Contents

1 Linear Algebra

Over the course of this power round, the main tools we leverage in every theorem and proof are mathematical objects. One such object you are likely familiar with is a **set**, a collection of any other things, for instance, numbers. What other mathematical objects are there? Often, we obtain new objects by bestowing an *additional structure* on sets. The set of rational numbers, $\mathbb{Q}$, is one example. By giving the rational numbers the operations of addition $(+)$ and multiplication $(\cdot)$, we can start to ask more specific questions about it, since the elements are no longer interchangeable.

The study of mathematical objects is very useful, because it helps us generalize results. If we prove some theorem about $\mathbb{Q}$ based on the fact that it has addition and multiplication, then any other object with those same operations would have the same theorem! So if we start with a precise definition for an "object" that has a minimal set of properties, in one fell swoop we prove the theorem for any possible instance of this object. Then, the results will have surprising generality and often represent vast common themes in the world of mathematics.

In this section, we introduce **vector spaces**, which are the currency of *linear algebra*. There are concrete, intuitive examples of vector spaces that you are probably already familiar with. While we encourage you to examine how our results manifest in these examples, try also to understand them abstractly. The abstract results apply to *anything* which is a vector space!

1.1 Vector Spaces

1.1.1 Definitions

We begin by defining the notion of a field. The most common examples of fields are $\mathbb{R}$ (the real numbers) and $\mathbb{Q}$ (the rational numbers) - intuitively, they are the familiar sets where we can add, subtract, multiply and divide. However, we must axiomatize all of the "familiar properties" that we usually take for granted. It may look overwhelming, but every property here is one that you use every time you do arithmetic. First, we recall that a *binary operation* on a set S is just a map $S \times S \to S$.

Definition 1.1. A *field* F is a set with at least two elements equipped with two binary operations:

- *Addition*: $+ : F \times F \to F,\ (a,b) \mapsto a+b$,
- *Multiplication*: $\cdot : F \times F \to F,\ (a,b) \mapsto a \cdot b$,

satisfying the following axioms:

1. **Associativity**: for all $a,b,c \in F$, $(a+b)+c = a+(b+c)$, and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

2. **Commutativity**: for all $a,b \in F$, $a+b = b+a$, and $a \cdot b = b \cdot a$.

3. **Identities**: there exist additive and multiplicative identities in F, referred to as $0,1 \in F$, respectively, such that for all $a \in F$, we have that $0+a = a$ and $1 \cdot a = a$.

4. **Additive inverse**: for all $a \in F$, there exists $b \in F$ such that $a+b = 0$. Such b is often denoted $-a$.

5. **Multiplicative inverse**: for all $a \in F$ such that $a \neq 0$, there $b \in F$ such that $a \cdot b = 1$. Such b is often denoted a^{-1}.

6. **Distributive property**: for all a,b,c, the following property holds: $a \cdot (b+c) = a \cdot b + a \cdot c$.

Remark. If you would like to prove that a certain object is a field, also **do not forget to check closure under addition and multiplication**! That is, do not forget to check that for all $a,b \in F$, we have $a+b \in F$ and $a \cdot b \in F$. "Closure" is not listed as one of the field axioms, but really what it comes from is from the *definition of addition and multiplication as binary operations*.

We have already seen some familiar examples of fields. The rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are fields (with the usual addition and multiplication operations). If you know some number theory, you might also be able to see that, for any prime number p, the set of integers modulo p (usually denoted by $\mathbb{Z}/p\mathbb{Z}$) equipped with the addition and multiplication operations for modular arithmetic, is a field (this is called a finite field, since the cardinality of $\mathbb{Z}/p\mathbb{Z}$ is p). Less obvious is that the following subset of the real numbers

$$\mathbb{Q}[\sqrt{2}+\sqrt{2023}] := \left\{ \sum_{k=0}^{3} a_k(\sqrt{2}+\sqrt{2023})^k : a_0, a_1, a_2, a_3 \in \mathbb{Q} \right\}$$

is also a field, again with the usual addition and multiplication operations.

Problem 1.1 (4 Points)**.** Let F be a field.

(a) (1 Point) Prove that the additive and multiplicative identity of F are both unique. *Hint: assume there are two additive identities, for instance. Play around with them, and try to prove they are equal!*

(b) (1 Point) Prove that for any $a \in F$, there is a unique additive inverse of a in F (justifying the notation $-a$). That is, there is exactly one element $b \in F$ such that $a+b=0$. Also prove that for any nonzero $b \in F$, there is a unique multiplicative inverse of b in F (justifying the notation b^{-1}).

(c) (2 Points) Prove that 0 cannot have a multiplicative inverse. *Hint: There is one property of a field you* ***must*** *use for this proof. Which property is it?*

Definition 1.2. A *vector space V* over a field F (we sometimes say: V is an *F-vector space*) is a set equipped with two operations:

- *Addition*: $+ : V \times V \to V,\ (v,w) \mapsto v+w$,
- *Scalar multiplication:* $\cdot : F \times V \to V,\ (a,v) \mapsto a \cdot v$,

satisfying the following axioms:

1. **Associativity of addition**: for all $u,v,w \in V$, $(u+v)+w = u+(v+w)$.
2. **Commutativity of addition**: for all $v,w \in V$, $v+w = w+v$.
3. **Additive identity**: there is a zero vector $\mathbf{0} \in V$, such that for all $v \in V$, we have $\mathbf{0}+v = v$.
4. **Additive inverse**: for all $v \in V$, there exists $w \in V$ such that $v+w = \mathbf{0}$. Such w is often denoted $-v$.
5. **Identity of scalar multiplication**: $1 \in F$ acts as the identity for all $v \in V$, i.e., $1 \cdot v = v$.
6. **Scalar multiplication and field multiplication commute**: for all $a,b \in F$ and $v \in V$, we have that $(a \cdot b) \cdot v = a \cdot (b \cdot v)$.
7. **Distributive property**: for all $a,b \in F$ and for all $v,w \in V$, we have $a \cdot (v+w) = a \cdot v + a \cdot w$ and $(a+b) \cdot v = a \cdot v + b \cdot v$.

Note that in the case of both vector spaces and fields, sometimes $\cdot$ is omitted when denoting multiplication. You should also take care to distinguish $0 \in F$ and $\mathbf{0} \in V$. Elements of fields are typically called *scalars*, and elements of vector spaces are called *vectors*. The zero vector is different from the scalar zero. Moreover, the same symbol $\cdot$ is used for field multiplication *and* scalar multiplication of vectors. The former is an operation between two scalars, and the latter is an operation between a scalar and a vector. So this abuse of notation is benign, since you should always be able to figure out which one is being used in context.

Another instance of benign abuse of notation occurs when we write $0 \in V$, instead of using the bolded form $\mathbf{0}$. Again, when reading algebraic expressions, it should be clear, by context, whether any 0 refers to that of a field or that of a vector space.

Example 1.1. Consider the set $F^n = \{(a_1, \dots, a_n) \mid a_i \in F \text{ for } i = 1, \dots, n\}$. We define addition on this set by $(a_1, \dots, a_n) + (b_1, \dots b_n) = (a_1 + b_1, \dots, a_n + b_n)$, and scalar multiplication by $c \cdot (a_1, \dots, a_n) = (ca_1, \dots, ca_n)$. Then, F^n equipped with these operations is a vector space over F: in fact, every vector space axiom follows from a corresponding axiom for the field F, as we encourage you to verify. For instance, $(b+c) \cdot (a_1, \dots, a_n) = ((b+c)a_1, \dots, (b+c)a_n) = (ba_1 + ca_1, \dots, ba_n + ca_n) = (ba_1, \dots, ba_n) + (ca_1, \dots, ca_n) = b \cdot (a_1, \dots, a_n) + c \cdot (a_1, \dots, a_n)$. This proof uses only field distributivity and definitions of operations to prove a distributivity axiom of vector spaces.

This example of a vector space is likely the most familiar to you. For instance, $\mathbb{R}^2$ is the space of all 2-dimensional vectors, with the first component being the x-coordinate and the second component the y-coordinate. However, there are various other vector spaces, as you will see in 1.3.

Let us briefly mention some easy-to-prove basic "properties" of vector spaces. Of course, given a vector space V, we have *uniqueness of the additive identity* $\mathbf{0} \in V$, and *uniqueness of additive inverses in* V. The method to prove these uniqueness statements is the same as in Problem 1.1 (which we urge you to review and/or ponder about if you are still confused!). These may seem obvious, but really they do require the vector space (and field) axioms to verify! As usual, F is a field, and V is an F-vector space. For every $a \in F, v \in V$:

- $0v = \mathbf{0}$, $a\mathbf{0} = \mathbf{0}$.
- $(-1)v = -v$.
- $av = \mathbf{0}$ implies $a = 0$ or $v = \mathbf{0}$.

To verify the last one, you would need to use the multiplicative inverse property of a field.

Problem 1.2 (5 Points)**.** Constructions on Vector Spaces (let V be an F-vector space):

(a) (1 Point) A *subspace* of a vector space V is a subset of V that is *also a vector space*. Show that a subset $W \subseteq V$ is a subspace if and only if $\mathbf{0} \in W$ and for any $v, w \in W$ and $a \in F$, we have that $av + (-w) \in W$. We say that W is a *trivial* subspace if $W = \{\mathbf{0}\}$ and *nontrivial* otherwise. Note that a subspace must "inherit" the addition and scalar multplication operations of the larger vector space it lies in.

(b) (1 Point) If U, W are subspaces of V, then the sum of the vector spaces, denoted $U + W$, is the subset $\{u + w \mid u \in U, w \in W\} \subseteq V$. Prove that $U + W$ is a subspace of V. Also convince yourself that this $+$ operation on vector spaces is associative, so we can use $U_1 + \cdots + U_n$ to denote the sum of n vector spaces (you do not need to prove this).

(c) (2 Points) Prove that $U_1 + \cdots + U_n$ is the smallest subspace of V containing U_i for all i, motivating the notation $U + W$ (by "smallest," I mean that for any subspace W of V containing U_i for all i, we have $U_1 + \cdots + U_n \subseteq W$).

(d) (1 Point) If any element in $U_1 + \cdots + U_n$ can be written **uniquely** as $\sum u_i$ for $u_i \in U_i$, then $U_1 + \cdots + U_n$ is called a *direct sum*, and is written as $U_1 \oplus \cdots \oplus U_n$. Prove that $U + W$ is a direct sum if and only if $U \cap W = \{\mathbf{0}\}$. *Hint: say $u_1 + w_1 = u_2 + w_2$. How can we find an element in both U and W?*

The above Problem 1.2 is **extremely important**!! Pretty much everything here is **as important** as the above definitions, and the concepts here will be referenced **many times in the future**. So, make sure to *read the problem statements* before you proceed.

Problem 1.3 (3 Points)**.** Here are some example vector spaces. You should define addition and scalar multiplication appropriately. Let F be a field.

(a) (1 Point) Prove that the set of polynomials with coefficients in F is a vector space (over F). Prove that the set of polynomials with degree at most d, for any $d \in \mathbb{N}$, is also a subspace. (Note: you should not need to define any multiplication between polynomial.)

(b) (2 Points) Prove that if S is any set, and V is any vector space over F, then the set of all functions from S to V is a vector space. *Hint: for addition and scalar multiplication, define* $f+g$ *by* $(f+g)(s) = f(s)+g(s)$, *and* $(cf)(s) = cf(s)$.

1.1.2 Dimension

From the remainder of Section 1, we will use F to denote a field. Additionally, U, W, and V will be used to denote any given vector space over F (potentially subspaces).

Many vector spaces are sets with an infinite number of elements. However, some of these vector spaces seem bigger than others. For instance, $\mathbb{R}$ seems smaller than $\mathbb{R}^2$. In terms of *cardinality*, the sizes of $\mathbb{R}$ and $\mathbb{R}^2$ are the same (we will not go into this). But, in terms of *vector spaces*, $\mathbb{R}$ is indeed “smaller” than $\mathbb{R}^2$. We will understand this rigorously by defining a *dimension* of a vector space. An expression of the form $a_1v_1 + \cdots + a_nv_n$ (with $a_i \in F, v_i \in V$) is called a *linear combination* of $v_1, \ldots, v_n$. Note that whenever we write $\{v_1, \ldots, v_n\}$, or (sometimes, omitting the brackets for brevity) we write $v_1, \ldots, v_n$, it is assumed implicitly that $v_1, \ldots, v_n$ are mutually distinct elements of the vector space V. Now, we provide the two key definitions:

Definition 1.3. The *span* of a finite set of vectors $S = \{v_1, \ldots, v_n\}$ in V is defined as follows. If S is non-empty, then the span of S is the subspace $\{a_1v_1 + \cdots + a_nv_n \mid a_1, \ldots, a_n \in F\}$ of V. If S is empty, then the span of S is $\{\mathbf{0}\}$. In other words, the span of S is the *smallest subspace of* V *that contains* S. We denote the span of S by $\operatorname{span}(S)$.

Example 1.2. Consider $\mathbb{R}^3$, the space of 3-dimensional real vectors. The span of the vector $(1,1,1)$, for instance, will be the line where $x = y = z$. The span of $(1,1,1), (1,0,0)$ is the plane formed by the x-axis and the line $x = y = z$. The span of $(1,1,1), (1,0,0)$ is the same as the span of $(0,1,1), (2,0,0)$.

Definition 1.4. A nonempty finite set of vectors $S = \{v_1, \ldots, v_n\}$ in V is *linearly independent* if any vector $v \in \operatorname{span}(S)$ can be written *uniquely* as a linear combination of S. S is called *linearly dependent* if it is not linearly independent.

Remark. You may verify that $S = \{v_1, \ldots, v_n\}$ is a linearly independent set in V is equivalent to saying that $\operatorname{span}(v_1, \ldots, v_n) = \operatorname{span}(v_1) \oplus \cdots \oplus \operatorname{span}(v_n)$. The definition can also be slightly simplified. Say that a vector v could be written distinctly as $a_1v_1 + \cdots + a_nv_n$ and $b_1v_1 + \cdots + b_nv_n$. Then, $(a_1 - b_1)v_1 + \ldots (a_n - b_n)v_n = \mathbf{0}$. So, $\mathbf{0}$, a vector in $\operatorname{span}(S)$, could then be written as this linear combination. However, $\mathbf{0}$ has an even simpler, different linear combination: $0v_1 + \cdots + 0v_n = \mathbf{0}$. We have just proved that if some vector in $\operatorname{span}(S)$ can be written in two different ways, then 0 can also be written in two different ways. And the converse is certainly true. This proof demonstrated a key technique in linear algebra: often global statements can be reinterpreted as statements about $\mathbf{0}$. Below, we restate this equivalent definition of linear independence.

Definition 1.5. A set of vectors $S = \{v_1, \ldots, v_n\}$ is called *linearly independent* if $a_1v_1 + \cdots + a_nv_n = \mathbf{0}$ implies that $a_1, \ldots, a_n = \mathbf{0}$.

Now, we define "finite-dimensional" vector spaces. Much of elementary linear algebra is the theory of vector spaces with finite dimension because it is easier to understand their structure.

Definition 1.6. A vector space V is finite-dimensional if there exists a finite set of vectors $S = \{v_1, \ldots, v_n\}$ that spans V. A finite set S of vectors is called a *basis* of V if it is linearly independent and spans V.

(The remainder of the Power Round will almost be always concerned with finite-dimensional vector spaces, unless otherwise indicated. In particular, the following definition of basis only applies to finite-dimensional vector spaces.)

At this point, we should mention that the proofs of Lemma 1.7 and Theorems 1.8, 1.9, and 1.10 do *rely on the fact that F be a field* (i.e., the field axioms). First, we present the Lemma, which will be crucial in proving the following Theorems. If you are interested, you may try to prove the Lemma yourself. You are welcome to cite the Lemma in any Problem on this Power Round that follows thereafter.

Lemma 1.7. If S is a linearly dependent set of vectors $\{v_1, \ldots, v_n\}$ that spans V, then there exists some v_i such that $v_i \in \operatorname{span}(v_1, \ldots, v_{i-1})$, and $S \backslash \{v_i\}$ spans V.

Using this lemma, the following key theorems can be proven:

Theorem 1.8. Any set $\{v_1, \ldots, v_n\}$ that spans V, a finite-dimensional vector space, can be reduced (by removing some elements) to a basis.

Proof. Problem 1.4. □

In particular, *every finite-dimensional vector space has a basis*.

Theorem 1.9. Any set that spans V, a finite-dimensional vector space, has at least as many elements as any other set that is linearly independent in V.

Proof. Problem 1.4. □

Theorem 1.10. Let V be a finite-dimensional vector space. Any linearly independent set $v_1, \ldots, v_n$ can be extended to a basis.

Proof. Problem 1.4. □

Theorem 1.9 proves a corollary that is foundational to finite-dimensional linear algebra:

Corollary. Every basis of a finite dimensional vector space V has the same cardinality. In particular, we can define the *dimension* of a vector space, $\dim V$, to be the cardinality of any of its bases.

Proof. Say that S_1 and S_2 are two bases of V. Then, S_1 is linearly independent and S_2 is spanning, which proves that S_2 is at least as big as than S_1. The same argument proves that S_1 is at least as big as S_2, proving they have the same cardinality. □

Problem 1.4 (7 Points)**.** Basis Theorems:

(a) (2 Points) Prove Theorem 1.8.

(b) (3 Points) Prove Theorem 1.9. *Hint: Use Lemma 1.7 and consider a maximal set of linearly independent vectors.*

(c) (2 Points) Prove Theorem 1.10. *Hint: Use Lemma 1.7 and Theorem 1.9.*

Now, in a very precise sense, we can see why $\mathbb{R}^2$ is "larger" than $\mathbb{R}$ as a vector space over $\mathbb{R}$. Note that $\{(1,0),(0,1)\}$ is a basis of $\mathbb{R}^2$ (verify this if it is not clear!), and $\{1\}$ is a basis of $\mathbb{R}$. So they have dimensions 2 and 1 respectively. It is a simple exercise that F^n is a vector space with dimension n, and the basis $\{(1,\dots,0),\dots,(0,\dots,1)\}$ is called the *standard basis*.

Example 1.3. Consider the fields $\mathbb{R},\mathbb{C}$. Observe that $\mathbb{C}$ is a 2-dimensional vector space over $\mathbb{R}$ in the obvious way. Indeed, the set $\{1,i\}$ (where $i=\sqrt{-1}$ as usual) is a basis of $\mathbb{C}$ as an $\mathbb{R}$-vector space.

On the other hand, $\mathbb{C}$ is certainly a 1-dimensional vector space over $\mathbb{C}$ (itself). This indicates that the concept of dimension also *depends on the base field*.

Problem 1.5 (4 Points)**.**

(a) (2 Points) Say that $V_1,\dots,V_n$ are subspaces of finite-dimensional V such that $V_1\oplus\cdots\oplus V_n$. Prove that $\dim(V_1\oplus\cdots\oplus V_n)=\sum_i \dim V_i$ by concatenating the bases of each space.

(b) (2 Points) Prove that given any vector space V and a nontrivial proper subspace U, there exists another subspace W such that $U\oplus W=V$.

Problem 1.6 (4 Points)**.** Prove that $\mathbb{R}$ is a vector space over $\mathbb{Q}$. Is it infinite-dimensional, or finite-dimensional? Justify (i.e. prove) your answer. If $\mathbb{R}$ is finite-dimensional, find the dimension. You may use the fact that π is transcendental (there is no polynomial p with rational coefficients such that $p(\pi)=0$).

1.2 Linear Maps

1.2.1 Definitions

A linear map $T:V\to W$ is a function from one vector space to another with additional structure. It is said to "respect the structure of vector spaces," which is made precise in the following sense:

Definition 1.11. Let V,W be vector spaces over the same field F. A map $T:V\to W$ is called a *linear transformation* or a *linear map* if it obeys the following properties:

1. $aT(v)=T(av)$ for any $a\in F$, $v\in V$.

2. $T(v+w)=T(v)+T(w)$ for any $a\in F$, $v\in V$.

When working with linear maps, you should make sure to distinguish elements of V with elements of W, and especially the $\mathbf{0}\in V$ with the $\mathbf{0}\in W$!

Why are linear maps defined in this manner? Well the theory of functions between vector spaces is only useful if the function connects the vector space properties of one space to another. If the vector spaces have completely different definitions of addition and multiplication that the function does not respect, then the function becomes rather weak from the perspective of linear algebra.

Example 1.4. Consider the fields $\mathbb{R},\mathbb{C}$. Recall that $\mathbb{C}$ is a 2-dimensional vector space over $\mathbb{R}$, and a 1-dimensional vector space over $\mathbb{C}$. Then, the map $T:\mathbb{C}\to\mathbb{C}, z\mapsto iz$ is a linear map, whether we view $\mathbb{C}$ as an $\mathbb{R}$-vector space or as a $\mathbb{C}$-vector space! Sometimes, we say that this is either an $\mathbb{R}$-*linear map* or a $\mathbb{C}$-*linear map* to be clear on the base field!

For instance, to check $\mathbb{C}$-linearity, let $z_1, z_2 \in \mathbb{C}$ (as vectors) and $a \in \mathbb{C}$ (as a scalar). We have $T(z_1+z_2) = i(z_1+z_2) = iz_1 + iz_2 = T(z_1)+T(z_2)$, and $T(az_1) = iaz_1 = a(iz_1) = aT(z_1)$. Thus, T is $\mathbb{C}$-linear. The exact same proof works for $\mathbb{R}$-linearity, because $\mathbb{R}$ is contained in $\mathbb{C}$ ($\mathbb{R}$ is called a *subfield* of $\mathbb{C}$).

In the following two definitions and the following theorem, $T : V \to W$ is a linear map.

Definition 1.12. The set $\{v \in V \mid T(v) = \mathbf{0}\}$ is a subset of V, called the *kernel* of T, or $\ker(T)$.

Definition 1.13. The set $T(V) := \{T(v) \mid v \in V\}$ is a subset of W, called the *image* of T, or $\operatorname{im}(T)$.

Theorem 1.14.

1. The image of T is a subspace of W.
2. The kernel of T is a subspace of V.

Proof. Let $v_1, v_2 \in V, a \in F$. If $w_1 := T(v_1)$ and $w_2 := T(v_2)$ are two elements in the image of T, then $aw_1 - w_2 = aT(v_1) - T(v_2) = T(av_1 - v_2)$ is in the image. Moreover, since $T(\mathbf{0}) = T(\mathbf{0}+\mathbf{0}) = T(\mathbf{0}) + T(\mathbf{0})$, $T(\mathbf{0}) = \mathbf{0}$. So, $\mathbf{0} \in \operatorname{im}(T)$. Now by Problem 1.2, the image is a subspace.

Now, say that $T(v_1) = \mathbf{0}$ and $T(v_2) = \mathbf{0}$. Then, $T(av_1 - v_2) = a \cdot \mathbf{0} - \mathbf{0} = \mathbf{0}$. Moreover, $T(\mathbf{0}) = \mathbf{0}$, so $\mathbf{0} \in \ker(T)$. So, once again by 1.2, the kernel is a subspace. □

The dimension of the image of a linear map $T : V \to W$ is called the *rank* of T, and the dimension of the kernel of T is called the *nullity* of T.

1.2.2 Injectivity, Surjectivity, and Isomorphisms

Let $T : V \to W$ be a linear map. Recall that T is called *injective* if T does not send any two elements to the same vector: i.e., if $v \neq w$ ($v, w \in V$), then $T(v) \neq T(w)$. Also recall that T is called *surjective* if $\operatorname{im}(T) = W$. The intuitive significance of injective functions (in general) is that they fit a full copy of the domain into the codomain, without collapsing any points. For linear maps, however, we can show the following additional result!

Theorem 1.15. A linear map $T : V \to W$ is injective if and only if $\ker(T) = \{\mathbf{0}\}$.

Proof. Note T injective implies that only $\mathbf{0}$ gets mapped to $\mathbf{0}$ by T, and therefore $\ker(T) = \{\mathbf{0}\}$ (the nullity of T is 0). Conversely, if $\ker(T) = \{\mathbf{0}\}$, then for $v, w \in V$, note $T(v) = T(w) \implies T(v-w) = \mathbf{0} \implies v - w = \mathbf{0} \implies v = w$. This is precisely injectivity. □

Definition 1.16. A linear map $T : V \to W$ which is both injective and surjective is called a *linear isomorphism*.

Theorem 1.17. A linear map $T : V \to W$ is a linear isomorphism if and only if it has a (two-sided) inverse linear map T^{-1}.

Linear isomorphisms are a key concept. If there is an isomorphism between two vector spaces, they are called isomorphic and you should think of those spaces as being, for all intents and purposes, **the same**. As vector spaces, they are effectively identical, with the linear isomorphism T just being a "translation" between two different representations of the same object.

Problem 1.7 (3 Points)**.** In this problem, you may use the fact that *the composition of two linear maps is again linear*. This is straightforward to show, and you may assume this fact witihout proof. This is yet another essential property of linear maps!

(a) (1 Point) Prove that the composition of two injective linear maps is again an injective linear map. That is, if $T_1 : U \to V,\ T_2 : V \to W$ are injective linear maps, then $T_2 \circ T_1 : U \to W$ is again an injective linear map.

(b) (1 Point) Prove that the composition of two surjective linear maps is again a surjective linear map. That is, if $T_1 : U \to V,\ T_2 : V \to W$ are surjective linear maps, then $T_2 \circ T_1 : U \to W$ is again a surjective linear map.

(c) (1 Point) Now, let $T_1 : U \to V,\ T_2 : V \to W$ be linear maps, and suppose $T_2 \circ T_1 : U \to W$ is a linear isomorphism. Prove that T_1 is injective and T_2 is surjective.

Problem 1.8 (2 Points)**.** Let $T : V \to W$ be a linear isomorphism of finite dimensional vector spaces V, W. Prove that $\dim V = \dim W$. You may not directly cite the Rank-Nullity Theorem.

1.2.3 Foundational Theorems

We now prove a fundamental result about the rank and nullity of a linear map out of a finite-dimensional vector space.

Theorem 1.18 (Rank-Nullity Theorem)**.** Say $T : V \to W$ is a linear map, where V is a finite-dimensional vector space. Then $\dim \operatorname{im}(T) + \dim \ker(T) = \dim V$.

Proof. Take some basis $v_1, \ldots, v_m$ of $\ker(T)$. Since it is linearly independent in V, it can be extended to a basis $v_1, \ldots, v_n$ of V. Then, note that $T(v_1), \ldots, T(v_n)$ span $\operatorname{im}(T)$. After, all, any $T(v) = T(a_1v_1 + \cdots + a_nv_n) = a_1T(v_1) + \cdots + a_nT(v_n)$. But since $T(v_1), \ldots, T(v_m)$ are all zero, this implies that $T(v_{m+1}), \ldots, T(v_n)$ span $\operatorname{im}(T)$.

Now, assume by way of contradiction that $T(v_{m+1}), \ldots, T(v_n)$ are linearly dependent in $\operatorname{im}(T)$ (i.e., not a basis). Then, we would have that $\mathbf{0} = a_{m+1}T(v_{m+1}) + \cdots + a_nT(v_n) = T(a_{m+1}v_{m+1} + \cdots + a_nv_n)$, for some $a_1, \ldots, a_n \in F$ with $a_i \neq 0$ for some i. This would imply that the nonzero vector (since one of the a_i is nonzero and $v = v_{m+1}, \ldots, v_n$ are linearly independent) $a_{m+1}v_{m+1} + \cdots + a_nv_n$ is in the kernel of T. However, this implies that v can be written as a nonzero linear combination of $v_1, \ldots, v_m$, and we already know it can be written as a linear combination of $v_{m+1}, \ldots, v_n$—this contradicts linear independence. So, we have that the dimension of $\operatorname{im}(T)$ is $n - m$, completing the proof! □

For an intuition for this proof, imagine that the size of V can be deduced from two quantities - how big its image is (when it is mapped into W by T), and how redundant the map T is. The latter is made precise by the number of elements that "hit zero," i.e. the dimension of the kernel.

Problem 1.9 (3 Points)**.** Consider the function $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by $T((v_1, v_2, v_3)) = (v_1, v_2)$. Prove it is a linear map, describe the kernel and image, and prove that the kernel is dimension 1 and the image dimension 2 (explicitly verifying that the rank-nullity theorem holds in this example). Describe both vector spaces geometrically.

Another crucial result is the following: the linear maps from one space to another form a vector space.

Definition 1.19. Let V,W be vector spaces over the same field F. Then, the set of all linear maps from V to W is denoted by $\mathrm{Hom}(V,W)$.

Theorem 1.20. The set $\mathrm{Hom}(V,W)$ is a vector space over F.

Proof. In Problem 1.3, we prove that the set of functions from S, a set, to W, for any fixed vector space W, is a vector space itself. Taking S to be V, we have that the set of functions from V to W is a vector space, where addition of functions is pointwise addition ($T_1+T_2 : v \mapsto T_1(v)+T_2(v)$) and scalar multiplication is pointwise scalar multiplication ($aT : v \mapsto aT(v)$).

To prove this result, therefore, we must show that $\mathrm{Hom}(V,W)$ is a subspace of such functions. Using the condition in 1.2, say we have two linear maps T and S. Then, $aT-S : v \mapsto aT(v)-S(v)$ is a linear map. To see this, we observe that:

$$(aT-S)(v+w) = aT(v+w)-S(v+w) = aT(v)-S(v)+aT(w)-S(w) = (aT-S)v+(aT-S)w)$$
$$(aT-S)(cv) = aT(cv)-S(cv) = c(aT(v)-S(v)) = c(aT-S)v$$

This shows that $\mathrm{Hom}(V,W)$ is a subspace, completing the proof. □

Unfortunately, we still don't know how to understand $\mathrm{Hom}(V,W)$ from a practical standpoint. This is addressed in the following section.

1.3 Coordinates

1.3.1 Vectors

All of the discussion so far has been very abstract - but when we talk about finite-dimensional vector spaces, a vector space can be made much more explicit.

Theorem 1.21. A finite-dimensional vector space V with dimension n is isomorphic to F^n.

Proof. Take some basis $v_1,\dots,v_n$ of V. Define a map $T : F^n \to V$ that maps $(a_1,\dots,a_n)$ to $a_1v_1+\cdots+a_nv_n$. This map is clearly linear, and $\ker(T) = \{\mathbf{0}\}$, since if $a_1v_1+\cdots+a_nv_n = \mathbf{0}$, then by linear independence $(a_1,\dots,a_n) = \mathbf{0}$. So, T is injective. Moreover, for any $v \in V$, it can be expressed as $a_1v_1+\cdots+a_nv_n$ for some $a_1,\dots,a_n \in F$, which implies that T is surjective. So, by Theorem 1.17, T is a linear isomorphism. □

Corollary. All vector spaces over F of the same dimension are isomorphic.

Proof. If V and W both have dimension n, There is an isomorphism from V to F^n and one from F^n to W. Since the composition of two isomorphism is an isomorphism, we find that V is isomorphic to W. □

The theorem allows us to write any vector $v \in V$ "with respect to a basis" as a coordinate vector $(a_1,\dots,a_n)$. In a sense, every isomorphism of a vector space to F^n represents another choice of basis. The choice allows us to express a vector in terms of scalars. But it is crucial to remember that *a vector exists independent of its basis*. We use a certain basis to manipulate and express vectors in a convenient manner, but no one basis is best. That is why this section, about coordinates, explores the choice of basis in detail.

1.3.2 Linear Maps and Matrices

If we can write vectors in terms of some basis, is it also possible to express linear maps in terms of a basis? The answer to this question is yes, using the notion of *matrices*. For positive integers m,n, an *$m \times n$ matrix with entries in F* is simply a table with m rows and n columns, where each of the mn entries is an element in F. For instance, $\begin{pmatrix} 1 & 2 & 3 \\ 1 & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$ is a 2×3 matrix with entries in $\mathbb{Q}$.

We denote by $M_{m\times n}(F)$ the *vector space of $m \times n$ matrices with entries in F*—as a vector space, it is isomorphic to F^{mn}, since its dimension is mn.

Now, given matrices $A \in M_{l\times m}(F), B \in M_{m\times n}(F)$. We may form a matrix $AB = C \in M_{l\times n}(F)$ using a formula called *matrix multiplication*. If

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lm} \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix},$$

then C is the matrix

$$C = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{l1} & c_{l2} & \cdots & c_{ln} \end{bmatrix},$$

where for $i = 1,2,\ldots,l$ and $j = 1,2,\ldots,n$:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj}.$$

Example 1.5. Matrices with entries in $\mathbb{Q}$: $A := \begin{pmatrix} 1 & 2 & 3 \\ 1 & \frac{1}{2} & \frac{1}{3} \end{pmatrix}$ is a 2×3 matrix, and $B := \begin{pmatrix} 3 & 1 & 5 & 6 \\ 4 & 2 & 3 & 1 \\ 5 & 3 & 1 & 9 \end{pmatrix}$ is a 3×4 matrix. Then, AB is the 2×4 matrix $\begin{pmatrix} 26 & 14 & 14 & 35 \\ \frac{20}{3} & 3 & \frac{41}{6} & \frac{19}{2} \end{pmatrix}$.

Note that matrix multiplication is *associative*. That is, for integers $l,m,n,p \geq 1$, if $A \in M_{l\times m}(F), B \in M_{m\times n}(F), C \in M_{n\times p}(F)$, then $(AB)C = A(BC)$ as $l \times p$ matrices. Here, on the LHS, we first carry out the matrix multiplication AB, whereas on the RHS, we first carry out the matrix multiplication BC. Associativity of matrix multiplication can be proven by a direct computation, i.e. without any reference to the theory of vector spaces or linear maps. In what follows, you may use this fact without proof.

A final additional remark. In what follows, it helps to regard vectors in F^n as *column vectors*. That is, given $v = (a_1, a_2, \ldots, a_n) \in F^n$, it helps to visualize v as $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$.

Theorem 1.22. Take a linear map $T : V \to W$, and assume both V and W are finite-dimensional. Fix bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$ of V and W.

Denote by $\phi_V : V \to F^n$ the linear isomorphism that takes a given $v \in V$ to the column vector $(a_1, \ldots, a_n) \in F^n$, where $v = a_1v_1 + \cdots + a_nv_n$ (cf. Theorem 1.21), and define $\phi_W : W \to F^m$ the same way with respect to the basis $w_1, \ldots, w_m$. There is a linear isomorphism ϕ from $\mathrm{Hom}(V,W)$ to $M_{m\times n}(F)$, such that for any $T \in \mathrm{Hom}(V,W)$ and any $v \in V$, $\phi_w(T(v)) = \phi(T) \cdot \phi_v(v)$, where the operation $\cdot$ is *matrix multiplication*.

Here, we say that $a_1, \dots, a_n$ are the *coefficients* of v with respect to the $v_1, \dots, v_n$ basis.

Before we prove this theorem, what is it telling us? It says that with respect to a basis on both the input and output spaces, linear maps can be represented as matrices—and scaling and adding the matrices is the same as scaling and adding the linear maps. And, just as the linear maps act on vectors, matrices act on the coordinate vectors (by matrix multiplication)! Matrix multiplication is precisely equivalent to the application of a linear map on a vector, once bases are chosen.

Proof. Say we have a linear map T. Then, we know that for any v_i in our basis, $T(v_i) = a_{1i}w_1 + \cdots + a_{mi}w_m$, for unique coefficients a_{ji}, since $w_1, \dots, w_m$ are a basis. We define:

$$\phi : \mathrm{Hom}(V,W) \to M_{m\times n}(F), \quad \phi(T) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

Here, the ith column are the coefficients of the output $T(v_i)$. The map ϕ is linear, because if $T, S \in \mathrm{Hom}(V,W)$ are linear maps, then the coefficients (with respect to the $w_1, \dots, w_n$ basis) of $T(v_i) + S(v_i) = (T+S)(v_i)$ are the sum of the coefficients of $T(v_i)$ and of $S(v_i)$. Similarly, for $a \in F$, the coefficients of $aT(v_i)$ are the coefficients of $T(v_i)$ scaled by a.

Now we must prove that $\phi_W(T(v)) = \phi(T) \cdot \phi_V(v)$. In the following, we use the subscript notation $*_i$ to mean the ith component of the column vector. Then, the ith component $\phi_W(T(v))_i$ may be expressed as:

$$\begin{aligned} \phi_W(T(v))_i &= \phi_W(T(b_1v_1 + \cdots + b_nv_n))_i \\ &= \phi_W(T(v_1))_i b_1 + \cdots + \phi_W(T(v_n))_i b_n \\ &= \sum_j a_{ij} b_j \end{aligned}$$

This formula gives precisely the output of the matrix multiplication

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

at every index, proving the result. □

Remark. The sums and expansions in this proof are bound to get confusing, so it is worth going through the proof carefully until it starts to make sense. The reason why this proof is a little laborious is the same reason why matrix multiplication is tedious! It turns out that both linear maps and matrix multiplication are representing the same process.

Corollary. If V and W are finite-dimensional vector spaces of dimension n and m, respectively, then $\mathrm{Hom}(V,W)$ is a finite-dimensional vector space of dimension mn.

Definition 1.23. In the notation of Theorem 1.22, we say that $\phi(T)$ is the *matrix of the linear map T* with respect to the bases $v_1, \dots, v_n$ of V and $w_1, \dots, w_m$ of W.

Remark. Sometimes we consider linear maps $T : V \to V$ from the a vector space to itself. In this case, we may obtain a matrix of T *with respect to a single basis $v_1, \dots, v_n$ of V* (just that this basis is applied to both the domain V and the codomain V in the situation of Theorem 1.22).

Proof. Isomorphic vector spaces have the same dimension, and $M_{m\times n}(F)$ has dimension mn. □

Here is another beautiful correspondence between linear maps and their matrix representations—simply multiplying the matrices corresponds to composing the linear maps. For this reason, we often denote composition of linear maps simply as if it was multiplication. Observe that the proof of the following theorem relies on associativity of matrix multiplication!

Theorem 1.24. Say that $u_1,\dots,u_n$, $v_1,\dots,v_m$, and $w_1,\dots,w_l$ are the bases for U,V,W. Let T be a linear map from U to V and S a linear map from V to W. Let ϕ_{UV} be the isomorphism $\mathrm{Hom}(U,V)\to M_{m\times n}(F)$ using the above bases, and similarly let ϕ_{VW} be the isomorphism $\mathrm{Hom}(V,W)\to M_{l\times m}(F)$. Then $\phi_{VW}(S)\cdot\phi_{UV}(T)=\phi_{UW}(S\circ T)$, where the former $\cdot$ is matrix multiplication and the latter $\circ$ is the (abstract) composition of linear maps.

Proof. By Theorem 1.22, $\phi_{UV}(T)\phi_U(u)=\phi_V(T(u))$. Similarly, $\phi_{VW}(S)\phi_V(T(u))=\phi_W((S\circ T)(u))$ Substituting the first equation in the second, we get that $\phi_{VW}(S)\phi_{UV}(T)\phi_U(u)=\phi_W((S\circ T)(u))$. Remember that $\phi_{VW}(S)\phi_{UV}(T)$ here is a matrix, and we've just shown that it acts just like $\phi_{UW}(S\circ T)$ should. After all, this latter matrix also satisfies the equation $\phi_{UW}(S\circ T)\phi_U(u)=\phi_W((S\circ T)(u))$. Since ϕ_U is an isomorphism, every vector in F^n equals $\phi_U(u)$ for some $u\in U$, which means that the two matrices $\phi_{VW}(S)\phi_{UV}(T)$ and $\phi_{UW}(S\circ T)$ act the same on every column vector in F^n. Hence, they are equal. □

Problem 1.10 (2 Points)**.** Complete the above proof by showing that if $A,B\in M_{m\times n}(F)$ and $Av=Bv$ for all $v\in F^n$ then $A=B$.

The above theorem basically says that if matrices act on vectors like linear maps do, then that implies that matrices multiplying together is like function composition. After all, a matrix multiplying with a matrix is like the left matrix multiplying n columns vectors on the right, so it makes sense they are roughly equivalent. Note that it also proves that the inverse of a linear map with respect to a basis is the matrix inverse! Here is a formal definition of matrix inverse:

Definition 1.25. Let $n\geq 1$ be a positive integer. The $n\times n$ *identity matrix*, (usually) denoted by I, is the $n\times n$ matrix $\begin{pmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & 1\end{pmatrix}$. Namely, every entry on the "main" diagonal is a 1, while every off-diagonal entry of I is a 0.

Let $A\in M_{n\times n}(F)$. Then, we say that A is *invertible* if there is some $B\in M_{n\times n}(F)$ such that $AB=BA=I$, where I is the $n\times n$ identity matrix. Such B is usually denoted A^{-1}.

Sometimes I is used to denote the identity linear map $I:V\to V$. This abuse of notation is generally OK, because the matrix of the the identity map is exactly the $n\times n$ identity matrix *regardless of choice of basis of* V!

Using Theorems 1.22 and Theorem 1.24, it is not hard to see that $A\in M_{n\times n}(F)$ is invertible if and only if it is the matrix of an invertible linear map $T:V\to V$ (where $\dim V=n$) with respect to any basis of V.

Problem 1.11 (3 Points)**.** Here we consider a special class of linear maps from $\mathbb{R}^2$ to $\mathbb{R}^2$, rotations! They can be expressed in the following form:

$$R_\theta\left(\begin{bmatrix}a\\b\end{bmatrix}\right)=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}\begin{bmatrix}a\\b\end{bmatrix}$$

Convince yourself that this is the same as saying that if we choose the basis of the input and output space to be the standard basis, then $\phi(R_\theta)=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$.

(a) (2 Points) Prove that $R_{\theta_1}\circ R_{\theta_2}=R_{\theta_1+\theta_2}$, and $R_{-\theta}=R_\theta^{-1}$, where the $\circ$ is composition of functions.

(b) (1 Point) Verify that R_θ acts as a rotation by θ on the vector $\begin{bmatrix}1\\0\end{bmatrix}$, for any θ.

1.3.3 Change of Basis

Say we have some finite-dimensional vector space V and two bases $v_1,\dots,v_n$ and $w_1,\dots,w_n$. Say we have a vector $(a_1,\dots,a_n)$ with respect to the first basis. How could we transform it to the second? If we say that $v_i=a_{1i}w_1+\cdots+a_{ni}w_n$, then the matrix

$$B=\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}$$

maps the ith standard basis vector to $(a_{1i},\dots,a_{ni})$. This precisely converts the basis vectors in the $v_1,\dots,v_n$ basis to the $w_1,\dots,w_n$ basis! By linearity, this then extends to express any $v=b_1v_1+\cdots+b_nv_n$ in terms of $w_1,\dots,w_n$. So, B is the matrix that changes basis from $v_1,\dots,v_n$ to $w_1,\dots,w_n$. To convert in the other direction, you can simply use the matrix B^{-1}.

Problem 1.12 (3 Points)**.** Find the change of basis matrix from $\{(1,0),(0,1)\}$ to $\{(2,3),(1,4)\}$ in $\mathbb{R}^2$. *Hint: This direction may be trickier than the reverse direction. You may then use the fact that*

$$\begin{bmatrix}a & b\\ c & d\end{bmatrix}\begin{bmatrix}d & -b\\ -c & a\end{bmatrix}=\begin{bmatrix}ad-bc & 0\\ 0 & ad-bc\end{bmatrix}.$$

Now express $(6,5)$ *as a linear combination of* $(2,3)$ *and* $(1,4)$.

Now, say we have a linear map $T:V\to W$, and we write it as the matrix M with input basis $v_1,\dots,v_n$ and output basis $w_1,\dots,w_m$. Say we would like to change the output basis to $u_1,\dots,u_m$. Using the basis change matrix above, it is not too bad! Say that B is the matrix that transforms a vector in the $w_1,\dots,w_m$ basis to $u_1,\dots,u_m$. Then, BM will take a vector of coefficients with respect to v basis, act on it with T, and then convert it to the u basis. This is precisely matrix for T from the v basis to the u basis.

Now, say we would like to take our matrix M that has input basis v and output basis w. Then, let B be a basis change matrix from some u' basis to the v basis. Now, MB is the new matrix with respect to the u' basis and w basis.

Specifically in the case where T maps a vector space to itself, we see that if T is an operator on V with respect to the $v_1,\dots,v_n$ basis, and B is the basis change matrix from v to w, *then BMB^{-1} is the new matrix for T with respect to the new basis*!

1.4 Eigenvalues and Eigenvectors

We now focus on linear maps from a vector space to itself, sometimes called linear operators. In this context, unless otherwise specified, when a linear map is written as a matrix, the input and output bases are the same. There are two types of actions on vectors that we have introduced—multiplication by scalars, and acting on the vector by a linear map. Sometimes, these two actions overlap heavily. For instance, there exist scalar linear maps $cI : v \mapsto cv$ (here I is the identity map, which is linear), which act on vectors exactly the same as scalar multiplication. Considering linear maps in $\operatorname{Hom}(V,V)$ with respect to the some basis $\{v_1,\dots,v_n\}$ (the same basis on the input and output space), there is a larger class of matrices that act a lot like scalars. Consider linear maps S such that their matrix representations are:

$$\phi(S) = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

For such a linear map, we have that $Sv_i = \lambda_i v_i$, so it acts like a scalar on every basis vector (though potentially a different scalar on each). The matrix $\phi(S)$ being diagonal makes it much simpler to work with. For instance, $\phi(S^n)$ is just:

$$\phi(S^n) = \begin{bmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^n \end{bmatrix}$$

However, in general, exponentiating a matrix is far more tedious - you have to repeatedly do messy matrix multiplication. This idea motivates a question: for every linear map, does there exist a basis with respect to which it is diagonal? In a sense, this would be a key to understanding a linear map, a distinguished basis that makes the map far easier to understand.

Definition 1.26. Let T be a linear map from V to V. A vector $v \neq \mathbf{0}$ is called an eigenvector of T if $T(v) = \lambda v$ for some $\lambda \in F$. Then λ is called the corresponding eigenvalue to v.

An eigenvector, eigenvalue pair is a step in the right direction - after all, if S is diagonal with respect to a basis, then each basis vector is an eigenvector, and the values along the diagonal are the corresponding eigenvalues. For a given eigenvalue λ, the set of all vectors such that $T(v) = \lambda v$ (including 0) is a subspace of V, since it is closed under addition and scalar multiplication. This subspace is called the *eigenspace* of λ.

Theorem 1.27. λ is an eigenvalue of $T : V \to V$ if and only if $T - \lambda I$ has a nontrivial kernel.

Proof. If v is an eigenvector of T with eigenvalue λ, then $T(v) = \lambda v = \lambda I v$. This implies that $(T - \lambda I)v = T(v) - \lambda I v = \mathbf{0}$. Since $v \neq \mathbf{0}$, this implies $T - \lambda I$ has a nontrivial kernel.

Conversely, if $T - \lambda I$ has a nontrivial kernel, then there exists a $v \neq \mathbf{0}$ such that $(T - \lambda I)v = \mathbf{0}$, implying that $T(v) = \lambda I v = \lambda v$. □

Theorem 1.28. Let T be a linear map from V to V. Say $V_1, \dots, V_m$ are the eigenspaces for distinct $\lambda_1, \dots, \lambda_m$. Then $V_1 + \cdots + V_m$ is a direct sum.

Proof. This result is vacuously true for $m = 1$. Assume by induction that the result is true for $m-1$. Now, if there are two ways of writing an element $v \in V_1 + \cdots + V_m$, then there is some $v_i \in V_i$ where $v_1 + \cdots + v_m = \mathbf{0}$, and at least one of the v_i is nonzero, which we will assume is not v_1. Then we have that $\mathbf{0} = T(v_1 + \cdots + v_m) = \lambda_1 v_1 + \cdots + \lambda_m v_m$. Now subtracting $\lambda_1 \mathbf{0} = \lambda_1 v_1 + \cdots + \lambda_1 v_m$, we find that $(\lambda_2 - \lambda_1)v_2 + \cdots + (\lambda_m - \lambda_1)v_m = \mathbf{0}$. This is another sum of values in $V_2, \dots, V_m$, at least one of which is nonzero. By induction, however, this is impossible! □

It turns out that the number of distinct eigenvalues is actually at most the dimension of the vector space.

Theorem 1.29. Let $T : V \to V$ be a linear map, where V has dimension n. T has at most n eigenvalues.

Proof. Assume by contradiction that there are more than n distinct eigenvalues. Then, we know that the first n eigenvalues correspond to eigenvectors $v_1, \dots, v_n$. Since $\operatorname{span}(v_i)$ is a subspace of the eigenspace for λ_i, we have that $\operatorname{span}(v_1, \dots, v_{n+1}) = \operatorname{span}(v_1) \oplus \cdots \oplus \operatorname{span}(v_{n+1})$ (an exercise shows that if the sum of a collection of vector spaces is a direct sum, then a sum of a collection of their subspaces is also a direct sum). This implies, as mentioned above, that $v_1, \dots, v_{n+1}$ are linearly independent. However, this is impossible, since it is longer than the dimension, so clearly cannot be extended to a basis. □

Problem 1.13 (3 Points)**.** Prove that if V is a vector space, $V_1, \dots, V_m \subseteq V$ are subspaces, and $U_i \subseteq V_i$ are subspaces, then

$$V_1 \oplus \cdots \oplus V_n \implies U_1 \oplus \cdots \oplus U_n.$$

Note: this notation for direct sum, while somewhat opaque, is standard. This problem, translated, reads: If the sum of V_i is a direct sum, then the sum of U_i is a direct sum.

Problem 1.14 (2 Points)**.** Prove that a linear map $T : V \to V$ has a basis of eigenvectors $v_1, \dots, v_n$ if and only if T has a basis with respect to which it is diagonal.

Definition 1.30. The **minimal polynomial** of a linear map T is a polynomial $p(x)$ with coefficients in F such that $p(T) = 0$, p is monic (the leading coefficient is 1) and p is the polynomial of minimum degree with this property. Remember that multiplication of linear maps is composition.

We call $p(x)$ "the" minimal polynomial because there can only be one such polynomial. If there were two, then their difference would be a polynomial of smaller degree for which T evaluates to zero.

Theorem 1.31. Every linear operator T from on a finite-dimensional vector space V (that is, $T : V \to V$) of dimension n has a minimal polynomial.

Proof. Consider the linear maps $I, T, \dots, T^{n^2}$. They correspond to $n^2 + 1$ vectors in the space $\operatorname{Hom}(V,V)$, and therefore they must be linearly dependent (since the dimension of $\operatorname{Hom}(V,V)$ as a vector space is n^2). So, we have some $a_0, \dots, a_{n^2}$ such that $a_{n^2} T^{n^2} + \cdots + a_0 I = 0$. Then, we know that $p(T) = 0$, where $p(x) = a_{n^2} x^{n^2} + \cdots + a_0$. The existence of one polynomial that T satisfies shows that there must be a polynomial of minimal degree, the minimal polynomial. □

Theorem 1.32. A value $\lambda \in F$ is a root of the minimal polynomial $p(x)$ if and only if λ is an eigenvalue of T.

Proof. Assume by contradiction λ is a root of the polynomial $p(x)$ and λ is not an eigenvalue. Because λ is a root, $p(x)$ is divisible by $x-\lambda$. So, $p(T)$ can be written as $(T-\lambda I)q(T)$. Since λ is not an eigenvalue, then $(T-\lambda I)$ is injective. If $q(T) \neq 0$, then $q(T)v \neq \mathbf{0}$, which implies that $(T-\lambda I)q(T)v \neq \mathbf{0}$. So, $q(T) \neq \mathbf{0}$ implies that $p(T) \neq 0$. We then have that since $p(T)=0$, $q(T)=0$. So, $p(T)$ is not the minimal polynomial, a contradiction.

Conversely, assume that λ is an eigenvalue of T. Then, $T-\lambda I$ has a nontrivial nullspace - say that $(T-\lambda I)v = \mathbf{0}$ with $v \neq \mathbf{0}$. We can, by polynomial division, write $p(x) = q(x)(x-\lambda)+r$, where $r \in F$. Then, $p(T)v = q(T)(T-\lambda I)v + rv = rv$. So, we must have that $r=0$, implying that $p(\lambda)=0$. □

Now, assume that $F = \mathbb{C}$, so we can leverage the following property, the *fundamental theorem of algebra*: any polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$. From this theorem, we can immediately factor the minimal polynomial, and the roots of the minimal polynomial are the eigenvalues.

A polynomial is said to be separable if it has no repeated roots. Then, we have a convenient characterization of diagonalizability. First, a lemma for the result.

Lemma 1.33. Say that $\lambda_1, \dots, \lambda_n$ are distinct values. Then consider the polynomials $p_i(x) = (x-\lambda_1)\dots(x-\lambda_{i-1})(x-\lambda_{i+1})\dots(x-\lambda_n)$. There exists a linear combination $\sum_i a_i p_i(x) = 1$.

Proof. We proceed by induction. For the base case $n=1$, $p_1(x)=1$, so the statement is clearly true. Now, assume the statement is true for n. We prove it for $n+1$. Note that $\frac{1}{\lambda_n-\lambda_1}(p_{n+1}(x)-p_i(x)) = (x-\lambda_1)(x-\lambda_2)\dots(x-\lambda_{i-1})(x-\lambda_{i+1})\dots(x-\lambda_n)$. So these terms enumerate the p_i for the n case, and we know a linear combination of these terms yields 1. So, a linear combination of p_1 through p_{n+1} can yield 1 as well, once the coefficients are expanded! □

Theorem 1.34. The minimal polynomial of T is separable if and only if T is diagonalizable.

Proof. First, say T is diagonalizable. Say ϕ is the map that takes T to its matrix in its diagonal basis. Then, consider $\phi(\Pi_i(T-\lambda_i I))$, where each distinct eigenvalue is counted once. Every term along the diagonal of this matrix is seen to be zero.

So, $\phi(\Pi_i(T-\lambda_i I)) = \mathbf{0}$, implying that for $p(x) = \Pi_i(x-\lambda_i)$, $p(T)=0$. Since we proved that each $x-\lambda_i$ must also be a factor of the minimal polynomial, this proves that $p(x)$, the minimal polynomial, is separable!

Say that $p(x) = \Pi_i(x-\lambda_i)$ is the minimal polynomial, where each λ_i is distinct (so it is separable). Then we would like to prove that if $V_1, \dots, V_n$ are the eigenspaces for $\lambda_1, \dots, \lambda_n$, then $V_1 \oplus \cdots \oplus V_n = V$. Then, by 1.5, we would have that T is diagonalizable. To prove it, take a vector $v \in V$. Then, note that we can write as in the above lemma $1 = \sum_i a_i p_i(x)$. Then, $v = Iv = \sum_i a_i p_i(T)v$, and note that since $(T-\lambda_i I)p_i(T) = p(T)$, $(T-\lambda_i I)(a_i p_i(T)v) = a_i p(T)v = \mathbf{0}$. So, since $a_i p_i(T)v$ is in the null space of $T-\lambda I$, $a_i p_i(T)v \in V_i$, so we have expressed an arbitrary element in V as a sum of the V_i. □

Using this theorem, we can see that there exist linear maps that cannot possibly be diagonalized. Consider the linear map $T : \mathbb{C}^2 \to \mathbb{C}^2$ whose matrix with respect to the standard basis is:

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

The minimal polynomial of this matrix clearly cannot be degree 1, since it is not a multiple of the identity, and we can calculate a degree 2 polynomial (which is therefore the minimal polynomial) to be $p(x) = x^2 - 2x + 1 = (x-1)^2$. This polynomial is not separable, so this linear map cannot be diagonalizable.

1.5 Inner Products

In $\mathbb{R}^2$, there are many possible bases. As you can verify, $\{(1,0),(0,1)\}$, $\{(1,0),(1,1)\}$, and $\{(2,1),(1,1)\}$ are all bases of $\mathbb{R}^2$. However, when we think intuitively of $\mathbb{R}^2$, one basis seems special: the standard basis. Or at least, when we draw axes for $\mathbb{R}^2$, we ensure that they are perpendicular. In this section, we introduce a structure on vector spaces, making them “inner product” spaces, such that notions like perpendicular make sense. In order to even understand “orthogonality,” we need our objects to have a notion of angles. In this section, we assume that F is either $\mathbb{R}$ or $\mathbb{C}$, and use a^* to denote the complex conjugate of a.

Definition 1.35 (Inner Products)**.** An inner product $\langle \cdot, \cdot \rangle$ on a vector space V is an operation that takes two vectors in V and outputs a scalar in F. It obeys the following properties:

1. Positivity: $\forall v \in V \, \langle v, v \rangle$ is a nonnegative real number.
2. Definiteness: $\langle v, v \rangle = 0$ if and only if $v = \mathbf{0}$.
3. Additivity: $\langle v_1, w \rangle + \langle v_2, w \rangle = \langle v_1 + v_2, w \rangle$ and $\langle v, w_1 \rangle + \langle v, w_2 \rangle = \langle v, w_1 + w_2 \rangle$ for all $v, w \in V$.
4. Homogeneity: $\langle cv, w \rangle = c\langle v, w \rangle$ and $\langle v, cw \rangle = c^* \langle v, w \rangle$
5. Conjugate Symmetry: $\langle v, w \rangle = \langle w, v \rangle^*$.

Example 1.6. $\mathbb{R}^n$ and $\mathbb{C}^n$. The standard inner product on $\mathbb{R}^n$ is the dot product: we define $\langle (a_1, \dots, a_n), (b_1, \dots, b_n) \rangle = \sum_i a_i b_i$. This dot product is equal to zero precisely when two vectors are orthogonal, and $\langle v, v \rangle = \sum_i a_i^2$ is the length of a vector. It is clear that this latter property implies positivity and definiteness, and the other properties can be proven quickly as well.

For $\mathbb{C}^n$, we define something that is almost the dot product: define $\langle (a_1, \dots, a_n), (b_1, \dots, b_n) \rangle = \sum_i a_i b_i^*$. Why add the conjugate? Well then, for $\langle (a_1, \dots, a_n), (a_1, \dots, a_n) \rangle = \sum_i a_i a_i^*$. Since $a_i a_i^*$ is a nonnegative real number, the dot product satisfies positivity. This is why conjugation is necessary, because the magnitude of a complex number is not a^2 but aa^*. As a direct result of this, we obtain the conjugate symmetry in the definition of inner products, rather than plain old symmetry.

Definition 1.36. An orthonormal set of vectors is a set of vectors such that every pair v_i, v_j is orthogonal ($\langle v_i, v_j \rangle = 0$), and every vector is normal ($\langle v_i, v_i \rangle = 1$). An orthonormal basis is a basis which is also orthonormal.

Note that any orthogonal set of vectors can easily be made normal, by taking each vector v_i and redefining it as $v_i' = \frac{v_i}{\sqrt{\langle v_i, v_i \rangle}}$. Then, $\langle v_i', v_j' \rangle = \frac{1}{\sqrt{\langle v_i, v_i \rangle \langle v_j, v_j \rangle}} \langle v_i, v_j \rangle = 0$, $\langle v_i', v_i' \rangle = \left\langle \frac{v_i}{\sqrt{\langle v_i, v_i \rangle}}, \frac{v_i}{\sqrt{\langle v_i, v_i \rangle}} \right\rangle = \frac{1}{\sqrt{\langle v_i, v_i \rangle}^2} \langle v_i, v_i \rangle = 1$.

Theorem 1.37. An orthogonal set of vectors $v_1, \dots, v_n$ is linearly independent.

Proof. Say that $0 = a_1 v_1 + \dots + a_n v_n$. Taking the inner product on either side with v_i, we get that $0 = \langle 0, v_i \rangle = \langle a_1 v_1 + \dots + a_n v_n, v_i \rangle = \sum_j a_j \langle v_j, v_i \rangle = a_i$. So, every $a_i = 0$, proving that $v_1, \dots, v_n$ must be linearly independent. □

Theorem 1.38 (Gram-Schmidt Procedure)**.** For any basis $v_1, \dots, v_n$, there exists an orthonormal $w_1, \dots, w_n$ such that $\operatorname{span}(v_1, \dots, v_i) = \operatorname{span}(w_1, \dots, w_i)$ for every $1 \leq i \leq n$.

Proof. We iteratively define $w_1, \dots, w_n$, such that the subsequences $w_1, \dots, w_i$ are orthonormal and the same span as $v_1, \dots, v_i$. Firstly, $w_1 = \frac{v_1}{\sqrt{\langle v_1, v_1 \rangle}}$ is a normalized vector, and so is orthonormal. Now, define inductively $w_i' = v_i - \langle v_i, w_1 \rangle w_1 - \cdots - \langle v_i, w_{i-1} \rangle w_{i-1}$. $w_1, \dots, w_{i-1}, v_i$ and $w_1, \dots, w_{i-1}, w_i'$ have the same span since v_i is in the latter span and w_i' is in the former span. This proves that $v_1, \dots, v_i$ and $w_1, \dots, w_{i-1}, w_i'$.

Now, note that $\langle w_i', w_j \rangle$, when we expand the definition of w_i' and expand by linearity, gives $\langle v_i, w_j \rangle - \langle v_i, w_j \rangle \langle w_j, w_j \rangle = 0$. Now since, w_i' is nonzero, we can define $w_i = \frac{w_i'}{\sqrt{\langle w_i', w_i' \rangle}}$ to give the new orthonormal basis vector. It satisfies all the conditions necessary, and completes this iterative step. □

Problem 1.15 (7 Points)**.**

(a) (2 Points) Let $v_1, \dots, v_n$ be an orthonormal basis for V. For $v \in V$, show that $v = \sum_i \langle v, v_i \rangle v_i$. Orthonormal bases are very well behaved!

(b) (1 Point) Let $v_1, \dots, v_n$ be an orthonormal basis for V. If $v = a_1 v_1 + \cdots + a_n v_n$, prove that $\langle v, v \rangle = a_1 a_1^* + \cdots + a_n a_n^*$ (this is the Pythagorean theorem).

(c) (4 Points) Say B is the basis change matrix from an orthonormal basis $v_1, \dots, v_n$ in V to another orthonormal basis $w_1, \dots, w_n$ in V. Prove that B's rows are orthonormal vectors in F^n and B's columns are orthonormal vectors in F^n, where the inner product on F^n is as defined in Example 1.6.

2 Group Theory

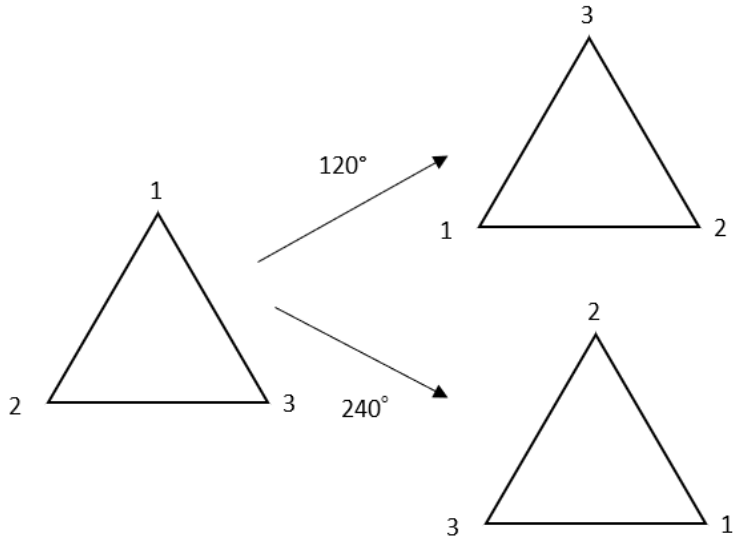
We will now be introducing ourselves to the world of *group theory*. Groups are the abstract way in which mathematicians describe symmetries of arbitrary objects. The birth of the concept came from a French mathematician named Evariste Galois who wanted to understand polynomial equations. He made observations such as the following. Suppose $3+2i$ ($i=\sqrt{-1}$) is a solution to some polynomial equation $f(x)=0$, where $f(x)$ is a polynomial with real coefficients. Then, one may deduce that $3-2i$ is *another* solution to $f(x)$ (you probably know how to prove this already from high school algebra!). Galois tried understanding this phenomenon of transforming known solutions into new ones and exploited these properties to understand polynomials. It was many years before this idea was molded into what is presently known as a group, but group theory has since become one of the deepest and most prevalent parts of modern mathematics.

We will not be dealing so much with how groups deal with polynomial equations, but instead arrive at the notion of a group through a separate route of motivation that is geometrically inspired. We will start this journey by studying group actions, which amount to the transformations of an object or how we can change an object without altering its main properties. We will then see that seemingly different objects actually have the same set of transformations. This will lead us into describing a transformation abstractly, and lead us to the definition of a group.

In this section, Problems 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.14 do not require proof. All other Problems require proof.

2.1 The Triangle

We begin our exploration by studying arguably the simplest geometric object: the equilateral triangle. The equilateral triangle is pretty "symmetric," for instance, you can rotate an equilateral triangle by 120° around its center and you get back an equilateral triangle. This doesn't work for scalene triangles though, so there's something special to explore. Let's try and quantify the different symmetries of the triangle. Let's start by labelling the vertices of the equilateral triangle 1, 2, 3 like this: $\overset{1}{\underset{2\ \ 3}{\triangle}}$. This is also shown by the leftmost triangle in the diagram below. We call this the *original* or *initial* labelling of the triangle's vertices. When we rotate counterclockwise by 120°, the shape we get is still a triangle, but it's labeled differently. Similarly, if we rotate the original triangle counterclockwise by 240°, we end up with yet another labelling.



Are these all the symmetries of our triangle? What if we rotated clockwise instead?

Problem 2.1 (1 Point)**.**
(a) Draw the labelling on the triangle that results when the original triangle is rotated *clockwise* by 120°.
(b) Do the same for a clockwise 240° rotation.

As you can see, we don't get anything new when we rotate clockwise, they're just duplicates of our old rotations. We will now call the labelling of our triangle rotated 120° counterclockwise as R_{120} and 240° counterclockwise as R_{240}. Note that going forward, whenever we say rotation, we always mean counterclockwise, it's just what mathematicians have grown to prefer.

We have not yet completed the list of symmetries, however, since we can also "flip" our triangles:

Problem 2.2 (1 Point)**.**
(a) Draw the labelling on the triangle that results when we flip the original triangle across the perpendicular bisector that passes through the point labelled 1 (in the initial labelling $\overset{1}{\underset{2\ \ 3}{\triangle}}$).
(b) Do the same for the resulting triangle when we flip the original triangle across the perpendicular bisectors that pass through the points labelled 2 and 3.

These labellings are different than what we had before. We will call these new labellings F_A, F_B, and F_C, where F_A corresponds to flipping across the vertical perpendicular bisector, F_B corresponds to the perpendicular bisector which bisects the right side of the triangle, and F_C the perpendicular bisector which bisects the left side of the triangle.

Now, we have constructed 5 new labellings starting from our original one: $R_{120}, R_{240}, F_A, F_B, F_C$. Can we transform our triangles in other ways? Up to now, we have only applied single transformations to our triangle, what if we do multiple in sequence? For example, if we rotate the original triangle by 120°, we get R_{120}, but if we rotate R_{120} by 120° again, we get R_{240}. Of course, rotating twice by 120° just rotates by 240°. But, we can do more interesting sequences of transformations, like what if we rotate and then flip?

Problem 2.3 (1 Point)**.** Draw the resulting labelling of the triangle after applying the following sequences of transformations:

(a) A 120° counterclockwise rotation followed by a flip across the vertical perpendicular bisector.
(b) A flip across the vertical perpendicular bisector followed by a 120° counterclockwise rotation.
(c) A 120° counterclockwise rotation followed by a flip across the vertical perpendicular bisector followed by a 240° counterclockwise rotation.

As you can see, we never get any labellings that we haven't gotten before. However, we do sometimes apply multiple transformations and get back to where we started, so we should give our initial labelling a special name as well, let's call it I for "initial." It's the labelling that results when we "do nothing" to our original triangle.

Although we the process of doing multiple transformations hasn't given us any new symmetries, it does tell us that the symmetries interact with one another. To put it more formally, we shouldn't think about our labellings as just labellings, but rather as the *transformations* we did to get there. For instance, F_A now doesn't represent the labelling of the triangle's vertices after flipping, but instead the **action** of flipping the triangle itself! From this point of view, it makes sense to compose such transformations. For example, performing F_A and then F_B gives us the same labelling as R_{240}. We write this as $R_{240} = F_B \circ F_A$. The reason for the seemingly backwards notation is that it's really function composition. We can think of F_A, F_B as functions, and we may write stuff such as $F_A(\triangle)$ to show that F_A is acting on a triangle. Then, we're saying $R_{240}(\triangle) = F_B(F_A(\triangle))$.

Problem 2.4 (1 Point)**.** Write the following compositions of transformations as a single transformation

(a) $F_B \circ F_C$.
(b) $F_C \circ R_{120} \circ F_A$.
(c) $F_A \circ R_{120} \circ I$.

Now, what if we wanted to compose 5 transformations and see the result? It would be tedious to do manually with a triangle, but luckily mathematicians have found a clever way around this by treating function composition as a sort of multiplication on the transformations. To see how this is used, we'll start by constructing a multiplication table, and then use that to compose many transformations. A word of caution: as the examples above showed, this multiplication is *not commutative* unlike normal multiplication, in other words, the order in which functions are composed matters. For instance, notice that $F_A \circ R_{120} = F_B$, whereas $R_{120} \circ F_A = F_C$.

Problem 2.5 (3 Points)**.** Complete the multiplication table. If the column has transformation S and the row has transformation T, the entry in that box should be $T \circ S$. No proof is needed.

	I	R_{120}	R_{240}	F_A	F_B	F_C
I						
R_{120}		R_{240}		F_C		
R_{240}						
F_A						R_{240}
F_B					I	
F_C						

Problem 2.6 (1 Point)**.** Write the following composition of transformations as a single transformation It may be helpful to use the table from the previous problem.

$$R_{120} \circ F_A \circ F_B \circ R_{240} \circ F_B \circ R_{120}.$$

An interesting thing to note is that even though we physically perform a sequence of transformations starting from the rightmost and working left, it actually doesn't matter what order we compose in. This is because transformations are functions, and function composition is *associative*: the multiplication operations can be carried out in any order. This can simplify many things, for example, $R_{240} \circ (R_{120} \circ F_A) = (R_{240} \circ R_{120}) \circ F_A = I \circ F_A = F_A$ (you should do the calculation the other way and verify it's still F_A).

An aside: in abstract algebra, and other branches of higher mathematics, *associativity* is actually a much more natural condition than *commutativity* when defining mathematical structures.

We now have a more or less complete understanding of the triangle: we know all possible transformations, and we can easily compute what will happen if we compose them in different orders.

2.2 Other Polygons

Now we analyze the square □, which is also a very symmetric shape. There are the $90°, 180°$, and $270°$ rotational symmetries given by R_{90}, R_{180}, R_{270} (remember, we always rotate counter-clockwise). We can also flip the square, either vertically, horizontally, or across the diagonals. These are denoted as F_A for flipping across the vertical perpendicular bisector, F_B for flipping across the horizontal perpendicular bisector, F_C for flipping across the diagonal connecting the bottom left and top right vertices, and F_D for flipping across the diagonal connecting the top left and bottom right vertices. There is also I, the "do nothing" symmetry. Let's try and understand how these compose. Here is a partially completed 8×8 *multiplication table* for the symmetries of the square, with the same conventions as Problem 2.5. If you would like, you can try filling in some more cells to this multiplication table.

	I	R_{90}	R_{180}	R_{270}	F_A	F_B	F_C	F_D
I		R_{90}	R_{180}	R_{270}				
R_{90}			R_{270}					
R_{180}						F_A		
R_{270}				R_{180}				
F_A	F_A							
F_B	F_B			F_C		I		
F_C					R_{270}			
F_D								I

We'll take some time to extract important patterns that have shown up in the multiplication tables of Problems 2.5 and the one above. Notice how every row and every column contains exactly one I. The interpretation of this is that every transformation has an "opposite" or *inverse* transformation such that composing the two will result in doing nothing.

Problem 2.7 (3 Points)**.** Give a general rule for the inverse for rotations R_* and flips F_*.

Another important thing to notice is that the row to the right of I is the same as the input row above it, and the column right below I is the same as the input column to its left. Moreover, this doesn't happen for any other row or column, they all change the inputs. The element I has a very important role, it's analogous to multiplying by the number 1, and inverse transformations are analogous to multiplicative inverses. This element I will be called the *identity element*. Also notice how every transformation appears in every row exactly once, and every transformation appears in every column exactly once.

Let's see how the ideas we developed for the equilateral triangle and the square may be generalized to arbitrary regular polygons. Let P_n denote the regular polygon with n sides, with $n \geq 3$. Take a copy of the regular n-gon P_n in $\mathbb{R}^3$ such that its circumcenter lies at the origin of $\mathbb{R}^3$. A *symmetry transformation* of the regular n-gon is a bijective[1] map of P_n onto itself which is effected by a rotation of $\mathbb{R}^3$ with respect to an axis passing through $\mathbf{0} \in \mathbb{R}^3$. This definition clearly does not depend on the "embedding" of P_n in $\mathbb{R}^3$.

Problem 2.8 (3 Points)**.**

(a) (2 Points) Show that there are $2n$ symmetry transformations for the regular polygon P_n.

(b) (1 Point) Label the vertices of P_n with the vertex labeled 2 directly counterclockwise to the vertex labeled 1. Show that there are exactly n symmetry transformations T of P_n in which, upon applying T to P_n and considering the resulting labelling of the vertices of P_n, 2 is still counterclockwise to 1. Thus, there are exactly n symmetry transformations F of P_n in which, upon applying F to P_n, 2 is clockwise to 1 in the new labelling.

Any symmetry transformation of P_n which gives rise to a new labelling with 2 counterclockwise to 1 is called *orientation preserving*, whereas any symmetry transformation of P_n which gives rise to a new labelling with 2 clockwise to 1 is called *orientation reversing*.

Definition 2.1. The collection of transformations of a regular n-gon will be denoted D_n, the *dihedral group* with $2n$ elements. Transformations which preserve orientations are called *rotations*, and transformations which reverse orientations are called *reflections* or *flips*.

In particular, under this definition, the identity transformation I of the regular n-gon is a rotation.

In the following problem, the identity (do nothing transformation) is denoted I and T denotes an arbitrary transformation in D_n:

Problem 2.9 (3 Points)**.** Prove the following

(a) (1 Point) Every element in D_n has an inverse, that is for every $T \in D_n$, there is some $T' \in D_n$ so that $T' \circ T = I$.

(b) (1 Point) For $T \in D_n$, show that also $T \circ T' = I$.

(c) (1 Point) In the "multiplication table" for D_n, show that every element appears in each row and each column exactly once.

The inverse of a transformation T will henceforth be denoted T^{-1}. You may have noticed that there are some "simple" transformations which you can apply repeatedly to get more complicated transformations. Let's formalize this.

[1] Recall that a map $f : S \to T$ is called *injective* if for any $x, y \in S$, $f(x) = f(y)$ implies that $x = y$, is called *surjective* if the image of f, i.e. the set $\{f(x) \mid x \in S\}$, is all of T, and is called *bijective* if it is injective and surjective.

Problem 2.10 (3 Points)**.** Show that there are two transformations $R, F \in D_n$ where R is a rotation and F is a flip such that for any $T \in D_n$, we can write $T = F^j \circ R^i$ for some $0 \leq i \leq n-1,\ j \in \{0,1\}$, where $R^i = R \circ R \circ \cdots \circ R$ with i total compositions. Furthermore, show that T is a flip if and only if $j = 1$.
Is there only one possibility for the initial choice of transformations R and F?

We remarked earlier that multiplications (compositions) do not generally commute. Let's try and understand better when they do.

Problem 2.11 (3 Points)**.** All of the transformations that follow will be in D_n.

(a) (1 point) Show that for two rotations R and R', $R \circ R' = R' \circ R$.

(b) (1 point) Show in general that for a rotation R and a flip F, $F \circ R \circ F = R^{-1}$. When is $R \circ F = F \circ R$?

(c) (1 point) If F and F' are two flips, when is $F \circ F' = F' \circ F$? (Hint: use the previous part and the previous problem)

In this last part, we can see that we get some utility out of an abstract way to represent transformations. We concluded something geometric (whether it matters what order you flip something) using algebraic manipulations!

Let C_n be the subset of all rotations in D_n. We call C_n the *rotation subgroup* of D_n. Observe that the set C_n is closed under the operations of composition and inverse. Now, for any flip $F \in D_n$, the subset $\{I, F\}$ of D_n is called a *reflection subgroup* of D_n. There are n such reflection subgroups. Each reflection subgroup is closed under the operations of composition and inverse. We will eventually formalize the idea of a subgroup.

2.3 Symmetric Groups

In the following section, we fix the following notation. For any positive integer n, set $\underline{n} = \{1, 2, \ldots, n\}$.

We now explore our next incarnation of transformations. Suppose you're playing the game where someone has 3 cups upside down with a dice underneath one and you need to guess which one has the dice after they're shuffled. The total number of cups hasn't changed, but they have been moved around, i.e. they have been **permuted**. We could try detecting this by labelling every cup with a number and seeing where each label ends up. This motivates the following.

Definition 2.2 (Permutations)**.** A *permutation* of $\underline{n} = \{1, 2, \ldots, n\}$ is a bijection $p : \underline{n} \to \underline{n}$. The collection of all such permutations will be denoted S_n, the *symmetric group on n elements*.

You can think of the expression $p(a) = b$ as meaning the object labeled a has been moved to where b used to be.

Problem 2.12 (2 Points)**.**

(a) (1 Point) Show that for permutations $p, q \in S_n$, $p \circ q$ is another permutation.

(b) (1 Point) Show that for any $p \in S_n$ and any $a \in \underline{n}$, there is some k so that $p^k(a) = a$. Here, p^k denotes the k-fold composition $p \circ p \circ \cdots \circ p$.

Definition 2.3 (Orbit)**.** For a given $a \in \underline{n}$ and $p \in S_n$, the set $\{p^n(a) \mid n \in \mathbb{Z}^{\geq 0}\}$ is denoted $\mathrm{Orb}_p(a)$ and called the *orbit* of a under p.

In other words, the orbit of an object is all the possible places a ends up by applying p to a. (note p^0 is the identity map, so that $p^0(a) = a$).

Problem 2.13 (3 Points)**.**

(a) (1 Point) Show that given a permutation p, for all $b \in \mathrm{Orb}_p(a)$, $\mathrm{Orb}_p(b) = \mathrm{Orb}_p(a)$.

(b) (2 Points) Show that for a given permutation $p \in S_n$, $\underline{n}$ can be broken up into disjoint sets $n = A_1 \cup A_2 \cup \cdots \cup A_k$ where a and b are in the same set if and only if $b \in \mathrm{Orb}_p(a)$. We call the sets A_i the orbits of $\underline{n}$ under p.

This problem gives us a very nice way to represent permutations. It would be very tedious if every time we wanted to talk about a permutation p we had to specify $p(1), p(2), \ldots, p(n)$. Instead, we can use what is called the cycle decomposition of a permutation:

Definition 2.4. Let $A = a_1, \ldots, a_m$ be an orbit for $p \in S_n$ such that $p(a_i) = a_{i+1}$ for $i = 1, 2, \ldots, m-1$ and $p(a_m) = a_1$. Then $C := (a_1\ a_2\ \cdots\ a_m)$ is called the *cycle* corresponding to A. We say that C is an m-cycle; here $m = |A|$ is the *length* of the cycle C. If $A_1, \ldots, A_k$ are the orbits for p, the *cycle decomposition* for p is denoted by the concatenation of the cycles $C_1, \ldots, C_k$ corresponding to the respective orbits $A_1, \ldots, A_k$. If the cycle $A_i = \{a\}$ is a singleton set, the corresponding cycle is sometimes omitted. The *cycle type* of p is the list of the cardinalities $|A_1|, |A_2|, \ldots |A_k|$, i.e. the list of the lengths of the cycles of p. The cycle type is usually listed in non-increasing order.

A *transposition* (or more informally, a *swap*) is the simplest type of non-trivial permutation whose cycle decomposition consists of just one cycle of length 2, and otherwise singletons. In other words, a transposition swaps two elements and leaves all other elements fixed.

We have seen how to recover the cycle decomposition of any permutation $p \in S_n$. Conversely, let $C_1, \ldots, C_k$ be mutually disjoint cycles with elements in $\underline{n}$. By mutually disjoint, we mean the following: if $C_i = (a_1\, a_2\ \cdots\ a_l)$, $C_j = (b_1\, b_2\ \cdots\ b_m)$ for some $1 \leq i < j \leq k$, then $a_s \neq b_t$ for all $1 \leq s \leq l, 1 \leq t \leq m$. Then, the *cycle product* $C_1 C_2 \ldots C_k$ defines a permutation in S_n. Sometimes we use the term *m-cycle* to refer to a permutation in S_n whose cycle decomposition consists of one cycle of length m and all other cycles of length 1. In particular, a 1-cycle is just the identity map $\mathrm{id} : \underline{n} \to \underline{n}$.

Example 2.1. The permutation $p = (1\ 3\ 6)(2\ 4) \in S_7$ is given explicitly by $p(1) = 3, p(2) = 4, p(3) = 6, p(4) = 2, p(5) = 5, p(6) = 1, p(7) = 7$. The cycle type of p is $3, 2, 1, 1$. Note that the cycle decomposition $(1\ 3\ 6)(2\ 4)$ could represent a valid element of S_n for any $n \geq 6$, and $(1\ 3\ 6)(2\ 4) \in S_n$ would fix any element $i \geq 6$.

As another example, the permutation $q = (2\ 3) \in S_4$ is a transposition, given explicitly by $q(1) = 1, q(2) = 3, q(3) = 2, q(4) = 4$.

Note that $(2\ 4)(3\ 6\ 1)$ is also a valid cycle decomposition of the permutation $p \in S_7$ defined in the above example. So is $(6\ 1\ 3)(4\ 2)$. This highlights the following property: the cycle decomposition of a permutation $p \in S_n$ is *unique* up to (1) *permuting cycles* and (2) *cyclically permuting the elements of each cycle*. The proof of this fact is omitted from the power round.

We shall now consider how to compose permutations. Cycles can be multiplied, where multiplication is just given by the usual function composition. Multiplication is typically denoted by concatenating cycles. In lieu with the usual convention for function composition, the convention is that the cycles on the right "act" first. For example, $(3\ 5)(1\ 2)(2\ 3)$ denotes the composition $(3\ 5) \circ (1\ 2) \circ (2\ 3)$ of the three transpositions $(2\ 3), (1\ 2)$, and $(3\ 5)$ (recall that function composition is associative). In particular, note that $(3\ 5)(1\ 2)(2\ 3)$ and $(1\ 2\ 5\ 3)$ are the same permutation of $\underline{n}$ $(n \geq 5)$. Now, let us see how permutations can be multiplied via cycle decomposition. Suppose $p \in S_n$ has cycle decomposition $C_1 C_2 \dots C_k$. Then, interpreting each cycle C_i as a permutation of $\overline{n}$, we have in fact $p = C_1 \circ C_2 \circ \cdots \circ C_k$ as elements of S_n. Equality here does not depend on the choice of cycle decomposition of p (i.e. does not depend on the order in which we write the cycles of p). In particular, if $q \in S_n$ is another permutation with cycle decomposition $C'_1 C'_2 \dots C'_l$, then the permutation $q \circ p \in S_n$ is the product of $k+l$ cycles $C'_1 C'_2 \dots C'_l C_1 C_2 \dots C_k$.

Example 2.2. Let $p = (1\ 3\ 6)(2\ 4), q_1 = (2\ 4), q_2 = (1\ 3\ 6)$, all permutations in S_7. Then, $p = q_2 \circ q_1 = q_1 \circ q_2$. Thus, $(1\,3\,6)(2\,4)$ may either be viewed as a permutation in S_7, or the product of the two permutations $(1\,3\,6), (2\,4)$ in S_7.

Another example: let $p = (2\,3\,6)(7\,1\,5\,8), q = (1\,3\,5\,6\,7\,8)(2\,4)$. Then, $q \circ p = (1\,6\,4\,2\,5)(3\,7)$.

Problem 2.14 (3 Points)**.** Simplify the following permutations into their cycle decomposition:

(a) (1 point) $(1\,3\,4)(2\,3\,1)$

(b) (1 point) $(1\,2\,4)(4\,5\,3)(5\,4\,3)(4\,2\,1)$

(c) (1 point) $\tau \circ \sigma$, where $\sigma = (6\,1\,3)(4\,2)$ and $\tau = (1\,2\,5\,6\,3\,7)$

There are some very close similarities in the structure of the permutations S_n with the set of transformations D_n explored earlier. Composition of permutations is associative. There is a "do nothing" permutation of $\underline{n}$, namely, the identity map $\mathrm{id} : \underline{n} \to \underline{n}$. Now we explore inverses. Recall that for any permutation $p \in S_n$ that there is an *inverse permutation* $p^{-1} \in S_n$.

Problem 2.15 (3 Points)**.** Given a cycle decomposition for a permutation p, describe the cycle decomposition for its *inverse* (that is, a permutation q such that $p \circ q = q \circ p = I$ where I is the permutation that fixes every element).

The symmetric groups give an example of a subgroup called the *alternating group* which is more subtle than the rotation and reflection subgroups we encountered earlier. Similar to how we simplified cycles in the previous problem, we can also write permutations as products of the simplest cycles, transpositions. Any permutation $p \in S_n$ may be written as a product of finitely many transpositions. For example, $p = (1\,5\,2\,3\,8)(4\,6\,11\,10)(7\,9)$ may be written as the product of 8 transpositions:

$$p = (1\,5)(5\,2)(2\,3)(3\,8)(4\,6)(6\,11)(11\,10)(7\,9),$$

while $(4\ 3\ 9\ 7)(6\ 2\ 8)$ may be written as the product of 5 transpositions. Unlike cycle decomposition, this product is not unique; for example, $(2\,3\,1) = (1\,2)(2\,3) = (1\,3)(2\,1)$. However, the following problem asserts that it is *unique up to the parity of the number of transpositions*.

Now we state some facts without proof. Let $p \in S_n$ be a permutation. If p can be written as the product of an even number of transpositions, then any way to express the permutation as a product of transpositions will also involve an even number of transpositions. In this case, p is called an *even permutation*. Any other permutation is called an *odd permutation*.

Furthermore, if $p, q \in S_n$ are permutations, each of which may be written as the product of an even number of transpositions, the composition $p \circ q \in S_n$ and the inverse $p^{-1} \in S_n$ have the same property. You are welcome to verify these facts if you like (the proof is combinatorial in nature).

The collection of all permutations in S_n which may be written as the product of an even number of transpositions is called the *alternating group* A_n and is a subgroup of S_n, a concept that will be formalized later. For now, think of it as being analogous to the subgroup of rotations C_n of D_n in the sense that it is a "subset of the transformations that satisfy some additional property."

Here, we also have a concept “dual” to a subgroup, namely a quotient group. By the above problem, every element of S_n can be written as either an even number or an odd number of transpositions. We now explore this idea further:

Problem 2.16 (2 Points)**.** Suppose $n \geq 2$. Let A_n be the alternating subgroup of S_n and $B_n = S_n \setminus A_n$.

(a) (1 point) Show that A_n and B_n have the same size and show that for any fixed $\sigma \in B_n$, we can write $B_n = \{\sigma \circ \tau \mid \tau \in A_n\}$.

(b) (1 point) Show that if $\tau_i \in A_n$ and $\sigma_i \in B_n$, then $\tau_1 \circ \tau_2 \in A_n$, $\tau_1 \circ \sigma_1 \in B_n$, $\sigma_1 \circ \tau_1 \in B_n$, and $\sigma_1 \circ \sigma_2 \in A_n$.

The punchline is that we can create a “multiplication table” with the elements being the **sets** A_n, B_n as follows:

	A_n	B_n
A_n	A_n	B_n
B_n	B_n	A_n

The “multiplication of sets” is as follows: the product of two sets X, Y (where X, Y are each either A_n or B_n) is found by taking an element of each set, i.e. $x \in X, y \in Y$, and identifying which of the two sets A_n, B_n the product $x \circ y$ lies in. The previous problem shows that the above multiplication table is indeed the correct one. Of course, this multiplication makes sense only if the choices of elements x, y *do not matter*. That is, if we picked some other $x' \in X, y' \in Y$, then the product $x' \circ y'$ lies in A_n if and only if $x \circ y$ lies in A_n. This construction is an example of a **quotient group**, a concept which will be formalized later.

2.4 Some Coincidences

To motivate the abstract definition to come, we will see that some of the groups we have encountered in two different contexts are really the same. Namely, there are bijections which ”preserve composition laws”.

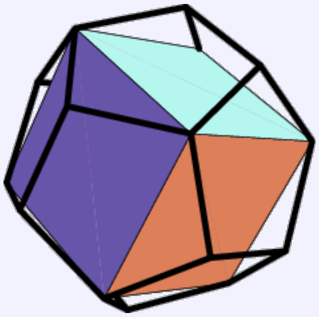
Problem 2.17 (4 Points)**.** Show that there is a bijection $\varphi : D_3 \to S_3$ such that $\varphi(a \circ b) = \varphi(a) \circ \varphi(b)$ for all $a, b \in D_3$.

Similar to how we defined D_n, we can also define transformation groups of three-dimensional regular polyhedra as being the transformations which preserve the shape, but may change the labelling of the vertices. Rigorously speaking: take a copy of any regular polyhedron P in $\mathbb{R}^3$ such that its circumcenter lies at the origin of $\mathbb{R}^3$. A (rigid) *symmetry transformation* of P is a bijective map of P onto itself which is effected by a rotation of $\mathbb{R}^3$ with respect to an axis passing through $\mathbf{0} \in \mathbb{R}^3$. As usual, this definition does not depend on the “embedding” of P in $\mathbb{R}^3$. Any symmetry transformation of P determines a permutation of the vertices of P (or the edges of P, or the faces of P). Furthermore, if S, T are any two symmetry transformations of P, then $S \circ T$ is again a symmetry transformation of P.

The transformations which can be obtained as a rotation (as opposed to a flip) are called orientation preserving.

Problem 2.18 (4 Points)**.** Let G be the set of symmetry transformations of a cube. Show that there is a bijection $\varphi : G \to S_4$ such that $\varphi(a \circ b) = \varphi(a) \circ \varphi(b)$ for all symmetry transformations $a, b \in G$.

Problem 2.19 (5 Points)**.** Let G be the group of symmetry transformations of a dodecahedron. Show that there is a bijection $\varphi : G \to A_5$ such that $\varphi(a \circ b) = \varphi(a) \circ \varphi(b)$ for all $a, b \in G$. *Hint: consider the following figure*:



2.5 Abstract Definitions and Properties

As the above problems show, the collection of symmetries of two seemingly different objects actually turn out to be the same. This is known as an *isomorphism*, which will be formalized later. In modern mathematical fashion, we now abstract out the important aspects of what symmetries are. We use this abstract notion to derive powerful statements about the transformations. The important aspects are: the existence of a "do nothing" transformation, a composition law, and the existence of inverses. Using these aspects, we are finally ready to give the "official" formal definition of a **group**.

Definition 2.5 (Group)**.** A *group* is a nonempty set G with a distinguised *identity element*, often denoted $e \in G$, and a binary operation $\cdot : G \times G \to G,\ (g,h) \mapsto g \cdot h$ called the *multiplication map* (i.e., $g \cdot h$ is the product of g and h in that order), satisfying the following axioms:

- **Associativity**: for all $x, y, z \in G$, $x \cdot (y \cdot z) = (x \cdot y) \cdot z$.
- **Identity**: for all $x \in G$, $e \cdot x = x \cdot e = x$.
- **Inverse**: for all $x \in G$, there exists $y \in G$ such that $x \cdot y = y \cdot x = e$. Such a y is usually denoted x^{-1}).

The data of a group will be written as $(G, \cdot, e)$, or simply G for brevity.

If in addition we have

- **Commutativity**: For all $x, y \in G$, $x \cdot y = y \cdot x$,

then the group G is called *commutative* or *abelian*. Otherwise, the group is called *non-commutative* or *non-abelian*.

The cardinality of the set G, denoted by $|G|$, is called the *order* of the group G. We say that G is a *finite group* if $|G|$ is finite.

Notes on notation:

- The identity element may also be written as $I, 1, 0, e$ depending on the context.
- The multiplication may also be written as $x \times y$, $x \circ y$, $x + y$ or simply as xy.
- Usually 0 and $+$ are used in conjunction when working with an abelian group and here, the "multiplication map" is usually called *addition*.

To explain the definition, a group can be thought of as a collection of symmetries with the multiplication map acting as composition. The identity element is the "do nothing" symmetry while the inverse is performing the opposite symmetry. The associativity says that it doesn't matter what order we compose symmetries, we will end up getting the same end result. The groups that we have encountered so far have not been abelian, but we will later encounter abelian groups.

The advantage of making this abstract definition is that it allows us to talk about a group as an independent object separate from any set that its acting on. Hence, we can make overarching statements that apply to many, sometimes even all, groups and get a more wholistic understanding of them. The remainder of this section will deal with abstract properties of groups, and later sections will deal with understanding how they act on different objects.

Now let's explore some basic properties:

Problem 2.20 (3 Points)**.** Let G be a group.

(a) (1 point) Show that the inverse of any element of G is unique, that is, if $x, y_1, y_2 \in G$ are such that $x \cdot y_1 = x \cdot y_2 = e$, then $y_1 = y_2$ (hence, we are justified in saying "the inverse").

(b) (1 point) Given $x, y, z \in G$. Show that if $x \cdot y = x \cdot z$, then $y = z$. Show that if $z \cdot x = y \cdot x$, then $y = z$.

(c) (1 point) Show that $f_x : G \to G$ defined as $f_x(y) = x \cdot y$ is a bijection, and that the same holds for $g_x : G \to G$ where $g_x(y) = y \cdot x$.

Let's go over some common examples.

Example 2.3. Any singleton set may be viewed as a group, called the *trivial group*.

The integers under addition $(\mathbb{Z}, +, 0)$ is an abelian group. The identity element is $0 \in \mathbb{Z}$, and the inverse of any $n \in \mathbb{Z}$ is just $-n \in \mathbb{Z}$.

Let $\mathbb{Q}^\times$ be the set of non-zero rational numbers. Then, the non-zero rational numbers under multiplication $(\mathbb{Q}^\times, \cdot, 1)$ is an abelian group. The identity element is $1 \in \mathbb{Q}^\times$, and the inverse of a rational number $\frac{p}{q} \in \mathbb{Q}$, where p, q are coprime integers with $p, q \neq 0$, is $\frac{q}{p} \in \mathbb{Q}^\times$.

The set D_n of symmetry transformations of the regular polygon with n sides under the composition operation $\circ$ is a group, called the *dihedral group* with $2n$ elements (cf. Section 2.2). Its identity element is I, the "do nothing" transformation, or identity transformation. You have already seen in Problem 2.10 that every transformation $T \in D_n$ has an inverse. The group D_n is non-abelian.

Similarly, we have seen that the set S_n of all permutations of $\underline{n} = \{1, 2, \ldots, n\}$ under the function composition operation $\circ$ is a non-abelian group, called the *symmetric group on n elements* (cf. Section 2.3).

Problem 2.21 (2 Points)**.**

(a) (1 Point) Let m be a positive integer. Show that the set μ_m of complex solutions to the equation $z^m = 1$ is an abelian group under multiplication.

(b) (1 Point) Explain why the triples $(\mathbb{Z}, \cdot, 1)$ and $(\mathbb{Q}, \cdot, 0)$ aren't groups ($\cdot$ represents the usual multiplication).

Now we move to the definition of a *subgroup*:

Definition 2.6. A nonempty subset H of a group $(G, \cdot, e)$ is a *subgroup* of G if it is closed under multiplication and inverse. That is, for all $x, y \in H$, we have $x \cdot y, x^{-1} \in H$ (thus, we may speak of the group $(H, \cdot, e)$). We use the notation $H \leq G$ to say that H is a subgroup of G. We say that H is a *proper* subgroup of G if H is a proper subset of G (as sets).

A *left coset* of H in G is the set $gH = \{gh \mid h \in H\}$ where $g \in G$ is fixed. Similarly, a *right coset* of H in G is $Hg = \{hg \mid h \in H\}$ where $g \in G$ is fixed. In both cases, g is called a *representative* for the coset.

Example 2.4. For any group $(G, \cdot e)$, the subset $\{e\}$ is always a subgroup of G, called the *trivial subgroup*.

The subset C_n of all rotations in D_n is a proper subgroup of D_n. For any flip $F \in D_n$, the subset $\{I, F\}$ of D_n is a proper subgroup of D_n. We call C_n the *rotation subgroup* of D_n, and call $\{I, F\}$ a *reflection subgroup* of D_n (cf. Section 2.2).

The set A_n of all permutations in S_n which may be written as the product of an even number of of transpositions is a proper subgroup of S_n. The subgroup A_n is called the *alternating group on n elements* (cf. Section 2.3). The set $B_n := S_n \setminus A_n$ is a coset of A_n by Problem 2.16: for any fixed $\sigma \in B_n$, we can write $B_n = \sigma \circ A_n$.

Problem 2.22 (4 Points)**.** Let $H \leq G$ be a subgroup, and let $g, g' \in G$.

(a) (1 Point) Show that the left (resp. right) coset gH (resp. Hg) is a subgroup of G if and only if $g \in H$, in which case $gH = H$.

(b) (1 Point) Show that in general $|gH| = |H|$ for any $g \in G$. In other words, there is a bijection between the sets H and gH. Show also that $|Hg| = |H|$.

(c) (1 Point) Show that the left cosets of H in G *partition* G. That is, for the left cosets $gH, g'H$, we either have $gH = g'H$ or $gH \cap g'H = \emptyset$. Prove that the same is true for the right cosets of H in G.

(d) (1 Point) (*Lagrange's theorem*) Suppose G is finite. Show that the number of distinct left (resp. right) cosets of H is $\frac{|G|}{|H|}$.

Definition 2.7 (Index of a subgroup)**.** Let G/H be the set of left cosets of H in G (thus, G/H is a *set* of *subsets of* G). The *index* of H in G, denoted by $[G : H]$, is defined to be the cardinality of the set G/H.

Hence, Lagrange's theorem implies that when $H \leq G$ is a subgroup and G is finite, then $[G : H] = \frac{|G|}{|H|}$. **Lagrange's theorem is really important.**

Example 2.5. For any $n \geq 3$, the index of C_n in D_n is 2.

Remark. In the above definition, we could have replaced the word "left" with "right" and obtained the same definition of index. It is somewhat tricky to prove that these two definitions coincide, especially since Lagrange's theorem only addresses the case that G is finite!

Problem 2.23 (3 Points)**.** For a fixed positive integer m:

(a) (1 Point) Show that $(m\mathbb{Z}, +, 0)$ is a subgroup of $(\mathbb{Z}, +, 0)$ where $m\mathbb{Z}$ is the subset of $\mathbb{Z}$ consisting of all (positive and negative) multiples of m.

(b) (2 Points) Show that $(\mathbb{Z}/m\mathbb{Z}, +, \overline{0})$ forms a group where $\mathbb{Z}/m\mathbb{Z}$ is the set of all left cosets $\overline{k} = k + m\mathbb{Z}$ of $m\mathbb{Z}$ in $\mathbb{Z}$, where $k \in \mathbb{Z}$, and the group operation $+$ is defined as $\overline{k} + \overline{l} = \overline{k+l}$ (in particular, $\overline{k+l}$ denotes the *left coset* $(k+l) + m\mathbb{Z}$ of $m\mathbb{Z}$ in $\mathbb{Z}$). Oftentimes, mathematicians are slothful and just write k, removing the bar, to refer to an element of $\mathbb{Z}/m\mathbb{Z}$. Also show that $|\mathbb{Z}/m\mathbb{Z}| = m$. *Hint: you need to check that addition is "well-defined", that is, if k, l, k', l' are integers such that $\overline{k} = \overline{k'}$ and $\overline{l} = \overline{l'}$, then $\overline{k+l} = \overline{k'+l'}$.*

The group $\mathbb{Z}/m\mathbb{Z}$ (sometimes simply denoted Z_m) is called the *integers under addition modulo m*. It is also called the *cyclic group of order m*. You have probably already seen this structure in elementary number theory. The group $\mathbb{Z}/m\mathbb{Z}$ is of order m, equal to the index $[\mathbb{Z} : m\mathbb{Z}]$, and it is an example of a **quotient group**, which we now explore.

Definition 2.8. Given two groups $(G, \cdot_G, e_G), (G', \cdot_{G'}, e_{G'})$, a *group homomorphism* (or *homomorphism* for short) is a map $\varphi : G \to G'$ such that $\varphi(g \cdot_G h) = \varphi(g) \cdot_{G'} \varphi(h)$ for all $g, h \in G$. A homomorphism is an *isomorphism* if it is a bijection as a map of sets. The *kernel* of a homomorphism, denoted by $\ker(\varphi)$, is the set $\{g \in G \mid \varphi(g) = e_{G'}\}$.

Intuitively, a homomorphism φ is map between groups that *preserves the group structure*. You can either do group multiplication before or after you apply the map φ, and you'll get the same answer in the end! Two groups are *isomorphic* if they are basically the "same object."

Example 2.6. The map $\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z}$ defined by sending k to $\overline{k}$ is a group homomorphism.

The map $\mathbb{Z}/m\mathbb{Z} \to \mu_m$ (where μ_m is defined in Problem 2.23) defined by sending $\overline{k}$ to $e^{\frac{2\pi ik}{m}}$ is a well-defined group isomorphism (why is this map "well-defined?"). Thus, some sources use μ_m to refer to the cyclic group of order m.

The groups $\mathbb{Z}/m\mathbb{Z}$ and D_n cannot be isomorphic, even if $m = 2n$, since one of these groups is abelian and the other is non-abelian. However, $\mathbb{Z}/m\mathbb{Z}$ is isomorphic to the subgroup C_m of D_m.

Problem 2.24 (3 Points)**.** Suppose $(G, \cdot_G, e_G)$ $(G', \cdot_{G'}, e_{G'})$ are groups and $\varphi : G \to G'$ is a homomorphism.

(a) (1 Point) Show that $\varphi(g^{-1}) = (\varphi(g))^{-1}$ and $\varphi(e_G) = e_{G'}$.

(b) (1 Point) Show that for any subgroup $H \leq G$, the set $\varphi(H) := \{\varphi(g) \mid g \in H\}$ is a subgroup of G' ($\varphi(H)$ is called the *image* of H under φ).

(c) (1 Point) Show that for any subgroup $H' \leq G'$, the set $\varphi^{-1}(H') := \{g \in G \mid \varphi(g) \in H'\}$ is a subgroup of G ($\varphi^{-1}(H')$ is called the *pre-image* of H' under φ). In particular, $\ker(\varphi)$ is a subgroup of G.

Problem 2.25 (2 Points)**.** Let $m > 1$ be a positive integer. Explain why there is no nontrivial group homomorphism $\varphi : \mathbb{Z}/m\mathbb{Z} \to \mathbb{Z}$.

Definition 2.9. A subgroup $H \leq G$ is *normal*, denoted by $H \trianglelefteq G$, if for all $g \in G$, $gH = Hg$ (the left and right cosets coincide). Equivalently, $\forall h \in H, g \in G, gHg^{-1} \in H$ Given a normal subgroup $H \trianglelefteq G$, the quotient group G/H is a group whose set of elements is $G/H = \{gH \mid g \in G\}$, the set of left cosets of H in G, with multiplication $gH \cdot_{G/H} g'H = (g \cdot_G g')H$ and identity eH (where $e \in G$ is the identity).

We have already seen examples of the quotient group: S_n/A_n and $\mathbb{Z}/m\mathbb{Z}$. However, it turns out that it is not possible to quotient out by any arbitrary group when G is not abelian. Let's see what happens if we try and brute force multiply elements of cosets and call the product the coset of the resulting element. For example, consider $G = S_3, H = \{1, (12)\}$. The cosets are $eH = \{1, (12)\}$, $(123)H = \{(123), (13)\}$, $(132)H = \{(132), (23)\}$. Let's multiply elements of the second and third cosets: $(123)(23) = (12)$ but $(13)(23) = (132)$. The first product lands in the first coset, but the second product lands in the third coset. So there isn't a well defined way to multiply these (their answers are different). If left cosets are equal to right cosets however, $gHg'H = g(Hg')H = g(g'H)H = gg'HH = gg'H$ so multiplication is well defined.

We have a surjective group homomorphism from G to G/H if H is normal which sends each group element to the left (or right) coset of H generated by it. Conversely, it can be checked that the kernel of any homomorphism is normal: if $\varphi : G \to G'$ is a group homomorphism and $H = \ker(\varphi)$, for any $g \in G, h \in H$, $\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g)^{-1} = e$, so $ghg^{-1} \in H$ and H is normal. Thus being normal is the right condition to be able to construct a quotient group. It is easily checked that any subgroup of an abelian group is normal, and thus can be quotiented by.

Example 2.7. The map sgn : $S_n \to \{-1,+1\}$ defined by $\text{sgn}(\sigma) = +1,\ \sigma \in S_n$ if and only if σ is an even permutation, is a surjective group homomorphism. The kernel of this homomorphism is $A_n \leq S_n$. Hence, A_n is a normal subgroup of S_n, and we may form the quotient group S_n/A_n.

Some final important concepts:

Definition 2.10. For $g \in G$, the *order of* g, denoted by $|g|$, is the smallest positive integer n such that $g^n = e$ where $g^n = g \cdot \cdots \cdot g$ n many times. If there is no such g, we say the order of g is infinite. The *cyclic subgroup* of G generated by g is the set $\langle g \rangle = \{g^n \mid n \in \mathbb{Z}\}$, where we define $g^{-n} = (g^{-1})^n$.

In lieu of the above definition: let m be a positive integer, and let Z_m be the cyclic group of order m. We say that any element $g \in Z_m$ of order m is a *generator* of Z_m.

Notice that for any element $g \in G$ of a group, the order of g is the same as the order of $\langle g \rangle$. Thus, by Lagrange's theorem (Problem 2.24), the order of g divides the order of G.

Example 2.8. The order of any generator of the rotation subgroup $C_n \leq D_n$ is exactly n. For instance, the 90° rotation in D_4 is of order 4. The order of any flip in D_n is 2.

Let p be a permutation in S_n with cycle decomposition $C_1 C_2 \ldots C_k$. Let l_i be the length of cycle C_i. Then, it is not too difficult to verify that the order of p is exactly $\text{lcm}(l_1, l_2, \ldots, l_k)$. For instance, the order of $(1\,6\,4\,2)(9\,3\,5)$ is 12.

Enough of definitions! Time for some fun theorems.

Problem 2.26 (2 points)**.** Let $m > 1$ be a positive integer.

(a) (1 point) Show that $((\mathbb{Z}/m\mathbb{Z})^\times, \cdot, \overline{1})$ is a group defined as follows. As a set, $(\mathbb{Z}/m\mathbb{Z})^\times$ is a subset of $\mathbb{Z}/m\mathbb{Z}$ consisting of the left cosets $\overline{k} := k + m\mathbb{Z}$ of $m\mathbb{Z}$ in $\mathbb{Z}$ such that every integer in $k + m\mathbb{Z}$ is relatively prime to m. The group operation $\cdot$ is defined as $\overline{k} \cdot \overline{l} = \overline{k \cdot l}$.

(b) (1 point) Prove (one of many) Euler's theorem. If a is the number of integers in $0, 1, \ldots, m-1$ coprime to m, then $m \mid k^a - 1$ for any integer k coprime to m.

Problem 2.27 (5 Points)**.** Given groups $G_1, \ldots, G_n$, their *direct product* is a group, whose set is the Cartesian product

$$G_1 \times \cdots \times G_n = \{(g_1, \ldots, g_n) \mid g_i \in G_i\},$$

and group multiplication is defined componentwise. Namely, for $(g_1, \ldots, g_n), (h_1, \ldots, h_n) \in G_1 \times \cdots \times G_n$, the product $(g_1, \ldots, g_n) \cdot (h_1, \ldots, h_n)$ in $G_1, \times \cdots \times G_n$ is the ordered tuple $(g_1 h_1, \ldots, g_n h_n)$. Here, $g_i h_i$ is the product of g_i and h_i in G_i.

Let $m > 1$ be a positive integer, Suppose the prime factorization for m is $p_1^{a_1}, \ldots, p_n^{a_n}$, for $a_1, \ldots, a_n \geq 1$. Prove the following isomorphisms:

(a) (2 Points) $\mathbb{Z}/m\mathbb{Z} \cong \mathbb{Z}/p_1^{a_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/p_n^{a_n}\mathbb{Z}$. You may recognize this as the Chinese Remainder Theorem.

(b) (3 Points) $(\mathbb{Z}/m\mathbb{Z})^\times \cong (\mathbb{Z}/p_1^{a_1}\mathbb{Z})^\times \cdots \times (\mathbb{Z}/p_n^{a_n}\mathbb{Z})^\times$

3 More Group Theory

3.1 Group Actions

Earlier, we saw that a subgroup H of G is a normal subgroup precisely when $gHg^{-1} = H$ for each $g \in G$, suggesting that conjugation is an important tool for studying the structure of groups and that we should try to harness its power. We will make this formal by introducing the notion of *group actions* (of which conjugation will then be a special case, acting on subsets of the group G itself!).

Definition 3.1. Let G be a group with identity element $e \in G$ and S a set. Then, a (left) *group action* of G on S is a function $f : G \times S \to S$ such that $f(e,s) = s$ and $f(g, f(h,s)) = f(gh,s)$ for all $s \in S$ and $g,h \in G$.

Typically, $f(g,s)$ is just written as $g \cdot s$ (so, the second axiom says $g \cdot (h \cdot s) = (gh) \cdot s$ for all $s \in S$ and $g,h \in G$).

We say that a group G acts on S *faithfully* (or that the group action is faithful) if $g \cdot s = s$ for all $s \in S$ implies g is the identity $e \in G$.

We emphasize that the $\cdot$ above corresponds to the **group action**, not the *group operation*!!

Example 3.1. Let S_n be the symmetric group on n elements. Then, S_n acts on $\underline{n} = \{1, 2, \ldots, n\}$ in the obvious way: $\sigma \cdot i = \sigma(i)$ for all $\sigma \in S_n, i \in \underline{n}$. This is called the *permutation action*.

We see some examples of group actions in the case that the set S is the group G itself.

Example 3.2 (Left Regular Action)**.** Let G be a group. Then, the map $f : G \times G \to G$ defined by $f(g,h) = gh$ is a group action. Here, gh is the *product* of group elements g and h. Thus, any group G *acts on itself* by left multiplication. This is called the *regular group action* of G.

Example 3.3 (Conjugation)**.** For any group G, we see that G *acts on itself by conjugation*: $x \cdot g = xgx^{-1}$ for $x, g \in G$. In particular for $x,y,g \in G$, notice $(xy) \cdot g = xyg(xy)^{-1} = xygy^{-1}x^{-1} = x \cdot (y \cdot g)$.

We will now show how group actions induce group homomorphisms. Given a (left) group action G on a finite set S. For $g \in G, s_1, s_2 \in S$, if $g \cdot s_1 = g \cdot s_2$, then multiplying by g^{-1} implies that $s_1 = s_2$, which means that the map $f_g : S \to S$ given by $s \mapsto g \cdot s$ is injective. As S is finite, this implies that f_g is surjective. Hence f_g is a bijective map $S \to S$, making it an permutation of the set S. Hence we get an induced map $\phi : G \to \mathrm{Perm}(S)$ given by $g \to f_g$. To check this is a group homomorphism, note that $\phi(g_1g_2)(s) = f_{g_1g_2}(s) = (g_1g_2) \cdot s = g_1 \cdot (f_{g_2}(s)) = f_{g_1}(f_{g_2}(s)) = (f_{g_1} \circ f_{g_2})(s) = (\phi(g_1) \circ \phi(g_2))(s)$ for each $s \in S$, which implies that $\phi(g_1g_2) = \phi(g_1) \circ \phi(g_2)$, as desired. Note that we can also recover the group action from the induced homomorphism because it contains the data of $g \cdot s$ for each $g \in G$ and $s \in S$. Observe that G acts on S faithfully if and only if the kernel of $\phi : G \to \mathrm{Perm}(S)$ is trivial ($\ker(\phi) = \{e\}$) if and only if ϕ is injective.

Example 3.4 (An Induced Homomorphism)**.** Consider the permutation action (Example 3.1). The resulting homomorphism just identifies an element σ of S_n with its usual definition as a permutation of $\{1, \cdots, n\}$

Definition 3.2 (Orbits and Stabilizers)**.** Let G be a group acting on a set S. Given $s \in S$, the *orbit* of $s \in S$ is the subset of S given by

$$\mathrm{Orb}(s) := \{t \in S \mid g \cdot s = t \text{ for some } g \in G\}.$$

The *stabilizer* of $s \in S$ is the subset of G given by

$$\mathrm{Stab}(s) := \{g \in G \mid g \cdot s = s\}.$$

The above definition of orbit is consistent with the previous definition (Definition 2.3), if we take $G = S_n$, $S = \underline{n}$, and the permutation action.

In general, the *distinct orbits* of a group G acting on a set S define a *partition* of the set S. Namely, every $s \in S$ lies in exactly one distinct orbit of the group action!

Definition 3.3 (Conjugacy classes)**.** Let G be a group, acting on itself by conjugation. The orbits of this group action are called the *conjugacy classes* of G. Namely, for any $g \in G$, the conjugacy class of g is the subset $\{xgx^{-1} \mid x \in G\}$.

Note that conjugacy classes of G are not necessarily subgroups of G.

Example 3.5 (Centralizer)**.** Let G be a group and $g \in G$ an element. The *centralizer* of g, denoted by $C_G(g)$, is the subgroup of G commuting with g, i.e. $C_G(g) = \{x \in G \mid gx = xg\}$. Note that $C_G(g)$ is precisely the *stabilizer* of g under the conjugation action since $C_G(g)$ consists of the $x \in G$ with $xgx^{-1} = g$!

We then have the following important result:

Theorem 3.4 (The Orbit-Stabilizer Theorem)**.** Given a finite group G, a set S, and an element $s \in S$, we have that $|\mathrm{Orb}(s)| \cdot |\mathrm{Stab}(s)| = |G|$.

Problem 3.1 (4 Points)**.** Prove the Orbit-Stabilizer Theorem.

Using this result, we can prove the class equation for any finite group G.

The class equation says that given any finite group G with conjugacy classes $C_1, \cdots, C_n$ and a choice of elements $c_i \in C_i$ for each i, we have that:

Theorem 3.5 (The Class Equation)**.** $|G| = \sum_{i=1}^{n} [G : C_G(c_i)]$.

Proof. We return to the conjugation action on G of Example 3.4. We saw there that g_1 and g_2 are equivalent if and only if they are in the same conjugacy class, and thus the equivalence classes under this action are just the set of conjugacy classes of G. These equivalence classes partition G, and so we conclude that $|G| = \sum_{i=1}^{n} |C_i|$. Thus the result comes down to showing that $|C_i| = [G : C_G(c_i)]$. We can restrict the group action of G to the set C_i and then choose $c_i \in C_i$ arbitrarily as our s in the class equation. The stabilizer of c_i is the set of $g \in G$ with $gc_ig^{-1} = c_i$, which is just the centralizer of c_i by definition. Then by the Orbit-Stabilizer Theorem, we have that $|\mathrm{Orb}(c_i)||\mathrm{Stab}(c_i)| = |G|$, and $\mathrm{Orb}(c_i)$ is just C_i, allowing us to conclude that $|C_i| = |G|/|C_g(c_i)|$. Finally, we have that $|G|/|C_g(c_i)| = [G : C_g(c_i)]$, finishing the proof of the class equation. □

3.2 The Alternating Group

For the remainder of Section 3, let $n \geq 2$. The symmetric group S_n and the alternating group A_n are trivial when $n = 1$.

Recall that S_n is the group consisting of all permutations of $1, \cdots, n$. We also introduced the sign homomorphism $\mathrm{sgn} : S_n \to \{-1, +1\}$ in Section 2. That is, for $\sigma \in S_n$, we have $\mathrm{sgn}(\sigma) = +1$ if and only if σ may be written as the product of an even number of transpositions, and $\mathrm{sgn}(\sigma) = -1$ if and only if σ may be written as the product of an odd number of transpositions.

The kernel of this homomorphism is the group A_n, the alternating group on the set $\{1,\cdots,n\}$, consisting of the σ with $\mathrm{sgn}(\sigma)=+1$, which are known as the *even permutations*. Similarly, those with $\mathrm{sgn}(\sigma)=-1$ are known as *odd permutations*. We can determine the size of the subgroup A_n.

Problem 3.2 (1 Point)**.** Show that $|A_n| = n!/2$.

Here are some examples of even and odd permutations:

Example 3.6. (Even permutation in S_5) Consider the permutation $(1\ 2\ 3\ 4\ 5)$. It turns out that an n-cycle is an even permutation when n is odd and an odd permutation when n is even, a result which will be proven in the next subsection. This may seem a bit counterintuitive, but recall that a transposition is a 2-cycle, so one can already see the parity imbalance with this simple example. We will discuss the power of cycles much more later in this section.

Example 3.7. (Odd permutation in S_5) Consider the permutation $(1\ 2)(3\ 4\ 5)$. If we start with 1, we see that it is mapped to 2, which is then mapped back to 1, giving a cycle of length 2. On the other hand, starting at 3, we see it is mapped to 4, which is mapped to 5, which is mapped to 3, giving a second cycle of length 3. From the above result on cycles, 2-cycles are even while 3-cycles are odd, and this permutation is the product of the two cycles $(1\ 2)$ and $(3\ 4\ 5)$, and thus by the homomorphism property, it is odd.

It turns out that A_n is a normal subgroup of S_n (in fact, given a group G, any subgroup of index 2 is automatically normal). For $n \geq 5$, the alternating groups A_n have a special property: they are simple groups, which means that they have no nontrivial normal subgroups (their only normal subgroups are $\{e\}$ and A_n). We will assume this fact without proof.

Problem 3.3 (4 Points)**.** Use the simplicity of $A_n, n \geq 5$ to show that for any $n \geq 5$, besides $\{e\}$ and S_n itself, A_n is the only normal subgroup of S_n.

Although this just gives information about the normal subgroups of S_n, it soon becomes apparent that understanding the normal subgroups of S_n goes a long way in understanding general subgroups of S_n.

Problem 3.4 (5 Points)**.** Show that for any $n \geq 5$, if H is a proper subgroup of S_n and $H \neq A_n$, then H has index at least n in S_n.

This bound is tight because we can consider the subgroup of S_n with $\sigma(n) = n$, which is a subgroup of index n. To see why this latter fact is true, note that there are $(n-1)!$ permutations with $\sigma(n) = n$, and thus they form a subgroup of index $n!/(n-1)! = n$. This subgroup turns out to be isomorphic to S_{n-1}, which can be seen by mapping σ to the permutation of $\{1,\cdots,n-1\}$ it induces.

3.3 Conjugacy Classes of S_n, A_n

Recall from Section 2 that any element σ of S_n has a *cycle decomposition*, consisting of cycles of the form $\{i, \sigma(i), \sigma^2(i) \cdots \sigma^{k-1}(i)\}$, where k is the size of the orbit of $i \in \{1,2,\ldots,n\}$ under the subgroup $\langle\sigma\rangle \subseteq S_n$ acting on $\{1,\cdots,n\}$. Furthermore, σ has a unique *cycle type*.

Problem 3.5 (2 Points)**.** Show that if the cycle type of σ is $b_1,\cdots,b_r$ (i.e., the cycle lengths of σ, then the order of σ is $\mathrm{lcm}(b_1,\ldots,b_r)$.

Now we will see the connection between conjugacy classes in S_n and cycle type, which you can recall from Section 2.

Problem 3.6 (2 Points)**.** Let $(a_1 \cdots a_k)$ be a cycle. Show that $\tau(a_1 \cdots a_k)\tau^{-1} = (\tau(a_1)\cdots\tau(a_k))$.

Thus writing an element σ as $\sigma = \prod_{i=1}^{r}(a_{i1}\cdots \quad a_{ic_i})$ (a disjoint cycle decomposition), we have that $\tau\sigma\tau^{-1} = \prod_{i=1}^{r}(\tau(a_{i1})\cdots\tau(a_{ic_i}))$, and so any element conjugate to σ has the same cycle type. Now choose any other σ' with the same cycle type, say $\prod_{i=1}^{r}(b_{i1}\cdots b_{ic_i})$. Choosing τ so that $\tau(a_{ic_i}) = b_{ic_i}$ for each i, j pair with $1 \le j \le c_i$, we see that any element of S_n with the same cycle type as σ is actually conjugate to σ, and thus the conjugacy classes are precisely sets of elements with particular cycle type. A cycle type is precisely a set of positive integers summing to n, i.e. *a partition of* n.

Problem 3.7 (3 Points)**.** Let C be the conjugacy class consisting of the $\sigma \in S_n$ corresponding to a partition P of n. Given such a partition P, let a_i be the number of appearances of i in the partition for each $1 \le i \le n$. Show that

$$|C| = \frac{n!}{\prod_{i=1}^{n} a_i! i^{a_i}}$$

Since we know that the conjugacy classes partition S_n, we can swiftly prove a tricky combinatorial lemma using the machinery developed thus far.

Example 3.8 (A Combinatorial Lemma)**.** Recall that for a finite group G with conjugacy classes $C_1, \cdots, C_n$, we had that $|G| = \sum_{i=1}^{n}|C_i|$. Let T be the set of all partitions of n. Using the result of Problem 3.7 for the group $G = S_n$, and summing over the elements of T, we conclude that

$$n! = |S_n| = \sum_{P\in T} \frac{n!}{\prod_{i=1}^{n} a_i! i^{a_i}}.$$

This may be rewritten as

$$n! = \sum_{\sum_{i=1}^{n} ia_i = n} \frac{n!}{\prod_{i=1}^{n} a_i! i^{a_i}}$$

Dividing by $n!$ gives the purely combinatorial identity

$$1 = \sum_{\sum_{i=1}^{n} ia_i = n} \frac{1}{\prod_{i=1}^{n} a_i! i^{a_i}}$$

The cycle type of a permutation also determines its sign. As before, we first do the calculation for cycles.

Problem 3.8 (1 Point)**.** Let $(a_1 \cdots a_k)$ be a cycle. Then show that $\operatorname{sgn}((a_1 \cdots a_k)) = (-1)^{k-1}$.

Now let σ be a permutation with cycle lengths $k_1, \cdots k_r$. We have that $k_1 + \cdots + k_r = n$, and thus that $\operatorname{sgn}(\sigma) = \prod_{i=1}^{r}(-1)^{k_i-1} = (-1)^{n-r}$. Hence the sign of σ depends only on the parity of $n-r$. In particular, given any partition $a_1, \cdots, a_r$ of n with $n-r$ odd, the conjugacy class corresponding to that partition consists of only odd permutations and thus intersects A_n trivially. Now we investigate the case when $n-r$ is even. Any element corresponding to such a partition will be an element of A_n.

In the above, we saw that a conjugacy class of S_n is either entirely in A_n or disjoint from A_n. This embodies the fact that A_n is a normal subgroup of S_n in light of the result that given a group G and a subgroup H, H is normal in G iff H is the union of conjugacy classes of G (why?). Furthermore, an individual conjugacy class of G contained in H will itself be a union of conjugacy classes of H because if two elements are conjugate in H, then they are necessarily conjugate in G, and so we may repeatedly remove H-conjugacy classes of some G-conjugacy class until none remain, writing the G-conjugacy class as a disjoint union of H-conjugacy classes. If K is a conjugacy class of G such that K is the (disjoint) union of k conjugacy classes in H, we say that K splits into k conjugacy classes of H.

Problem 3.9 (8 Points)**.**

(a) (5 Points) Let G be a finite group and H a normal subgroup. Show that if K is a conjugacy class of G splitting into k conjugacy classes of H, then each of the k conjugacy classes has the same size.

(b) (3 Points) Now let $G = S_n$ and $H = A_n$. Show that $k = 1$ or $k = 2$ and that $k = 2$ iff no element of K commutes with an odd permutation.

The latter condition can be smoothly translated into a condition on cycle lengths:

Problem 3.10 (5 Points)**.** Let $g \in A_n$ be an even permutation. Then show that g commutes with no odd permutation iff its cycle lengths are distinct odd positive integers.

From this, we deduce that the conjugacy classes of S_n that split are the ones corresponding to partitions of n into distinct odd positive integers. Note that these satisfy the parity condition on conjugacy classes of A_n automatically since $n = k_1 + \cdots + k_r \equiv r \mod 2$ (since each k_i is odd).

Example 3.9. We demonstrate an example of the theory developed to conjugacy classes of S_5. Take the conjugacy class of 5-cycles, corresponding to the partition $\{5\}$ of 5. Since this partition consists of distinct odd parts, it consists only of even permutations, and by the previous two problems, it will split into exactly two conjugacy classes in A_5. These two conjugacy classes are given, for example, by the conjugacy classes generated by the elements (12345) and (21345).

We finish this section by giving an application of the theory developed so far to give a proof of a nice result about subgroups of S_n. Let G be a finite abelian group. The Structure Theorem for Finite Abelian Groups tells one that there exist a unique list of prime powers $(q_1, \cdots, q_n)$ (up to ordering) so that $G \cong \prod_{i=1}^n \mathbb{Z}/q_i\mathbb{Z}$. This gives a well-defined invariant of G, namely $\sum_{i=1}^n q_i$, which we know is determined by the uniqueness of the list $q_1, \cdots, q_n$ in the result. Denote this quantity by $f(G)$. Then we have that G is isomorphic to a subgroup of S_n iff $f(G) \leq n$.

For one direction, if $f(G) \leq n$, we may choose disjoint cycles of S_n of lengths q_i for each $1 \leq i \leq n$, which together generate a subgroup with the desired isomorphism type.

Conversely, suppose that G is isomorphic to a subgroup of S_n, which we call H. We will define a group action involving H. Let our set S just be $\{1, \cdots, n\}$ and consider the action of H on S just by permuting the elements of S according to H. We consider the orbits of H acts on S, which are equivalence classes defined by the relation $i \sim j$ iff for some $h \in H$, $h(i) = j$. Let the orbits of H acting on S be $S_1, \cdots, S_r$. For each $1 \leq i \leq r$, let H_i be the subgroup of $S_{|S_i|}$ (the subgroup of permutations of the elements of S_i) attained by restricting H to the set S_i. Then it follows that G injects into $\prod_{i=1}^r H_i$ upon composing the isomorphism with H and the restriction of the action of H on the S_i because if some element of H were to be the identity permutation on each of the S_is, this would mean it would fix all of S, implying it is the identity permutation of S.

Now we claim that $|H_i| = |S_i|$ for each i. By definition of the equivalence relation, H and thus H_i acts transitively on the individual S_i (i.e. given s_1, s_2 in some S_i, we can choose h so that $hs_1 = s_2$). In particular, the orbit of any element $s \in S_i$ under H_i is just $|S_i|$. Thus by the Orbit-Stabilizer Theorem, we have that $|H_i| = |S_i||\text{Stab}(s)|$. Thus we just need to show that $\text{Stab}(s)$ is trivial. For this, choose $h \in \text{Stab}(s)$ and note that for any $s' \in S_i$, since we may write $s' = h's$ for some $h' \in H_i$, we have that $hs' = hh's = h'(hs) = h's = s'$, which shows that each element of $\text{Stab}(s)$ fixes all of S_i and thus is the identity element of H_i. This shows that $\text{Stab}(s)$ is trivial and so $|H_i| = |S_i|$. Thus since the S_is partition S, we conclude that $n = \sum_{i=1}^r |H_i|$. Next we claim that $\sum_{i=1}^r |H_i| \geq f(G)$. Since G is isomorphic to a subgroup of $\prod_{i=1}^r H_i$, we have that $f(\prod_{i=1}^r H_i) \geq f(G)$. But then f is additive on direct products, so $f(\prod_{i=1}^r H_i) = \sum_{i=1}^r f(H_i)$. Finally, we have the bound that $f(H_i) \leq |H_i|$ since given positive integers $a_1, \cdots, a_n \geq 2$, $\sum_{i=1}^n a_i \leq \prod_{i=1}^n a_i$. Thus we conclude that $f(G) \leq \sum_{i=1}^r |H_i| = n$, as desired.

3.4 The Commutator Subgroup

In this section, unless stated otherwise, G is any group. Given elements $g, h \in G$, we can define $[g,h] = ghg^{-1}h^{-1}$ to be their commutator. The commutator subgroup G' is the subgroup of G generated by all the commutators. The commutator subgroup has many useful properties.

Problem 3.11 (1 Point)**.** Let G be an abelian group. Then show that G' is trivial.

In some sense, this result is telling one that there is no obstruction to being abelian for abelian groups. We will refine this idea in the following problems to show how one can use the commutator subgroup to obtain an abelian quotient of a group G.

Problem 3.12 (3 Points)**.** Show that G' is a normal subgroup of G.

Thus $G/G' = G^{\text{ab}}$ is a group, known as the abelianization of G. In some sense, it is the maximal abelian quotient of G. To make this more precise:

Problem 3.13 (5 Points)**.** Let H be a normal subgroup of G such that G/H is abelian.

(a) (2 Points) Show that $G' \subset H$.

(b) (3 Points) Show that G/H is isomorphic to a quotient of G^{ab}.

Example 3.10. Let us compute the commutator subgroup of S_n. For $n = 1,2$, it is trivial since S_n is abelian in these cases. For each $n > 2$, it is nontrivial because S_n itself is nonabelian, meaning that $xyx^{-1}y^{-1} = xy(yx)^{-1} = e$ is not always true. Furthermore, it is also always a subgroup of A_n because $\text{sgn}(xyx^{-1}y^{-1}) = \text{sgn}(x)\text{sgn}(y)\text{sgn}(x^{-1})\text{sgn}(y^{-1}) = \text{sgn}(xx^{-1})\text{sgn}(yy^{-1}) = 1$, and so since the generators are even, multiplicativity implies every element in the whole subgroup is even. This implies that for $n = 3$, it is A_3 since A_3 has only two subgroups and the commutator subgroup can't be the trivial one. Now for $n \geq 5$, Problems 3.3 implies that the commutator subgroup of S_n must be A_n because it is the only nontrivial normal subgroup of S_n contained in A_n. Thus it remains to handle the case of $n = 4$. For this, first note that there are 8 3-cycles in S_4, each of which are elements of A_4, so it follows that the 3-cycles necessarily generate all of A_4 by Lagrange's Theorem (the subgroup generated by them has size at least 9 while the group itself has order 12, and so since Lagrange implies the order of the subgroup they generate will divide the size of the group, 12, it follows they generate the whole group). Then given a 3-cycle (abc), we have that $[(ac),(bc)] = (ac)(bc)(ac)(bc) = (acb)(acb) = (abc)$, proving that each 3-cycle is in fact a commutator. Thus we conclude that A_4 is the commutator subgroup in this case as well. Hence the commutator subgroup of S_n is always A_n for any $n > 1$.

In fact, we have a **universal property for the commutator subgroup**:

Problem 3.14 (6 Points)**.** Let H be an abelian group and $\phi : G \to H$ be a homomorphism. Let p be the projection map $G \to G^{\text{ab}}$ (given by $g \to gG'$). Show there exists a unique map $h : G^{\text{ab}} \to H$ such that $\phi = h \circ p$.

Universal properties are quite useful because the ensure the existence of something up to (unique) isomorphism, which gives a very clean construction in a variety of cases. In this case, G^{ab} is the universal object in the category of groups for maps from a group G into abelian groups because by the above exercise, and the map factors through it ("goes through it") to get to H.

4 Representation Theory

Representation theory studies the action of (finite) groups on vector spaces over a field. While representation theory can be conducted over any field in general, for simplicity, we fix our field to be $\mathbb{C}$, the complex numbers. **Throughout this section, we take G to be a finite group. Unless otherwise stated, all vector spaces are finite dimensional over $\mathbb{C}$.**

4.1 Introduction

Before we dive into representation theory, we first introduce two important examples of groups related to finite dimensional vector spaces.

Let V be a finite dimensional vector space over $\mathbb{C}$. Recall from Section 1 that given any two linear isomorphism $T, T' : V \to V$, their composition $T \circ T' : V \to V$ is again another linear isomorphism. The composition of linear maps is associative, and any linear isomorphism has an inverse which is also a linear isomorphism. Taking the identity map id_V as the identity element, the set of all linear isomorphisms $V \to V$ forms a group under composition. This group is called the *general linear group* of V and denoted $GL(V)$.

Recall that under a fixed choice of ordered basis, any linear transformation $T : V \to V$ corresponds to a matrix $M_T \in \mathrm{Mat}_{n\times n}(\mathbb{C})$. Fix a basis of V and consider the map $GL(V) \to \mathrm{Mat}_{n\times n}(\mathbb{C})$ defined by taking a linear transformation $T : V \to V$ to its corresponding matrix M_T with respect to the given basis. You should convince yourself that T is a linear isomorphism if and only if M_T is an invertible matrix, hence the image is precisely the set of invertible $n \times n$ matrices over $\mathbb{F}$. The group structure of $GL(V)$ also carries over with the identity matrix I as the identity element and matrix multiplication as the group operation. As a group, we denote the set of invertible $n \times n$ matrices over $\mathbb{C}$ by $GL_n(\mathbb{C})$. Observe that the mapping $T \mapsto M_T$ then gives a group isomorphism $GL(V) \to GL_n(\mathbb{C})$.

Definition 4.1. Let V be a finite dimensional $\mathbb{C}$-vector space. A *representation of G on V* is a group homomorphism $\varphi : G \to GL(V)$. We also call V together with φ, written as a pair (V, φ), a *G-representation* or simply a *representation*. We will sometimes omit φ for brevity when it is clear that we are referring to V as a representation and not just a vector space. The *degree* or *dimension* of the representation is the dimension of V over $\mathbb{C}$. We say a representation (V, φ) is *faithful* if φ is injective. In other words, (V, φ) is a faithful representation if and only if the only element of G that gets mapped to the identity map on V is the identity element.

Problem 4.1 (2 Points)**.** Prove the following.

(a) (1 Point) Any representation $\varphi : G \to GL(V)$ defines a corresponding group action of G on the set V by $g \cdot v = \varphi(g)(v)$ for all $g \in G, v \in V$.

(b) (1 Point) The action defined in (a) is compatible with the vector space structure of V, i.e. for any $v_1, v_2 \in V$, $\lambda_1, \lambda_2 \in \mathbb{C}$, and $g \in G$,

$$g \cdot (\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 (g \cdot v_1) + \lambda_2 (g \cdot v_2) \tag{4.1}$$

The converse to Problem 4.1 is also true, that is any group action of G on V such that (4.1) holds determines a representation of G on V. Let us see why this is. Let n be the dimension of V over $\mathbb{C}$ and $\{e_1, \ldots, e_n\}$ a basis. For any element $g \in G$, define a linear map $M_g : V \to V$ by setting $M_g(e_i) := g \cdot e_i$ for $1 \le i \le n$. By (4.1), $M_g(v) = g \cdot v$ for all $v \in V$. As $g \cdot (g^{-1} \cdot v)) = v = g^{-1} \cdot (g \cdot v)$ for all $v \in V$, we see that M_g is invertible, hence $M_g \in GL(V)$. Define $\varphi : G \to GL(V)$ by $\varphi(g) = M_g$. As $(gh) \cdot v = g \cdot (h \cdot v)$ for any $g, h \in G$, $M_{gh} = M_g M_h$, so φ is a group homomorphism, hence a representation. We thus see that giving a representation of G on V is equivalent to giving a group action of G on V that such that (4.1) holds. Throughout the rest of this section, we will make use of this equivalent definition for a G-representation.

We now give some standard examples of representations. Recall that for any set S, $\mathbb{C}[S]$ is the complex vector space with basis given by the elements of S. In particular, we will denote by $\mathbb{C}[G]$ the vector space over $\mathbb{C}$ on the underlying set of G.

Example 4.1 (Trivial representation)**.** Consider the trivial homomorphism $\varphi : G \to GL(\mathbb{C})$ given by $\varphi(g) = \mathrm{id}_{\mathbb{C}}$ for all $g \in G$. This is a degree 1 representation called the *trivial representation* of G. When $|G| > 1$, the trivial representation is not faithful.

Example 4.2 (Regular representation)**.** Take $V = \mathbb{C}[G]$. We define an action of G on V by setting $g \cdot h = gh$ for all $g, h \in G$ and extending by linearity using (4.1). By the above, this gives a representation $\varphi : G \to GL(\mathbb{C}[G])$, which is called the *regular representation* of G. The regular representation is of degree $|G|$ and is faithful as any nonidentity element of G corresponds to a permutation of the basis of V which is not the identity.

Example 4.3 (Representation from a group action)**.** Let S be a finite set equipped with a G action. Consider the vector space $V = \mathbb{C}[S]$. We may then define an action of G on V by extending the action of G on S using (4.1). This endows V with a G-representation structure. In particular, if the action of G on S is faithful, the resulting representation will also be faithful.

Recall that we may identify $GL(V)$ with $GL_n(\mathbb{C})$, the group of invertible $n \times n$ matrices over $\mathbb{C}$. A *matrix representation* of G of dimension n is then a group homomorphism $\varphi : G \to GL_n(\mathbb{C})$. Note that (up to a choice of basis) we have a one-to-one correspondence between representations of V and matrix representations of V by post-composing with the appropriate isomorphism between $GL(V)$ and $GL_n(\mathbb{C})$. Thus, giving a matrix representation of V is equivalent to giving a representation of V, up to a choice of basis for V. Often times when computing specific representations of a group, it is easier to work with matrix representations.

Example 4.4. Let $G = S_3$. We may define a 2-dimensional matrix representation $\varphi : S_3 \to GL_2(\mathbb{C})$ by

$$\varphi((1\ 2)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \varphi((1\ 3)) = \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix}, \quad \varphi((2\ 3)) = \begin{pmatrix} -1 & 0 \\ -1 & 1 \end{pmatrix}$$

$$\varphi((1\ 2\ 3)) = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}, \quad \varphi((3\ 2\ 1)) = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$$

Example 4.5 (Representations of dimension 1)**.** When V is a vector space of dimension 1, all linear transformations of V to itself are given by scalar multiplication by some complex number. The invertible elements are thus the elements of $\mathbb{C}^\times = \mathbb{C} \setminus \{0\}$, the multiplicative group of $\mathbb{C}$, hence we may identify $GL(V) = \mathbb{C}^\times$. In particular, the matrix of a linear map $f : V \to V$ is a well-defined complex number $\alpha \in \mathbb{C}^\times$ for any choice of basis of V! Thus, the 1-dimensional representations of a finite group G are "equivalent" to group homomorphisms $\varphi : G \to \mathbb{C}^\times$.

Example 4.6 (Permutation representation)**.** Let V be an n-dimensional vector space over $\mathbb{C}$ and fix a basis $\{e_1, \ldots, e_n\}$. We define an action of S_n on V by setting $\sigma \cdot e_i = e_{\sigma(i)}$ for all $\sigma \in S_n, 1 \leq i \leq n$ and extending by linearity. The corresponding representation $\varphi : S_n \to GL(V)$ is called a *permutation representation.* Note that this is a specialization of the above example (obtaining a G-representation from an action of G on a set S). All permutation representations are faithful.

Using the same ordered basis $\{e_1, \ldots, e_n\}$, the representation φ corresponds to a matrix representation $\psi : S_n \to GL_n(V)$. Observe that any element of S_n then maps to a matrix that has exactly one 1 in every row and column and 0's elsewhere. Such matrices are called permutation matrices as they permute the basis vectors.

Given any group G and a finite set S equipped with a G-action, we may put Examples 4.3 and 4.6 together to obtain a group homomorphism $f : G \to S_n$, where $n = |S|$. This is another way to construct the induced group homomorphism discussed in Section 3.1!

Problem 4.2 (3 Points)**.** Let G be a group and S a finite set equipped with a G-action. Fix an ordering of the elements of $S = \{s_1, \dots, s_n\}$.

(a) (1 Point) Construct a matrix representation $\varphi : G \to GL_n(\mathbb{C})$ with respect to the ordered basis $s_1, \dots, s_n$ corresponding to the representation $\mathbb{C}[S]$ (see Example 4.3).

(b) (1 Point) Let $\psi : S_n \to GL_n(\mathbb{C})$ be the matrix representation for the permutation representation from Example 4.6. Show that $\operatorname{im}(\varphi)$ is a subgroup of $\operatorname{im}(\psi)$.

(c) (1 Point) Recall that ψ is injective, hence we may define a map $f : G \to S_n$ by $f(g) = \psi^{-1}(\varphi(g))$. Show that f is a group homomorphism.

In previous sections, we discussed linear maps between vector spaces and group homomorphisms between groups. We may combine the two to define G-linear maps between representations. Intuitively speaking, a G-linear map is a linear map of vector spaces that preserves the representation structure.

Definition 4.2. Let V, W be two finite-dimensional complex vector spaces and $\varphi : G \to GL(V)$, $\psi : G \to GL(W)$ representations on V and W respectively. A *G-linear map* is a linear map $T : V \to W$ satisfying

$$T \circ \varphi(g) = \psi(g) \circ T \tag{4.2}$$

for all $g \in G$, i.e. a linear map that preserves the G-action given by the representations. We say that the two representations φ, ψ are *equivalent* or *isomorphic* if there exists an invertible G-linear map between them, i.e. an invertible linear map $T : V \to W$ such that

$$T \circ \varphi(g) \circ T^{-1} = \psi(g) \tag{4.3}$$

for all $g \in G$. The map T is also called a *G-linear isomorphism*. We denote by $\operatorname{Hom}_G(V, W)$ the set of G-linear maps $V \to W$.

Problem 4.3 (3 Points)**.** Let V, W, Z be G-representations. Verify the following properties of G-linear maps.

(a) (1 Point) The identity $\mathrm{id}_V : V \to V$ is a G-linear map.

(b) (1 Point) If $T : V \to W, S : W \to Z$ are two G-linear maps, then $S \circ T : V \to Z$ is also a G-linear map.

(c) (1 Point) If $T : V \to W$ is an invertible G-linear map, then so is $T^{-1} : W \to V$, the linear transformation inverse.

Remark. As a brief aside, the previous problem tells us that we may define the category of G-representations, and equivalent G-representations are precisely those that are isomorphic as objects of the category.

Problem 4.4 (4 Points)**.** Let $(V, \varphi), (W, \psi)$ be two G-representations.

(a) (1 Point) Show that if G is abelian, then for all $T \in \operatorname{Hom}_G(V, W)$ and $g \in G$, $T \circ \varphi(g) \in \operatorname{Hom}_G(V, W)$.

(b) (1 Point) Find an example of a finite non-abelian group G and two finite-dimensional G-representations $(V, \phi), (W, \psi)$ such that $T \circ \varphi(g) \in \operatorname{Hom}_G(V, W)$ for all $T \in \operatorname{Hom}_G(V, W)$ and $g \in G$.

(c) (2 Points) Suppose G is a finite group such that for any two finite-dimensional G-representations $(V, \phi), (W, \psi)$, the condition $T \circ \varphi(g) \in \operatorname{Hom}_G(V, W)$ for all $T \in \operatorname{Hom}_G(V, W)$ and $g \in G$ holds. Show that G must be abelian.

Recall that for two vector spaces V, W, $\mathrm{Hom}(V,W)$ is the set of all linear maps $V \to W$ and carries a vector space structure. The following problem shows that $\mathrm{Hom}(V,W)$ may be endowed with a G-representation structure when V, W are both G-representations.

Problem 4.5 (5 Points)**.** Let $(V,\varphi),(W,\psi)$ be two G-representations.
(a) (1 Point) Check that $g \cdot T = \psi(g) \circ T \circ \varphi(g)^{-1}$ defines a group action of G on $\mathrm{Hom}(V,W)$.
(b) (1 Point) Show that the action of G on $\mathrm{Hom}(V,W)$ given by part (a) makes $\mathrm{Hom}(V,W)$ into a G-representation.
(c) (1 Point) Show that $\mathrm{Hom}_G(V,W)$ is a vector subspace of $\mathrm{Hom}(V,W)$.
(d) (2 Points) Define $\mathrm{Hom}(V,W)^G$ to be the subset of linear maps in $\mathrm{Hom}(V,W)$ that are fixed by the action defined in (a). Check that $\mathrm{Hom}(V,W)^G = \mathrm{Hom}_G(V,W)$. We say that $\mathrm{Hom}(V,W)^G$ is a *G-invariant subspace* of $\mathrm{Hom}(V,W)$.

4.2 Irreducible Representations

The main focus of this section is on irreducible G-representations (defined below), which can be thought of as the building blocks for all G-representations. In particular, every G-representation admits a direct sum decomposition into irreducible representations. In the later section on character theory, we shall see that this decomposition is unique up to equivalence, i.e. two G-representations are equivalent if and only if they have the same irreducible representation decomposition (up to reordering and equivalence of the summands).

Definition 4.3. Let (V,φ) be a G-representation. A vector subspace W of V is called *G-stable* if $\varphi(g)v \in W$ for all $v \in W$. We may then define $\varphi_W : G \to GL(W)$ by $\varphi_W(g)v = \varphi(g)v$ for all $v \in W$. (W,φ_W) is then said to be a *subrepresentation* of (V,φ). As the G-representation structure of W is completely determined by that of V, we will often take φ, φ_W to be implicit and simply define a subrepresentation of a G-representation V to be a G-stable vector subspace of V. In this case, one just says that W is a subrepresentation of V.

We give examples of subrepresentations in the exercises below.

Problem 4.6 (2 Points)**.** Let (V,φ) be a G-representation. Define

$$V^G := \{v \in V | \varphi(g)(v) = v \text{ for all } g \in G\}.$$

Show that V^G is a subrepresentation of V. We call V^G the *G-invariant subrepresentation* of V.

Example 4.7. In the context of problem 4.5, $\mathrm{Hom}_G(V,W)$ is the G-invariant subrepresentation of $\mathrm{Hom}(V,W)$.

Problem 4.7 (3 Points)**.** Let $(V,\varphi),(W,\psi)$ be two G-representations and $T : V \to W$ a G-linear map.
(a) (1 Point) Show that $\ker(T)$ is a subrepresentation of V.
(b) (2 Points) Show that $\mathrm{im}(T)$ is a subrepresentation of W.

Definition 4.4. A G-representation V is *irreducible* if it is nontrivial as a vector space (i.e. has dimension great than 0) and does not have any nontrivial, proper (as a subspace) subrepresentations. We say V is *reducible* if it is a nontrivial vector space that is not irreducible.

Example 4.8. All dimension 1 G-representations are irreducible as they have no nontrivial, proper subspaces. The converse is true when G is abelian, as we will see in the next section.

Problem 4.8 (4 Points)**.** Verify (without using character theory) that the matrix representation of S_3 given in Example 4.4 is an irreducible representation.

Recall that a vector space V is a direct sum of two subspaces U, W if $U \cap W$ is the trivial subspace and every vector $v \in V$ admits a decomposition $v = u + w$ with $u \in U, w \in W$. We denote the direct sum decomposition by $V = U \oplus W$. Similarly, a G-representation V is a direct sum of two subrepresentations U, W if $V = U \oplus W$ as vector spaces.

Example 4.9. Let U, W be subspaces of a finite-dimensional vector space V such that $V = U \oplus W$. Let $\varphi_1 : G \to GL(U), \varphi_2 : G \to GL(W)$ be representations on U, W respectively.

Fix bases $\{u_1, \dots, u_n\}$ of U and $\{w_1, \dots, w_m\}$ of W. Let $\psi_1 : G \to GL_n(\mathbb{C})$ and $\psi_2 : G \to GL_m(\mathbb{C})$ be the corresponding matrix representations of U, W respectively.

We then define a matrix representation $\psi : G \to GL_{n+m}(V)$ by the block diagonal matrix

$$\psi(g) = \begin{pmatrix} \psi_1(g) & \mathbf{0} \\ \mathbf{0} & \psi_2(g) \end{pmatrix}.$$

Here the notation means that writing the element of a matrix M in the ith row and jth column as $M_{i,j}$, we have

$$\psi(g)_{i,j} = \begin{cases} \psi_1(g)_{i,j} & 1 \leq i, j \leq n \\ \psi_2(g)_{(i-n),(j-n)} & n+1 \leq i, j \leq m \\ 0 & \text{otherwise} \end{cases}.$$

In particular, each bolded $\mathbf{0}$ is a "sub"-matrix, where every entry is the scalar $0 \in \mathbb{C}$ (do not confuse the $\mathbf{0}$ here with the $\mathbf{0}$ element of a vector space). More generally, a block diagonal matrix is a matrix that may be written in terms of submatrices as above with the nontrivial blocks on the diagonal.

Observe that $\{u_1, \dots, u_n, w_1, \dots, w_m\}$ is a basis of V. Then, U, W with their respective representations are subrepresentations of V with representation given by ψ.

Recall that given a finite-dimensional vector space V and a subspace U, we may always find another subspace W such that $V = U \oplus W$. The analogous statement is true for G-representations and is called Maschke's Theorem. Maschke's Theorem allows us to decompose representations into a direct sum of irreducible subrepresentations. This result is the foundation of character theory and is key to determining whether or not two representations are equivalent.

Theorem 4.5. *(Maschke's Theorem)* Let (V, φ) be a G-representation. If (U, φ_U) is any subrepresentation of V, then V has a subrepresentation (W, φ_W) such that $V = U \oplus W$.

Problem 4.9 (10 Points)**.** Prove Maschke's Theorem:

(a) (1 Point) Show that there exists a (not necessarily G-stable) subspace W_0 of V such that $V = U \oplus W_0$.

(b) (2 Point) Define $\pi_0 : V \to U$ by $\pi_0(u+w) = u$ for $u \in U, w \in W_0$. For each $g \in G$, define

$$g\pi_0 g^{-1} : V \to U \qquad v \mapsto \varphi_U(g)\pi_0(\varphi(g^{-1})v).$$

Verify that $g\pi_0 g^{-1}(u) = u$ for all $u \in U$.

(c) (4 Points) Define $\pi : V \to U$ by

$$\pi(v) = \frac{1}{|G|}\sum_{g\in G} g\pi_0 g^{-1}(v)$$

Prove the following properties of π.

(i) π is a linear transformation

(ii) $\pi(u) = u$ for all $u \in U$

(iii) $\pi(\pi(v)) = \pi(v)$ for all $v \in V$

(iv) π is a G-linear map

(d) (3 Points) Let $W = \ker(\pi)$. Show that W is a G-stable subspace of V and $V = U \oplus W$.

Problem 4.10 (3 Points)**.** Using Maschke's theorem, show that any G-representation V may be decomposed into a direct sum of irreducible subrepresentations.

We will see in the following section using character theory that this decomposition into irreducible subrepresentations is unique up to reordering and equivalence of the summands. A key result to studying irreducible representations is Schur's Lemma, which gives restrictions on the allowed G-linear maps between any two irreducible representations.

Theorem 4.6. *(Schur's Lemma)* Let $(V,\varphi), (W,\psi)$ be two irreducible G-representations.

(i) If $T : V \to W$ is a nontrivial G-linear map (i.e. there exists $v \in V$ such that $T(v) \neq 0$), then T is a G-linear isomorphism.

(ii) If $T : V \to V$ is a nontrivial G-linear map, then $T = \lambda I$ for some $\lambda \in \mathbb{C}$ where I is the identity map.

Remark. As a consequence of Schur's Lemma, given any two irreducible G-representations V,W, either V and W are equivalent or $\text{Hom}_G(V,W) = \{0\}$, where 0 denotes the zero map.

Problem 4.11 (6 Points)**.** Prove Schur's Lemma.

(a) (2 Point) Prove part (i) of Schur's Lemma. *Hint: use Problem 4.7.*

(b) (4 Points) Prove part (ii) of Schur's Lemma. *Hint: use something you learned from Section 1.*

4.3 Abelian Groups

A common question one may ask when given a finite group G is to compute all the irreducible G-representations up to equivalence. In the next section, we will describe general methods and compute explicit examples using character theory. As a precursor, in this subsection, we discuss some tools for doing so in relation to finite abelian groups. The representation theory of finite abelian groups is far simpler than that of finite arbitrary groups. We first illustrate this with the following problem, which is an application of Schur's Lemma.

Problem 4.12 (3 Points)**.** Let G be a finite abelian group. Prove that if (V,φ) is an irreducible G-representation, then V must be of dimension 1.

Remark. This result also holds for infinite abelian groups.

Now, let G be an arbitrary finite group. Let V be a 1-dimensional representation of G. Recall from Example 4.5 that we may identify $GL(V)$ with $\mathbb{C}^\times$, hence any representation $\varphi : G \to GL(V)$ induces an equivalent representation $\varphi' : G \to \mathbb{C}^\times$. Conversely, any homomorphism $\varphi : G \to \mathbb{C}^\times$ gives an equivalent representation of G on V using the same identification $GL(V) \cong \mathbb{C}^\times$. We leave it as an exercise to check that any two distinct homomorphisms $G \to \mathbb{C}^\times$ give rise to inequivalent representations. We thus see that the inequivalent 1-dimensional G representations are in one-to-one correspondence with group homomorphisms $G \to \mathbb{C}^\times$.

Recall that we may associate to G the abelian group G/G', where G' is the commutator subgroup. By the universal property of abelianization, any homomorphism $\varphi : G \to C^\times$ factors through the projection map $\pi : G \to G/G'$, i.e. there exists a unique group homomorphism $\psi : G/G' \to C^\times$ such that $\psi \circ \pi = \varphi$. We thus may associate to any 1-dimensional representation of G a unique corresponding 1-dimensional representation of G/G'. Conversely, given a 1-dimensional representation $\psi : G/G' \to \mathbb{C}^\times$, we may take the composition $\psi \circ \pi : G \to \mathbb{C}^\times$ to get a 1-dimensional representation of G. Altogether we have the following result.

Proposition 4.7. Let G be an arbitrary finite group. Then the inequivalent one-dimensional representations of G are in one-to-one correspondence with the inequivalent one-dimensional representations of the abelianization G/G'.

We will see examples of how Proposition 4.7 may be used to compute the 1-dimensional irreducible representations of a group in the following section. We conclude this section with a result that allows us to restrict the dimensions of the irreducible representations of a finite group G based on its abelian subgroups.

Problem 4.13 (7 Points)**.** Let G be an arbitrary finite group and H an abelian subgroup of G. Show that if (V,φ) is a finite irreducible G-representation, then $\dim_{\mathbb{C}}(V) \leq [G:H]$. Note that this generalizes the result of Problem 4.12.

Example 4.10. The symmetric group S_3 has a cyclic subgroup of order three given by $H = \{(123),(132),(1)\}$. As $[S_3 : H] = 2$, all irreducible representations of S_3 must be of either dimension 1 or dimension 2.

Example 4.11. Any dihedral group D_n (where $|D_n| = 2n$) has a cyclic subgroup of index 2 given by the rotations, so all irreducible representations of D_n are of dimension at most 2.

5 Character Theory

In this section, all groups denoted by G are finite, and all vector spaces and representations are over $\mathbb{C}$ and finite dimensional, unless otherwise stated.

We have seen that it is in general difficult to find or classify the irreducible representations of a (finite) group. We will now introduce the basics of *character theory*. The character of an irreducible representation is a numerical invariant that encodes the isomorphism type of a representation. In this way, it is possible to give a description of all the irreducible representations of a finite group without explicitly writing them out.

5.1 Introduction

We begin by introducing a few more concepts from linear algebra.

Definition 5.1. Let $n \geq 1$ be an integer, and let A be an $n \times n$ matrix with values in $\mathbb{C}$. The *trace* of A, denoted by $\mathrm{Tr}(A)$, is the sum of the diagonal values of A, i.e. $\mathrm{Tr}(A) = A_{1,1} + A_{2,2} + \cdots + A_{n,n}$.

Let V be an n-dimensional vector space over $\mathbb{C}$, and $f : V \to V$ a linear map. The *trace* of f, denoted by $\mathrm{Tr}(f)$ is the trace of the matrix of f with respect to any basis of V.

Example 5.1. The trace of $\left(\begin{smallmatrix} 1 & 2 & 3 \\ 5 & 4 & 6 \\ 7 & 8 & 9 \end{smallmatrix}\right)$ is $1 + 4 + 9 = 14$.

Remark. We need to check that the above definition of $\mathrm{Tr}(f)$ is well-defined, i.e. does not depend on the choice of basis of V. Indeed, suppose $v_1, \ldots, v_n$ and $w_1, \ldots, w_n$ are two bases of V. Every vector w_j may be written in terms of the basis of v_i's: let A be the (invertible) $n \times n$ matrix defined by $v_j = \sum_{i=1}^{n} A_{i,j} w_i$. In other words, A is the *change of basis matrix* from the v_i basis to the w_j basis. Then, if P is the matrix of f with respect to $v_1, \ldots, v_n$, then APA^{-1} is the matrix of f with respect to $w_1, \ldots, w_n$. We just need to check $\mathrm{Tr}(P) = \mathrm{Tr}(APA^{-1})$:

Problem 5.1 (3 Points)**.** Let $n \geq 1$ be an integer.
(a) (2 Points) Let A, B be square $n \times n$ matrices. Prove that $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.
(b) (1 Point) Let P, A be square $n \times n$ matrices with A invertible. Deduce that $\mathrm{Tr}(P) = \mathrm{Tr}(APA^{-1})$.

We are now ready to give the definition of a character:

Definition 5.2. Let G be a finite group.

(1) A (complex) *class function* (of G) is a function $\theta : G \to \mathbb{C}$ that is constant on conjugacy classes of G, i.e. $\theta(xgx^{-1}) = \theta(g)$ for all $x, g \in G$. The set of all class functions of G is denoted by $\mathbb{C}_{\text{class}}(G)$.
(2) Let V be a finite dimensional $\mathbb{C}$-vector space and $\varphi : G \to GL(V)$ a representation. The *character* of the representation (V, φ) is the function $\chi : G \to \mathbb{C}$ given by $\chi(g) = \mathrm{Tr}(\varphi(g))$.

We say that a class function $G \to \mathbb{C}$ is a character if it is the character of some representation of G. We say that a character is *irreducible* if it is the character of an irreducible representation.

Observe that for any representation (V, φ) of G with character χ, we have that $\chi(1) = \dim V$. To this end, we say that $\dim \chi := \chi(1)$ *is the degree or the dimension of the character* χ.

Example 5.2. A one-dimensional representation of a group G is just a homomorphism $\varphi : G \to \mathbb{C}^{\times}$, where (recall) $\mathbb{C}^{\times}$ is the multiplicative group of the nonzero complex numbers. Then, the character of φ is just φ itself! In particular, $\varphi(g)$ *is a complex root of unity for every* $g \in G$.

For any group G, there is the trivial representation: $\varphi : G \to \mathbb{C}^{\times}, g \mapsto 1$. The character of this representation is identically 1 and called the *trivial character* of G.

Example 5.3. The two-dimensional matrix representation of S_3 (denote by $\varphi : S_3 \to GL_2(\mathbb{C})$) determined by

$$\varphi((1\ 2)) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \varphi((1\ 2\ 3)) = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$$

has character $\chi : S_3 \to \mathbb{C}$ given by

$$\chi(1) = 2, \quad \chi((1\ 2)) = \chi((2\ 3)) = \chi((1\ 3)) = 0, \quad \chi((1\ 2\ 3)) = \chi((1\ 3\ 2)) = -1.$$

Example 5.4. Let $S := \{1, 2, \dots n\}$ be a finite set, and G a finite group acting on S. Let $V := \mathbb{C}[S]$ be the $\mathbb{C}$-vector space on S. In the previous section, we saw that $\mathbb{C}[S]$ is a representation of G of dimension $|S|$. The "obvious" basis of V, of course, is the one indexed by the elements of S. Denote by φ the corresponding homomorphism of this representation.

Suppose that $g \cdot j = i$ for some $i, j \in \{1, 2, \dots, n\}, g \in G$. With respect to the above basis of V, the matrix of $\varphi(g)$ has a 1 in row i, column j, and a 0 in every other row of column j. One deduces the character χ of V is given by

$$\chi(g) = \{\#\text{ of } i \in \{1, 2, \dots, n\} \text{ such that } g \cdot i = i\}, g \in G,$$

i.e. the number of fixed points of g acting on S. In the special case where $G = S_n$ with the natural permutation action on $\{1, 2, \dots, n\}$, then the character χ of V is

$$\chi(\sigma) = \{\#\text{ of } i \in \{1, 2, \dots, n\} \text{ such that } \sigma(i) = \sigma\}, \sigma \in S_n,$$

i.e. the number of fixed points of σ.

Problem 5.2 (1 Point)**.** Let χ be the character of the regular representation of G. Verify that for any $g \in G$,

$$\chi(g) = \begin{cases} |G| & g = 1 \\ 0 & g \neq 1 \end{cases}.$$

Now, is not immediately apparent why characters might be so useful, or why they are closely connected to class functions. Start by proving the following basic properties:

Problem 5.3 (2 Points)**.** Let G be a finite group.

(a) (1 Point) Verify that the character of a representation (V, φ) of G is a class function.

(b) (1 Point) Let $(V, \varphi_1), (W, \varphi_2)$ be equivalent representations of G. Prove that the characters of V and W are the same.

5.2 Class Functions and Orthogonality

One of the main results of this Section that we shall prove will be the converse to the above problem, i.e. that *two representations with the same character are equivalent*. Since we have seen that *characters are class functions*, this suggests we might want to study characters of representations by developing the theory of class functions. Namely, we shall introduce an inner product on the space of all class functions, and the *First Orthogonality Relation* will describe how irreducible characters are realized by the inner product.

The precise setup is as follows. Let G be a finite group, and recall that $\mathbb{C}_{\text{class}}(G)$ denotes the set of all class functions $G \to \mathbb{C}$. Notice that for any $\chi, \psi \in \mathbb{C}_{\text{class}}(G)$, and any scalars $\alpha, \beta \in \mathbb{C}$, the "linear combination"

$\alpha\chi+\beta\psi$ defined by $(\alpha\chi+\beta\psi)(g)=\alpha\chi(g)+\beta\psi(g)$ is again a class function $G\to\mathbb{C}$. Thus, $\mathbb{C}_{\text{class}}(G)$ is a vector space over $\mathbb{C}$, and its dimension is finite, equal to the number of conjugacy classes of G. Now, one puts a complex inner product on the set of class functions $G\to\mathbb{C}$. Let $\chi,\psi:G\to\mathbb{C}$ be class functions. Define a function $\langle\cdot,\cdot\rangle_G:\mathbb{C}_{\text{class}}(G)\times\mathbb{C}_{\text{class}}(G)\to\mathbb{C}$ by:

$$\langle\chi,\psi\rangle_G:=\langle\chi,\psi\rangle:=\frac{1}{|G|}\sum_{g\in G}\chi(g)\overline{\psi(g)}$$

Problem 5.4. (2 Points) Verify that $\langle\cdot,\cdot\rangle_G$ as defined above is a complex inner product on the vector space $\mathbb{C}_{\text{class}}(G)$ of class functions $G\to\mathbb{C}$.

Example 5.5. Let χ be a character of a one-dimensional representation of a finite group G, i.e. χ is a group homomorphism $G\to\mathbb{C}^\times$. Since, for every $g\in G$, $\chi(g)$ is a root of unity, we have that $\chi(g)\overline{\chi(g)}=1$. Therefore,

$$\langle\chi,\chi\rangle=\frac{1}{|G|}\sum_{g\in G}\chi(g)\overline{\chi(g)}=\frac{1}{|G|}\sum_{g\in G}1=1.$$

We are ready to state the *First Orthogonality Relation*, the powerful result which states that the irreducible characters of G form a *set of orthonormal vectors* in $\mathbb{C}_{\text{class}}(G)$ with respect to the above inner product. Actually, it is a generalization of the above example!

Theorem 5.3 (First Orthogonality Relation)**.** Let χ,ψ be two irreducible characters of a finite group G. Then, $\chi=\psi$ if and only if they correspond to equivalent representations of G. Furthermore,

$$\langle\chi,\psi\rangle=\begin{cases}1 & \chi=\psi\\ 0 & \chi\neq\psi\end{cases}.$$

Note in particular that the First Orthogonality Relation verifies that the distinct irreducible characters of G are in one-to-one correspondence with (a complete set of) the inequivalent irreducible representations of G. Furthermore, recall that any set of orthogonal vectors in an inner product space are linearly independent (see Section 1). Hence,

Corollary. The set of irreducible characters of a finite group G are linearly independent in $\mathbb{C}_{\text{class}}(G)$. The number of irreducible representations of G is finite and at most the number of conjugacy classes of G, $=\mathbb{C}_{\text{class}}(G)$.

Over the course of Problems 5.5, 5.6, 5.7, we will work through a proof of the First Orthogonality Relation.

Problem 5.5 (6 Points)**.** Let G be a finite group, let (V,φ) be a representation of G, and let χ be the character of V. Recall that

$$V^G := \{v \in V \mid \varphi(g)(v) = v \text{ for all } g \in G\}$$

is a subrepresentation of V, called the G-invariant subspace of V.

(a) (4 Points) Consider the linear map

$$P := \frac{1}{|G|}\sum_{g\in G}\varphi(g) \in \mathrm{Hom}(V,V).$$

Prove that $P = P^2$ and $\mathrm{im}(P) = V^G$. In other words, prove that P is a projection of V onto V^G.

(b) (2 Points) Deduce that

$$\dim V^G = \frac{1}{|G|}\sum_{g\in G}\chi(g).$$

Problem 5.6 (5 Points)**.** Let G be a finite group, and let $(V,\varphi_1),(W,\varphi_2)$ be representations of G. Recall in Problem 4.5 we defined a representation $\varphi : G \to GL(\mathrm{Hom}(V,W))$ as follows: for $g \in G$,

$$\varphi(g) : \mathrm{Hom}(V,W) \to \mathrm{Hom}(V,W)$$

is the linear isomorphism which takes a linear map $f : V \to W$ to the linear map $\varphi_2(g^{-1}) \circ f \circ \varphi_1(g) : V \to W$. We have also seen that $\mathrm{Hom}_G(V,W) = \mathrm{Hom}(V,W)^G$, so that $\mathrm{Hom}_G(V,W)$ is a subrepresentation of $\mathrm{Hom}(V,W)$. Now, let χ_V,χ_W be the characters of V,W, respectively, and let χ be the character of $(\mathrm{Hom}(V,W),\varphi)$. Prove that

$$\chi(g) = \chi_V(g)\chi_W(g^{-1})$$

for all $g \in G$. Deduce that

$$\dim \mathrm{Hom}_G(V,W) = \frac{1}{|G|}\sum_{g\in G}\chi_V(g)\chi_W(g^{-1}).$$

Take notation as in Problem 5.6, now assuming V,W are irreducible and inequivalent. By Schur's lemma, we have

$$\frac{1}{|G|}\sum_{g\in G}\chi_V(g)\chi_W(g^{-1}) = \dim \mathrm{Hom}_G(V,W) = 0, \qquad \frac{1}{|G|}\sum_{g\in G}\chi_V(g)\chi_V(g^{-1}) = \dim \mathrm{Hom}_G(V,V) = 1$$

(namely, $\mathrm{Hom}_G(V,V) \cong \mathbb{C}$ as vector spaces). In particular, we clearly cannot have $\chi_V = \chi_W$. Therefore, the proof of the First Orthogonality Relation is complete once we have verified the following fact:

Problem 5.7 (5 Points)**.** Let (V,φ) be a representation of a finite group G, and χ its character. Then, for any $g \in G$, prove that $\chi(g)$ is a sum of complex roots of unity, and $\chi(g^{-1}) = \overline{\chi(g)}$.

This completes the proof of the First Orthogonality Relation.

5.3 Decomposition of Reducible Characters

Having completed the proof of the First Orthogonality Relation and shown that the irreducible characters of G form a basis of $\mathbb{C}_{\text{class}}(G)$, we shall now discuss its relation to direct sum decompositions of representations into irreducible components. In particular, we shall prove the promised assertion that *two representations with the same character are equivalent*. As usual, let G be a finite group and (V,φ) a finite dimensional representation of G. Suppose that we have a direct sum decomposition

$$V = V_1 \oplus V_2 \oplus \cdots \oplus V_s,$$

where $V_1, V_2, \ldots, V_s$ are subrepresentations of V.

We may understand V and its irreducible components in terms of matrix representations. For each $1 \le i \le s$ we may fix a basis $\mathscr{B}_i$ of V_i, so that the union of $\mathscr{B}_i$'s is a basis $\mathscr{B}$ of V. For each $g \in G$, let $\varphi_i(g)$ be the matrix of $\varphi(g)|_{V_i}$ with respect to $\mathscr{B}_i$. Thus, for each $g \in G$, the matrix of $\varphi(g)$ with respect to $\mathscr{B}$ is the block diagonal matrix

$$\begin{pmatrix} \varphi_1(g) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \varphi_2(g) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \varphi_s(g) \end{pmatrix}.$$

Here, each $\mathbf{0}$ is a "sub"-matrix, where every entry is the scalar $0 \in \mathbb{C}$. Note this is just a minor generalization of Example 4.13. For each $1 \le i \le s$, the matrix representation φ_i is irreducible and equivalent to V_i (obviously). Having fixed a basis $\mathscr{B}$ of V, we readily observe that $\mathrm{Tr}(\varphi(g)) = \mathrm{Tr}(\varphi_1(g)) + \cdots + \mathrm{Tr}(\varphi_s(g))$ for each $g \in G$. In other words: *the character of the representation V of the group G is the sum of the characters of the components appearing in the direct sum decomposition of V.*

Now, we consider a direct sum decomposition of V into *irreducible* subrepresentations; that is to say, each $V_1, V_2, \ldots, V_s$ as defined above are irreducible. *By Maschke's theorem, proven in the previous section, this is always possible.* We claim that the irreducible components of V are uniquely determined up to isomorphism. This claim may be verified using the First Orthogonality Relation:

Problem 5.8 (4 Points)**.** Let (V,φ_1) be a representation of a finite group G, and χ its character. Let $U_1, \ldots, U_r$ be a complete set of inequivalent irreducible representations of G, and let $\chi_1, \ldots, \chi_r$ their respective characters.

(a) (2 Points) Prove that for any direct sum decomposition $V = V_1 \oplus V_2 \oplus \cdots \oplus V_s$, where $V_1, V_2, \ldots, V_s$ are irreducible subrepresentations of V, the number of V_j's equivalent to U_i (for any $1 \le i \le r$) is equal to $\langle \chi, \chi_i \rangle$. Deduce that the character of V is simply

$$\chi = \sum_{i=1}^{r} \langle \chi, \chi_i \rangle \chi_i.$$

(b) (2 Points) Now, let (W,φ_2) be another representation of G, and ψ its character. Prove that if $\chi = \psi$, then V, W are equivalent representations.

Therefore, given any representation V of G, and a complete set of inequivalent irreducible representations $U_1, \ldots, U_r$ of G, we may write

$$V = U_1^{\oplus n_1} \oplus U_2^{\oplus n_2} \oplus \cdots \oplus U_r^{\oplus n_r}$$

for some integers $n_1, \ldots, n_r \ge 0$. This notation is to say that V decomposes as $n_1 + \cdots + n_r$ irreducible components, where exactly n_i components are equivalent to U_i. We say that n_i is the *multiplicity of U_i in V*. Of course, we have already verified using character theory that the integer n_i is uniquely determined, i.e. *n_i does not depend on the choice of direct sum decomposition.*

One of the immediate consequences of the character theory of reducible representations is as follows. Let χ be a character of a group G. Write $\chi = \sum_{i=1}^{r} a_i \chi_i$ as a linear combination of irreducible characters of G.

Then, $\langle \chi, \chi \rangle = a_1^2 + \cdots + a_r^2$ by the First Orthogonality Relation. In particular, $\langle \chi, \chi \rangle = 1$ if and only if χ is irreducible. You might also observe that $\langle \chi, \chi \rangle = 2$ implies that χ is the sum of two distinct irreducible characters of G. This method allows us to study reducible representations of G without even knowing what the irreducible representations/characters are! Let us consider an example in which we exhibit the power of many these tools at once.

Example 5.6. Let W be the (usual) permutation representation of S_3. This means that we are letting S_3 act naturally on three points $\{1,2,3\}$; letting W be the complex vector space on $\{1,2,3\}$, this yields a representation $\varphi : S_3 \to GL(W)$. Let χ be the character of (W, φ). Thus, $\chi(\sigma)$ is the number of fixed points of σ on $\{1,2,3\}$ by Example 5.6. In particular,

$$\langle \chi, \chi \rangle = \frac{1}{6}(3^2 + 3 \cdot 1^2 + 2 \cdot 0) = 2,$$

so, as remarked above, $\chi = \chi_1 + \chi_2$, where χ_1, χ_2 are irreducible characters of S_3. Now, let ψ be the character of the matrix representation of S_3 in Example 5.5. Observe that

$$\langle \psi, \psi \rangle = \frac{1}{6}(2^2 + 3 \cdot 0 + 2 \cdot (-1)^2) = 1, \quad \langle \chi, \psi \rangle = \frac{1}{6}(3 \cdot 2 + 3 \cdot (1 \cdot 0) + 2 \cdot (0 \cdot (-1)) = 1,$$

so matrix representation in Example 5.5 is irreducible and of multiplicity 1 in W. In particular, $\chi = \chi_1 + \psi$ for some irreducible character χ_1. Observe that $\chi_1 = \chi - \psi \equiv 1$, i.e. χ_1 is the trivial character of S_3. This verifeis that W decomposes as the direct sum of the trivial representation of S_3 and an irreducible representation of S_3 of dimension 2.

We have already seen that the irreducible characters $\chi_1, \ldots, \chi_r$ are linearly independent and orthonormal vectors in $\mathbb{C}_{\text{class}}(G)$ via the First Orthogonality Relation. We shall now strengthen this corollary to

Theorem 5.4. Let $\chi_1, \ldots, \chi_r$ be the irreducible characters of a finite group G. Then, $\chi_1, \ldots, \chi_r$ form an orthonormal basis of $\mathbb{C}_{\text{class}}(G)$. In particular, the number of inequivalent irreducible representations of G equals the number of conjugacy classes of G.

Problem 5.9 (8 Points)**.** Prove Theorem 5.4. *Hint: the proof that we have in mind starts as follows. Let $\alpha : G \to \mathbb{C}$ be a class function. For any representation (V, φ) of G, consider the associated linear map*

$$\rho_{\alpha,V} : \sum_{g \in G} \overline{\alpha(g)} \varphi(g).$$

Check that $\rho_{\alpha,V}$ is G-linear.

Recall that for any finite abelian group G, a (finite dimensional) representation V of G is irreducible if and only if it is one-dimensional. Moreover, recall that the one-dimensional representations of any arbitrary finite group G are in one-to-one correspondence with those of G/G', where $G' \leq G$ is the commutator subgroup. Then, we obtain the following consequence to Theorem 5.4:

Corollary. The number of inequivalent one-dimensional representations of any finite group G equals the index $[G : G']$. In particular, any finite abelian group G has $|G|$ inequivalent one-dimensional representations.

5.4 Applications and Character Tables

We are now equipped with the wealth of tools and the theory of representations and characters, especially those that we have developed in the previous sections. Example 5.8 was only one simple application of this theory Now we are ready to apply this theory to prove interesting facts on group representations and study various explicit examples! One of the main tasks you will have in this section is the problem of *classifying all irreducible representations/characters of a finite group* G. Even with the First Orthogonality Relation on hand, this task is, in general, quite challenging (in some cases perhaps impossible without even more theory), so be prepared to use everything you know in the following series of problems!

First, we shall decompose the character of the regular representation of G using the First Orthogonality Relation:

Problem 5.10 (4 Points)**.** Let G be a finite group, and let χ be the character of the regular representation of G.

(a) (2 Points) Let $\chi_1, \dots, \chi_r$ be the complete set of distinct irreducible characters of G. Prove that

$$\chi = \sum_{i=1}^{r} (\dim \chi_i)\chi_i.$$

In particular, every irreducible representation of G has positive multiplicity in the regular representation of G. Deduce the formula

$$|G| = \sum_{i=1}^{r} (\dim \chi_i)^2.$$

(b) (2 Points) Prove that G is abelian if every irreducible representation of G is one-dimensional.

The result of Problem 5.8 is standard in the study of representations of finite groups. In particular, we have that $(\dim \chi)^2 \leq |G|$ for any irreducible character χ of the finite group G. This result, may be strengthened, however, as the following problem states. You probably need character theory to solve it.

Problem 5.11 (5 Points)**.** Let V be any irreducible representation of a finite group G. Show that $(\dim V)^2 \leq [G : Z(G)]$. Here, recall that $Z(G) \leq G$ is the center of G.

The *classification of the irreducible representations of the symmetric group* is just one of many bridges between the fields of algebra and combinatorics. We shall only begin to explore this beautiful theory in this Power Round. This is precisely the topic of the next problem, where we exhibit a well-known family of irreducible representations of S_n, known as the *standard representation of* S_n. Note we have already seen an example of what the standard representation looks like for the small $n = 3$.

Problem 5.12 (8 Points)**.**

(a) (3 Points) Let S be a finite set, and G a finite group acting on S. Let $V := \mathbb{C}[S]$ be the $\mathbb{C}$-vector space on S, the representation of G arising from this group action. Prove that $\dim V^G$ is the number of orbits of G on S.

(b) (5 Points) Let $n \geq 2$, and let W be the (usual) permutation representation of S_n. Prove that we have a direct sum decomposition $W = U \oplus V$ as representations of S_n, where U is (equivalent to) the trivial representation of S_n, and V is an irreducible subrepresentation of W of dimension $n-1$.

We shall now move onto the problem of classifying all irreducible characters of a finite group. To this end, we introduce the *character table of* G, a bookkeeping device to track the data of all the irreducible characters of G. Informally speaking, the character table is a table with r rows and r columns ($r = \#$ of conjugacy classes of G) which tracks all the possible values all the irreducible characters of G take, over the distinct conjugacy classes of G. The formal definition is:

Definition 5.5. Given a finite group G, let $\chi_1, \dots, \chi_r$ be the distinct irreducible characters of G, and let $g_1, \dots, g_r \in G$ be a set of representatives of the conjugacy classes of G. Then, the *character table* of G is the $r \times r$ square matrix

$$(\chi_i(g_j))_{1 \le i,j \le r}.$$

Since the character table of G is an $r \times r$ matrix with values in $\mathbb{C}$, we may study it using techniques from linear algebra. View the rows of this matrix as vectors in $\mathbb{C}^r$. Despite the First Orthogonality Relation, these vectors are in general not orthogonal under the usual *Euclidean inner product of* $\mathbb{C}^r$. However, the formula for the inner product of $\mathbb{C}_{\text{class}}(G)$ is not too different from that of $\mathbb{C}^r$. To this end, we recall what orthogonality/orthonormality means for the Euclidean inner product. Given any $n \times n$ square matrix P, let $P^\dagger$ be the conjugate transpose matrix of P. We have already seen that the rows of P form an orthonormal basis of $\mathbb{C}^n$ if and only if P is invertible with $P^{-1} = P^\dagger$ if and only if the *columns* of P form an orthonormal basis of $\mathbb{C}^n$. The intuition here may be applied to the similar but somewhat different situation of character tables. As the First Orthogonality Relation concerns the *rows* of a character table, we might believe that there is also a *column orthogonality* condition on the character table of G. This is indeed the case, the exact statement given by:

Problem 5.13 (6 Points)**.** Let $\chi_1, \dots, \chi_r$ be the distinct irreducible characters of a finite group G. Prove the *Second Orthogonality Relation*: for any $x, y \in G$, we have

$$\sum_{i=1}^{r} \chi_i(x)\overline{\chi_i(y)} = \begin{cases} 0 & x, y \text{ not conjugate in } G \\ |C_G(x)| & x, y \text{ conjugate in } G \end{cases}.$$

Here, recall that $C_G(x) := \{g \in G \mid gx = xg\}$ is the centralizer subgroup of x in G.

Hence, the columns of any character table are a set of r orthogonal (hence linearly independent) vectors in $\mathbb{C}^r$ with the standard Euclidean inner product, so the character table of any finite group is an invertible matrix! In particular,

Corollary. Elements $x, y \in G$ of a finite group are conjugate if and only if $\chi(x) = \chi(y)$ for all irreducible characters χ of G.

Now, let us illustrate the practical usefulness of the character table. In the character table of G, the rows are usually labelled with the irreducible characters of G, and the columns are usually labelled with the conjugacy classes of G along with their respective sizes. It is particularly helpful to track the sizes of the conjugacy classes to aid with the computation of the inner products of characters. We shall now present some example computations of character tables (a complete character table of G is a complete classification of the irreducible characters of G). In general, the step by step process is as follows (for a finite group G:

(1) Compute the conjugacy classes of G, labelling the rows of the character table of G (in particular, this determines the size of the character table).

(2) Compute the one-dimensional (irreducible) representations of G (i.e., by studying the abelian quotient G/G').

(3) Compute the remaining irreducible characters of G using the First Orthogonality Relation and other facts we know.

Example 5.7. We compute the character table of S_3. The conjugacy classes of S_3 are indexed by the distinct partitions of 3: they are precisely $\{1\}, \{(1\ 2), (2\ 3), (1\ 3)\}, \{(1\ 2\ 3), (1\ 3\ 2)\}$ (in particular, recall two elements of S_3 are conjugate if and only if they have the same cycle type). Since $S_3' = A_3$ (why?), there are $2 = [S_3 : A_3]$ one-dimensional representations of S_3, given by the two distinct homomorphisms $S_3 \to \{\pm 1\} \le \mathbb{C}^\times$; the nontrivial one is called the *sign representation*. We have already seen a two-dimensional irreducible representation of S_3, the standard representation. Hence, the character table of S_3 is

classes:	1	$(1\ 2)$	$(1\ 2\ 3)$
sizes:	1	3	2
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

,

where χ_1, χ_2 are the characters of S_3 corresponding to the trivial and the sign representation, respectively, and χ_3 is the character of the two-dimensional standard representation.

Example 5.8. We compute the character table of Z_4, the cyclic group of order 4. This group is abelian, so there are three conjugacy classes and three inequivalent irreducible representations of Z_4. We see that there are four distinct homomorphisms $Z_4 \to \{\pm 1, \pm i\} \leq \mathbb{C}^\times$, each determined uniquely by the choice of where the generator $\overline{1} \in Z_4$ maps to. This means we have found all irreducible characters, so the character table of Z_4 is

classes:	0	1	2	3
sizes:	1	1	1	1
χ_1	1	1	1	1
χ_2	1	i	-1	$-i$
χ_3	1	-1	1	-1
χ_4	1	$-i$	-1	i

.

Example 5.9. We compute the character table of D_4, the dihedral group of order 8. To write down the elements of D_4, let $r \in D_4$ be the 90° rotation, and let $s \in D_4$ be any flip. Thus, $r^4 = s^2 = 1$, and $rs = sr^{-1}$, and as a set, $D_4 = \{1, r, r^2, r^3, s, sr, sr^2, sr^3\}$. The conjugacy classes of D_4 are $\{1\}, \{r^2\}, \{r, r^3\}, \{s, sr^2\}, \{sr, sr^3\}$, and the commutator subgroup is $D_4' = \{1, r^2\}$. Note D_4/D_4' is isomorphic to $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$, the direct product of the cyclic group of order 2 by itself. In particular, we may define a surjective homomorphism $D_4 \to \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ by mapping r to $(1,0)$ and s to $(0,1)$; this is valid because the images of r and s satisfy the *relations* of D_4. The kernel of this homomorphism is exactly D_4', hence the requested isomorphism. Now, $D_4/D_4' \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ is order 4, and so there are exactly four distinct homomorphisms $D_4 \to \mathbb{C}^\times$. This yields four distinct one-dimensional characters of D_4, as listed in the following incomplete character table of D_4:

classes:	1	r^2	s	r	sr
sizes:	1	1	2	2	2
χ_1	1	1	1	1	1
χ_2	1	1	-1	1	-1
χ_3	1	1	1	-1	-1
χ_4	1	1	-1	-1	1
χ_5					

.

We just need to determine the remaining character, χ_5. We can do so without even finding the corresponding irreducible representation, simply by recalling results from Problem 5.10. We have $\sum_{i=1}^{5} (\dim \chi_i)^2 = 8$, so $\dim \chi_5 = 2$. Now, if χ is the character of the regular representation of D_4, then the fact

$$\chi = \chi_1 + \chi_2 + \chi_3 + \chi_4 + 2\chi_5$$

determines the value of χ on every conjugacy class of G. This completes our computation of the character table of D_4:

classes:	1	r^2	s	r	sr
sizes:	1	1	2	2	2
χ_1	1	1	1	1	1
χ_2	1	1	-1	1	-1 .
χ_3	1	1	1	-1	-1
χ_4	1	1	-1	-1	1
χ_5	2	-2	0	0	0

One can check by hand that for every one of these character tables we have computed so far, the rows satisfy the First Orthogonality Relation, while the columns satisfy the Second Orthogonality Relation.

In the last example, we have shown that it is possible to compute the character table of G without the full knowledge of the irreducible representations of G. Moreover, we have seen that finding irreducible representations of G explicitly are certainly helpful (crucial even) for computing the character table of G. However, given the character table of G, it is in general a far more challenging problem to find an irreducible representation affording each irreducible character.

We conclude this section with some exercises on computing the character tables of some familiar groups of small order. The challenge problem here is part (d).

Problem 5.14 (22 Points)**.** Write down the character tables of the following groups:

(a) (2 Points) Q_8.

(b) (3 Points) D_5.

(c) (5 Points) S_4 and A_4.

(d) (12 Points) A_5.

No explanations are needed for parts (a), (b), (c). A proof is required for part (d).

6 Real Character Tables of A_n and S_n

In this section, we seek to understand when the character tables of A_n and S_n consist of only real entries. We start by characterizing what it means for $\chi_i(g)$ to be real for each $1 \leq i \leq r$.

Problem 6.1 (4 Points)**.** Let $g \in G$ be an element. Show that $\chi_i(g)$ is real for each $1 \leq i \leq r$ iff g and g^{-1} are in the same conjugacy class.

It turns out that the character table of S_n always consists of real entries:

Problem 6.2 (2 Points)**.** Let n be a positive integer. Show that each entry of the character table of S_n is real.

The case of A_n is more subtle, but we can now use the machinery developed to handle it. For A_n, by Problem 6.1, the problem reduces to determining when g and g^{-1} are in the same conjugacy class for all $g \in A_n$. Firstly, we only need to consider certain cycle types:

Problem 6.3 (2 Points)**.** Suppose that $g \in A_n$ has cycle type not consisting of distinct odd integers. Then show that g and g^{-1} are conjugate.

In particular, we only need to consider the g with cycle types consisting of distinct odd integers.

Problem 6.4 (6 Points)**.** Let $g \in A_n$ be any element with cycle type consisting of k distinct odd integers. Show that g and g^{-1} are conjugate in A_n iff $n \equiv k \mod 4$.

With this, we now have all the information we need to complete the investigation of when A_n has real character table:

Problem 6.5 (3 Points)**.** Let $n > 1$ be a positive integer. Show that A_n has real character table iff $n = 2, 5, 6, 10, 14$.

Given a conjugacy class C of a group G, note that the inverses of the elements of C also form a conjugacy class of G. We call this conjugacy class C^{-1} and say that a conjugacy class C is self-inverse if $C = C^{-1}$. Let r be the number of conjugacy classes of a finite group G. From Problem 6.1, we know that a finite group G has r real irreducible characters iff it has real character table iff g and g^{-1} are conjugate for each $g \in G$ iff G has r self-inverse conjugacy classes. We can prove a stronger result of this form:

Problem 6.6 (10 Points)**.** Show that for any finite group G, the number of real irreducible characters of G is equal to the number of self-inverse conjugacy classes of G.