0 \end{pmatrix} \tag{281}$$

We now present the following theorem without proof.

Theorem 6.23

Any map $M : X \longrightarrow Y$ can be written as a surjective map followed by an injective map.

This theorem implies that any map, when given the right choice of basis, can be written as

$$\Sigma_{inj}\Sigma_{surj} = \begin{pmatrix} & & & \cdots & 0 \\ & I_p & & \cdots & 0 \\ & & & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & & & & 0 \\ & \cdots & & & \cdots \\ & & 1 & & 0 \\ & & & \cdots & \cdots \\ 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \tag{282}$$

where $\text{rk}(M) = p =$ the number of 1's in $\Sigma_{inj}\Sigma_{surj}$. As for choosing the proper set basis for X and Y, we can find these passive transformations in the unitary groups $U(n)$ and $U(m)$.

We now present a geometric description of the singular value decomposition. Think of the unit n-ball being rotated and flipped ($V^\dagger$ applied) under the unitary transformation. Then, it is stretched along its othogonal axes to result in an ellipsoid living in an m-dimensional space. The factor of stretching and compressing the axes are precisely the singular values. Finally, this ellipsoid is rotated and flipped (U applied) back to its original basis.

Theorem 6.24

Geometrically, we can see that the largest singular value is the matrix norm, also called the operator norm.

$$\|M\| = \sigma_1 \tag{283}$$

Theorem 6.25 (Properties of Singular Values)

Given linear mapping A from a n-dimensional inner product space to m-dimensional inner product space,

1. $\sigma_i(A) = \sigma_i(A^T) = \sigma_i(A^\dagger) = \sigma_i(\bar{A})$
2. $\forall\, U \in U(m), V \in U(n),\ \sigma_i(A) = \sigma_i(UAV)$
3. Relation to eigenvalues
$$\sigma_i^2(A) = \lambda_i(A^\dagger A) = \lambda_i(AA^\dagger) \tag{284}$$

We now present the (not the best) process of computing SVD of small matrices by hand. Given matrix M, $M = U\Sigma V^\dagger \implies M^\dagger M = V\Sigma^2 V^\dagger$. The eigenvalues of $M^\dagger M$ are σ_i^2 with corresponding eigenvectors being the columns of V, which can all be found by putting $M^\dagger M$ into JNF. We repeat this process for $MM^\dagger = U\Sigma^2 U^\dagger$ to find the eigenvectors that make up the column vectors of U.

Theorem 6.26

Let $A : X \longrightarrow Y$, with $\dim X = n, \dim Y = m$, and let $k \leq \min\{m, n\}$, with $A = U\Sigma V^\dagger$. Then, amongst all rank k $m \times n$ matrices B, the matrix $A^{(k)}$ minimizes

$$\|A - B\|_2,\ A^{(k)} = U\Sigma^{(k)} V^\dagger \tag{285}$$

and $\Sigma^{(k)}$ is Σ with $\sigma_{k+1} = \sigma_{k+2} = \cdots = 0$. Therefore, to see how "close" B is to A, we can compare the singular values of A and B, given that they both have the same unitary matrices U and V.

The singular value decomposition has many applications in high dimensional data analysis and data compression. For example, in a set of m data points in $\mathbb{R}^n$ that each lie in the rows of matrix A, if the singular values of A suddenly drops (e.g. $120, 118, 107, 98, 2, 1, 0.3, \cdots$) then we can determine that the points "almost" lie in a subspace in $\mathbb{R}^n$. Knowing this allows us to compress high dimensional data to $A^{(k)}$, which is a more manageable form. This is especially useful in the data compression of electronic images, where each pixel is treated as a single number to form a matrix.

It can also be used to define the "pseudo-inverse" of a matrix that may not be invertible.

Definition 6.13 (Pseudo-Inverse)

Given matrix $M = U\Sigma V^\dagger$ in SVD, we define the **pseudo-inverse** $M^+ = V\Sigma^+ U^\dagger$, where Σ^+ is Σ with entries σ_i^{-1}, or 0 if $\sigma_i = 0$. For example,

$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \implies \Sigma^+ = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{286}$$

$\implies M^+ M = V\Sigma^+\Sigma V^\dagger$. If M is square and all $\sigma_i \neq 0$, then $M^\dagger M = VV^\dagger \implies M^\dagger M = I \implies M^+ = M^{-1}$.

By computing the SVD of M, where $\sigma_p \neq 0, p = \operatorname{rk} M = \operatorname{rk} \Sigma$, we can automatically compute the 4 fundamental spaces.

$$M = U\Sigma V^\dagger = \begin{pmatrix} & & | & \\ & & | & \\ & U & | & U' \\ & & | & \\ & & | & \end{pmatrix} \begin{pmatrix} \sigma_1 & & & & 0 \\ & \sigma_2 & & & 0 \\ & & \cdots & & \cdots \\ & & & \sigma_p & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} & & & & \\ & & V^\dagger & & \\ & & & & \\ — & — & — & — & — \\ & & V^{\dagger\prime} & & \end{pmatrix} \tag{287}$$

1. $\operatorname{Im} M = C(U)$
2. $\ker M = R(V^{\dagger\prime}) = C(V')$
3. $\ker M^\dagger = C(U')$
4. $\operatorname{Im} M^\dagger = C(V) = R(V^\dagger)$

One of the main differences between the JNF and SVD of a matrix A lies in how they are affected by perturbations in the elements of A. For example, take the small change

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \longrightarrow A' = \begin{pmatrix} 1 & 1 \\ 0 & 1.00001 \end{pmatrix} \tag{288}$$

The SVD of A' will "change" continuously for changes in the elements of A, but the JNF of A is completely different from the JNF of A'. More specifically, the JNF of A is A itself, but the JNF os A' is now diagonalizable, meaning that the 2-dimensional eigenspace $E(1)$ "breaks up" into two 1-dimensional eigenspaces from small perturbations.

Definition 6.14 (Frobenius Norm)

The **Frobenius norm** of a $m \times n$ matrix A is defined

$$\|A\|_F \equiv \sqrt{\operatorname{Tr}(A^\dagger A)} = \sqrt{\operatorname{Tr} \Sigma^2} = \bigg(\sum_{i,j} a_{ij}^2 \bigg)^{\frac{1}{2}} \tag{289}$$

By Singular Value Decomposition, we can reduce its calculations to

$$\|A\|_F = \sqrt{\sum_i \sigma_i^2} \tag{290}$$

where σ_i's are the singular values. Clearly,

$$\|A\|_2 \leq \|M\|_F \tag{291}$$

In quantum mechanics, the Frobenius norm is also called the **Hilbert Schmidt norm** in the context of infinite dimensional Hilbert spaces.

We end by defining two more common decompositions of square matrices.

Theorem 6.27 (Schur Decomposition)

Every $n \times n$ matrix A over $\mathbb{C}$ can be decomposed into

$$A = QTQ^\dagger \tag{292}$$

where $Q \in U(n)$ and T is upper triangular.

Proof. This is an obvious result of the Grahm-Schmidt algorithm.

Theorem 6.28 (Polar Decomposition)

Every complex $n \times n$ matrix A can be factored into the form

$$A = UP \tag{293}$$

where $U \in U(n)$ and P is a positive semidefinite self-adjoint matrix. If A is a real matrix, then $U \in O(n)$.

Proof. We take the SVD to get

$$A = W\Sigma V^\dagger \tag{294}$$

and we can assign

$$U = WV^\dagger,\ P = V\Sigma V^\dagger \tag{295}$$

Since V, W are unitary, this confirms that P is positive definite and self-adjoint along with U being unitary. Thus, the existence of the SVD implies the existence of the polar decomposition.

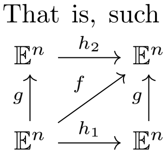
6.5 Polar Decomposition

Theorem 6.29 (Polar Decomposition)

Given a Euclidean space $\mathbb{E}^n$ and any linear endomorphism f of $\mathbb{E}^n$, there are two positive definite self-adjoint linear maps $h_1, h_2 \in \text{End}(\mathbb{E}^n)$ and $g \in \text{O}(n)$ such that

$$f = g \circ h_1 = h_2 \circ g \tag{296}$$

That is, such that f can be decomposed into the following as shown in this commutative diagram.



7 Normed Vector Spaces

Given a vector space V, we can induce different structures on it to allow us to conduct different measurements on it. For example, the endowment of the basis on V allows us to represent vector as an n-tuple of scalars, and from topology, we can also define an arbitrary topology or metric on a vector space as well.

Definition 7.1 (Norm)

A **norm** on a vector space V over field $\mathbb{C}$ is a mapping

$$\rho : V \longrightarrow \mathbb{R} \tag{297}$$

satisfying three properties

1. $\rho(x) \geq 0$, with $\rho(x) = 0 \iff x = 0$
2. For $a \in \mathbb{C}$, $\rho(ax) = |a|\rho(x)$
3. $\rho(x + y) \leq \rho(x) + \rho(y)$

A norm allows us to define some notion of a magnitude or length on each vector in V. A vector space V with a norm is called a **normed space**, denoted (V, ρ).

Example 7.1 (Absolute Value)

The absolute value function $|\cdot| : \mathbb{C} \longrightarrow \mathbb{R}_+$ is an example of a norm on the 1 dimensional space $\mathbb{C}$.

Example 7.2 (Euclidean Norm, L_2-Norm)

The Euclidean norm of a vector $x \equiv (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n$ is defined

$$||x||_2 \equiv \left(\sum_{i=1}^n x_i^2 \right)^{\frac{1}{2}} \tag{298}$$

This is the most commonly used norm in $\mathbb{R}^n$.

Example 7.3 (Taxicab Norm, Manhattan Norm)

The Taxicab norm of $x \equiv (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n$ is defined

$$||x||_1 \equiv \sum_{i=1}^n |x_i| \tag{299}$$

Example 7.4 (Infinity Norm, L_∞-Norm)

The Infinity norm of vector $x \equiv (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n$ is defined

$$||x||_\infty \equiv \max \{|x_1|, |x_2|, \cdots, |x_n|\} \tag{300}$$

Example 7.5 (p-norm, L_p-Norm)

Let $p \geq 1$ be a real number. The p-norm of a vector

$$x \equiv (x_1, x_2, \cdots, x_n)^T \in \mathbb{R}^n \tag{301}$$

is defined

$$||x||_p \equiv \bigg(\sum_{i=1}^n x_i^p \bigg)^{\frac{1}{p}} \tag{302}$$

For $0 < p < 1$, this function could be of some use, but it is not considered a norm since it violates the triangle inequality. When $p = 1$ and $p = 2$, the norm is the Taxicab norm and Euclidean norm, respectively, and

$$\lim_{p \rightarrow \infty} || \cdot ||_p = || \cdot ||_\infty \tag{303}$$

Theorem 7.1 (Norm Induces Metric)

Every norm induces a metric in the following way

$$d(x, y) \equiv \rho(x - y) \tag{304}$$

However, a metric does not necessarily induce a norm because the definition

$$\rho(x) \equiv d(x, 0) \tag{305}$$

is not guaranteed to have all properties of the norm.

7.1 Dual Spaces

7.2 Norms of Linear Mappings

Since the algebra of linear operators is itself a vector space, we can also define structures on it, too. We focus on matrix norms.

Definition 7.2 (Operator Norm)

Let $A : X \longrightarrow U$ be linear. Then, we define

$$\|A\| = \sup_{\|x\|=1} \|Ax\| \tag{306}$$

Note that $\|Ax\|$ is measure with respect to the norm of U and $\|x\|$ the norm of X.

There is a very nice visualization of this.

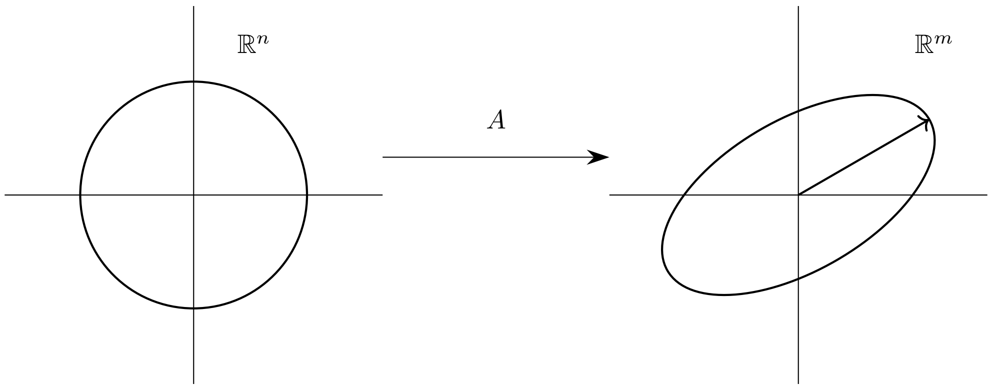


Figure 7: The norm of A is the length of the major axis of the ellipsoid. Given that $\dim X = n$, imagine the n-dimensional unit ball in X being transformed under A. The image of the ball should be an ellipsoid (of dimension $\leq m$) in U.

Theorem 7.2

$$\|Az\| \leq \|A\|\|z\| \text{ for all } z \in X \tag{307}$$
$$\|A\| = \sup_{\|x\|,\|v\|=1} (Ax, v) \tag{308}$$

Proof.

$$\|Az\| \leq \sup \|Az\| = \sup \bigg\| A \frac{z}{\|z\|} \bigg\| = \|A\|\|z\|$$
$$\|u\| \equiv \max_{\|v\|=1} (u, v) \implies \|Ax\| \equiv \max_{\|v\|=1} (Ax, v) \implies \|A\| \equiv \sup_{\|x\|,\|v\|=1} (Ax, v)$$

Theorem 7.3 (Properties of Matrix Norm)

Let there exist any $k \in \mathbb{F}$, with any $A, B : X \longrightarrow U$, $C : U \longrightarrow V$. Then,

1. $\|kA\| = |k|\|A\|$
2. $\|A + B\| \leq \|A\| + \|B\|$
3. $\|CA\| \leq \|C\|\|A\|$
4. $\|A\| = \|A^\dagger\|$

Theorem 7.4

A simple lower and upper bound of $\|A\|$ can be defined

$$r(A) \leq \|A\| \leq \left(\sum_{i,j} a_{ij}^2 \right)^{\frac{1}{2}} \tag{309}$$

Matrix norms have extremely useful applications in determining the existence of inverses.

Theorem 7.5

Let A be invertible and

$$\|A - B\| < \frac{1}{\|A^{-1}\|} \tag{310}$$

in the sense that B is "close" to A. Then B is invertible.

8 Numerical Methods

General complexity of eigendecomposition, multiplication, etc.

8.1 Matrix Multiplication

8.1.1 Strassen Algorithm

When computing the product two $n \times n$ matrices A and B to another $n \times n$ matrix C, since each entry of C is the product of a row of A with a column of B, and since C has n^2 entries, we need n^3 scalar multiplications to compute (as well as $n^3 - n^2$ additions). In order words, the computing efficiency of the algorithm is at $O(n^3)$. However, there are faster algorithms than this. This is algorithm is known as the **Strassen Algorithm** (however, there do exist faster algorithms).

Theorem 8.1 (Strassen Algorithm)

Let A, B be 2×2 matrices such that $AB = C$. That is, component-wise,

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \tag{311}$$

where for $i, j = 1, 2$,

$$c_{ij} = a_{i1} b_{1j} + a_{i2} + b_{2j} \tag{312}$$

Then, let us define

$$\begin{aligned}
P_1 &= (a_{11} + a_{22})(b_{11} + b_{22}) \\
P_2 &= (a_{21} + a_{22}) b_{11} \\
P_3 &= a_{11}(b_{12} - b_{22}) \\
P_4 &= a_{22}(b_{21} - b_{11} \\
P_5 &= (a_{11} + a_{12}) b_{22} \\
P_6 &= (a_{21} - a_{11})(b_{11} + b_{12}) \\
P_7 &= (a_{12} - a_{22})(b_{21} + b_{22})
\end{aligned}$$

Then, the theorem states that we the entries of C are

$$\begin{aligned}
c_{11} &= P_1 + P_4 - P_5 + P_7 \\
c_{12} &= P_3 + P_5 \\
c_{21} &= P_2 + P_4 \\
c_{22} &= P_1 + P_3 - P_2 + P_6
\end{aligned}$$

This algorithm for multiplying 2×2 matrices requires 7 scalar multiplications, while regular multiplication requires 8. Using block multiplication, we can use this algorithm to calculate any matrix of order 2^k. That is, to calculate $2^k \times 2^k$ matrices, we have to perform seven multiplications of blocks of size $2^{k-1} \times 2^{k-1}$, and doing this recursively, it reduces it down to

$$7^k = 2^{k \log_2 7} = n^{\log_2 7} \tag{313}$$

where n is the order of the matrices being multiplied.

Additionally, the number of scalar additions or subtractions needed is bounded by

$$6 \times 7^k = 6 \times 2^{k \log_2 7} = 6n^{\log_2 7} \tag{314}$$

Since $\log_2 7 \approx 2.807 < 3$, this algorithm does indeed have more computational efficiency. Note that matrices whose order is not a power of 2 can be turned into one by adjoining a suitable number of 1s on the diagonal.

Theorem 8.2 (Conjecture)

For any positive number ε, there is an algorithm that computes the product of two $n \times n$ matrices with computational efficiency of $O(n^{2+\varepsilon})$.

8.2 Solving System of Linear Equations

In this section, we will concern ourselves with a system of equations with only *one solution*, represented by the matrix equation

$$Ax = b \tag{315}$$

where A is an invertible square matrix, b some given vector, and x the vector of unknowns to be determined.

An algorithm for solving the system takes as inputs the matrix A and vector b and outputs some approximation to the solution x. However, with billions of arithmetic operations on top of each other, the errors can accumulate. Algorithms for which this does not happen are said to be **arithmetically stable**.

The use of finite digit arithmetic places an absolute limitation on the accuracy with which the solution can be determined. To demonstrate this, let us imagine a change δb being made in the vector b, which causes a corresponding change in x, denoted δx.

$$A(x + \delta x) = b + \delta b \implies A\delta x = \delta b \tag{316}$$

To compare the changes in x with the changes in b, we define the following variable.

Definition 8.1

The **relative change in x with the relative change in** b is the quantity

$$\frac{|\delta x|}{|x|} \bigg/ \frac{|\delta b|}{|b|} \tag{317}$$

where the norm is convenient for the problem (usually a numerical approximation of the Euclidean norm for floating point numbers). We will assume the use of the Euclidean norm from now on. We can rewrite the value as the expression below with the following upper bound, denoted by $\kappa(A)$, called the **condition number**.

$$\frac{|b|}{|x|}\frac{|\delta x|}{|\delta b|} = \frac{|Ax|}{|x|}\frac{|A^{-1}\delta b|}{|\delta b|} \leq |A||A^{-1}| \equiv \kappa(A) \tag{318}$$

where $|A|$ is the matrix norm of A.

A high value of this relative change would mean that small perturbations in b would cause large changes in x.

Note that $\kappa(A) \geq 1$. Notice also that the higher the condition number $\kappa(A)$, the harder it is to solve the equation $Ax = b$, and $\kappa(A) = \infty$ when A is not invertible. For a k-digit floating point arithmetic, the relative error in b can be as large as 10^{-k}, meaning that the relative error in x can be as large as $10^{-k}\kappa(A)$.

Let β be the largest absolute value of the eigenvalues of A and α as the smallest absolute value of the eigenvalues of A. Then

$$\beta \leq |A|, \frac{1}{\alpha} \leq |A^{-1}| \implies \frac{|\beta|}{|\alpha|} \leq \kappa(A) \tag{319}$$

An algorithm that generates an exact solution after a finite number of arithmetic steps is called a *direct method* (e.g. Gauss Elimination). An algorithm that generates successive approximations that converge onto the solution is called an *iterative method*.

The methods mentioned in this section will be iterative.

Definition 8.2

Let $\{x_n\}$ be the sequence of approxmations generated by such an algorithm. The deviation of x_n from the true value x is caelled the **error at the nth stage**, denoted by e_n.

$$e_n \equiv x_n - x$$

The amount by which the nth approximation fails to satisfy the equation $Ax = b$ is called the nth residual, denoted by r_n.

$$r_n \equiv Ax_n - b$$

Error and residual are related by the equation.

$$r_n = Ae_n$$

Note that since we do not know x, the error cannot be calculated, but the residuals can be. We further restrict our scope to solving linear systems in which A is real, positive, and self-adjoint. Clearly, we already know that $|A| = \beta$, and since A is positive, we can conclude that

$$|A^{-1}| = \frac{1}{\alpha} \tag{320}$$

which implies that

$$\kappa(A) = \frac{\beta}{\alpha} \tag{321}$$

8.2.1 Method of Steepest Descent

Theorem 8.3

Assume that $n \times n$ matrix A is self-adjoint. The solution of $Ax = b$ minimizes the functional

$$E(y) \equiv \frac{1}{2}(y, Ay) - (y, b) \tag{322}$$

where $(\cdot, \cdot)$ is the Euclidean dot product. That is, the solution x is

$$x = \min \big\{E(y)\big\} = \min \Big\{\frac{1}{2}(y, Ay) - (y, b)\Big\} \tag{323}$$

Proof. Add to $E(y)$ a constant, that is a term independent of y to define a new function F.

$$F(y) \equiv E(y) + \frac{1}{2}(x, b)$$

Then, by self adjointness of A, we can express $F(y)$ as

$$\begin{aligned} F(y) &= \frac{1}{2}(y, Ay) - (y, b) + \frac{1}{2}(x, b) && (324) \\ &= \frac{1}{2}(y, Ay) - \frac{1}{2}(y, Ax) + \frac{1}{2}(Ax, x) - \frac{1}{2}(Ay, x) && (325) \\ &= \frac{1}{2}\big(y, A(y - x)\big) + \frac{1}{2}\big(A(x - y), x\big) && (326) \\ &= \frac{1}{2}\big(y - x, A(y - x)\big) && (327) \end{aligned}$$

Since $F(x) = 0$ and $F(x) \geq 0$ (since it is an inner product with repsect to $y - x$), $F(y)$, and also $E(y)$, takes a minimum at $y = x$.

Now to actually compute the value of x, we us the method of steepest descent. Note that $E : \mathbb{R}^m \longrightarrow \mathbb{R}$, so we can utilize ordinary calculus on it. The gradient of E can be computed by the formula

$$\operatorname{grad} E(y) = Ay - b \tag{328}$$

So, if our nth approximation is x_n, then the $(n+1)$st approximation, x_{n+1}, is calculated as

$$x_{n+1} = x_n - s(Ax_n - b) \tag{329}$$

where s is the step length in the direction $-\operatorname{grad} E$. By calculating the residual $r_n = Ax_n - b$, we can rewrite the above to

$$x_{n+1} = x_n - sr_n \tag{330}$$

Rather than keeping s constant, we can actually determine an optimal value of s at the nth step, denoted s_n, which minimizes $E(x_{n+1})$. This quadratic minimum problem is easily solved, since

$$E(x_{n+1}) = \frac{1}{2}\big(x_n - sr_n, A(x_n - sr_n)\big) - \big(x_n - sr_n, b\big) \tag{331}$$

$$= E(x_n) - s(r_n, r_n) + \frac{1}{2}s^2(r_n A r_n) \tag{332}$$

By taking the derivative and computing the value of s where $E(x_{n+1}) = 0$, we see that the minimum is reached when

$$s = s_n \equiv \frac{(r_n, r_n)}{(r_n, Ar_n)} \tag{333}$$

Theorem 8.4

The sequence of approximations $\{x_n\}$, with s optimized to be s_n, converges to the solution of $Ax = b$.

The error bound for this algorithm is

$$\|e_n\|^2 \leq \frac{2}{\alpha}\left(1 - \frac{1}{\kappa(A)}\right)^n F(x_0) \tag{334}$$

which shows that the error e_n tends to 0 in $\mathbb{R}^m$. However, this algorithm has a very slow rate of convergence for large $\kappa(A)$.

8.2.2 Method of Chebyshev Polynomials

The disadvantage of the method of steepest descent mentioned in the end of the last subsection renders it quite outdated and obsolete. this next method has a much better error bound that can handle large values of κ more efficiently. However, we will need a positive lower bound m for the smallest eigenvalue of a and an upper bound m for the largest eigenvalue. That is,

$$m \leq \alpha, \beta \leq m \tag{335}$$

and all the eigenvalues of a lie in the interval $[m, m]$. It follows that

$$\kappa = \frac{\beta}{\alpha} < \frac{m}{m} \tag{336}$$

We generate the same sequence of approximations $\{x_n\}$ by the same recursion formula

$$x_{n+1} = x_n - s(ax_n - b) \iff x_{n+1} = (i - s_n a)x_n + s_n b \tag{337}$$

Since the solution of x satisfies the formula; that is, since $x = (i - s_n a)x + s_n b$, we subtract this equation from the top to get

$$e_{n+1} = (i - s_n a)e_n \tag{338}$$

Doing this recursively, we can deduce an explicit formula

$$e_n = p_n(a)e_0 = \prod_{n=1}^{n} (1 - s_n a) \tag{339}$$

This allows us to estimate the size of e_n.

$$||e_n|| \leq ||p_n(a)|| ||e_0|| \tag{340}$$

The norm of a self adjoint matrix a is the largest $|a|$, where a is the eigenvalue, and the spectral mapping theorem states that the eigenvalues p of $p_n(a)$ are of the form $p = p_n(a)$, where a is an eigenvalue of a. this means that

$$||a|| \leq \max_{m \leq a \leq m} |a| \implies ||p_n(a)|| \leq \max_{m \leq a \leq m} |p_n(a)| \tag{341}$$

So, we are left with the bound

$$||e_n|| \leq ||e_0|| \max_{m \leq a \leq m} |p_n(a)| \tag{342}$$

To get the best estimate of e_n, we have to choose the $s_1, s_2, \cdots, s_n$ so that the polynomial p_n has a small maximum on the interval $[m, m]$. note that the polynomial p_n satisfies the normalizing condition

$$p_n(0) = 1 \tag{343}$$

Oo find such a polynomial, we must first define chebyshev polynomials.

Definition 8.3

The n**th chebyshev polynomial** t_n is defined for $-1 \leq u \leq 1$ by

$$t_n(u) = \cos(n\theta), \quad u = \cos(\theta) \tag{344}$$

Theorem 8.5

Among all polynomials p_n of degree n that satisfy $p_n(0) = 1$, the one that has the smallest maximum on $[m, m]$ is the *rescaled chebyshev polynomial* that rescales values from $[-1, 1]$ to the interval $[m, m]$ while preserving the condition that $p_n(0) = 1$. this polynomial is expressed as

$$p_n(a) \equiv t_n\bigg(\frac{m + m - 2a}{m - m}\bigg) \bigg/ t_n\bigg(\frac{m + m}{m - m}\bigg) \tag{345}$$

Now, assuming that $m/m \approx \kappa$,

$$t_n\bigg(\frac{m + m}{m - m}\bigg) = t_n\bigg(\frac{\frac{m}{m} + 1}{\frac{m}{m} - 1}\bigg) \approx t_n\bigg(\frac{\kappa + 1}{\kappa - 1}\bigg) \tag{346}$$

since $|t_n(u)| \leq 1$ for $|u| \leq 1$, this also implies that

$$t_n\bigg(\frac{m + m - 2a}{m - m}\bigg) \leq 1 \tag{347}$$

Combining this together, we get

$$||e_n|| \leq ||e_0|| \max_{m \leq a \leq m} |p_n(a)| = ||e_0|| \Bigg/ t_n\left(\frac{\kappa+1}{\kappa-1}\right) \tag{348}$$

It is a fact that higher order chebyshev polynomials tend to infinity faster once the value reaches out of $[-1,1]$, meaning that as $n \to \infty$, $t_n\big((\kappa+1)/(\kappa-1)\big)$ will also tend to infinity (note that $(\kappa+1)/(\kappa-1)$ is a constant, implying that e_n tends to 0 as n tends to infinity. the error bound for e_n is given by the following

$$||e_n|| \leq 2\left(1+\frac{2}{\sqrt{\kappa}}\right)^{-n} ||e_0|| \approx 2\left(1-\frac{2}{\sqrt{\kappa}}\right)^{n} ||e_0|| \tag{349}$$

Once again, this confirms that $e_n \to 0$ as $n \to \infty$. furthermore, when κ is large, the error bound works with $\sqrt{\kappa}$, which is must smaller than κ itself. so, e_n converges much faster through this algorithm than through the method of steepest descent.

8.3 Eigendecomposition