

THE CLASSICAL GROUPS

KEVIN MCGERTY

Date: March, 2006.

1. INTRODUCTION

These notes are the result of teaching Math 241 “Topics in Geometry” in the Spring of 2006 at the University of Chicago. They are a study of matrix groups and some of the geometry attached to them. Of course “geometry” is not a technical term, and in order to keep the prerequisites to a minimum the word is used in an essentially combinatorial sense here – the “geometry” of projective space for example is the poset of subspaces, not anything more advanced, such as the structure of a manifold or algebraic variety (though we describe things in enough detail that a more knowledgeable student should easily be able to find such a structure if they know the appropriate definitions).

We begin by studying isometries of \mathbb{R}^n , focusing on the low dimensional cases of $n = 2, 3$. We then discuss the quaternions and explain the connection with $\text{SO}(\mathbb{R}^3)$. Motivated by this, we classify composition algebras over \mathbb{R} , and define the compact classical groups as matrix groups of \mathbb{R} , \mathbb{C} and \mathbb{H} preserving the appropriate Hermitian form. We then formulate a description of these groups as subgroups of $\text{GL}(\mathbb{C}^n)$, allowing us to obtain the noncompact analogs in terms of bilinear forms, which make sense over any field. We then briefly study projective spaces, which have a rich enough structure that in the next section we can use the action of PGL_n on them to show that the projective special linear group is almost always simple. We end with a brief discussion of bilinear forms and the symplectic group.

The prerequisites are few: some basic linear algebra (the determinant is the most sophisticated notion used), and elementary group theory. In the latter sections, knowledge of finite fields is used, but only the most basic facts. When working with groups over \mathbb{R} , some basic notions of analysis are also mentioned (for example the connectedness $\text{SO}(\mathbb{R}^n)$ is discussed). That said, at various points we attempt to make connections to more advanced topics, but these are for “cultural enrichment” only.

Contents:

- Isometries of \mathbb{R}^n .
- The Quaternions.
- Composition Algebra.
- The Classical Groups.
- Projective Geometry.
- The General Linear Group.
- Bilinear Forms.
- The Symplectic Group.

2. ISOMETRIES OF \mathbb{R}^n

In this course we will study interactions between geometry and group theory. Our goal in this section is to examine a simple example – a group of symmetries of three dimensional Euclidean space. The symmetries we consider are the rigid motions of space which fix a point (the origin $\mathbf{0}$). By a rigid motion, we mean a transformation of space which preserves distance and orientation (we will return shortly to what we mean by the term orientation).

To start with we will work in \mathbb{R}^n for any n . Recall that the notions of distance and angle in \mathbb{R}^n can be given by the dot product:

$$(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^n v_i w_i$$

where the distance $\|\mathbf{v} - \mathbf{w}\|$ between \mathbf{v} and \mathbf{w} is just $\|\mathbf{v} - \mathbf{w}\|^2 = (\mathbf{v} - \mathbf{w}) \cdot (\mathbf{v} - \mathbf{w})$. An *isometry* of \mathbb{R}^n is a bijection $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ which preserves distances. A basis $\{w_1, w_2, \dots, w_n\}$ of \mathbb{R}^n of unit vectors which are mutually perpendicular is called an *orthonormal basis*, thus in terms of the inner product $\{w_1, w_2, \dots, w_n\}$ is an orthonormal basis if

$$\mathbf{w}_i \cdot \mathbf{w}_j = \delta_{ij} = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{otherwise.} \end{cases}$$

(indeed it follows automatically from this condition that the vectors form a basis of \mathbb{R}^n , since they must be linearly independent). For the proof of the next proposition we use the fact that, given any $\mathbf{v} \in \mathbb{R}^n$ and orthonormal basis $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n\}$ we have

$$\mathbf{v} = \sum_{i=1}^n (\mathbf{v} \cdot \mathbf{w}_i) \mathbf{w}_i.$$

Proposition 2.1. *Let $T: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be an isometry such that $T(\mathbf{0}) = \mathbf{0}$. Then T is a linear map.*

Proof. By assumption we have $\|\mathbf{v} - \mathbf{w}\| = \|T(\mathbf{v}) - T(\mathbf{w})\|$ for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$. Expanding this out in terms of the inner product, and using the fact that $\|T(\mathbf{u})\| = \|\mathbf{u}\|$, and $\|T(\mathbf{v})\| = \|\mathbf{v}\|$, since $T(\mathbf{0}) = \mathbf{0}$, we find that $T(\mathbf{u}) \cdot T(\mathbf{v}) = \mathbf{u} \cdot \mathbf{v}$, that is, T preserves the inner product.

Consider the standard basis of \mathbb{R}^n , $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n\}$. Then we know that the vectors $\{T(\mathbf{e}_1), T(\mathbf{e}_2), \dots, T(\mathbf{e}_n)\}$ must also form an orthonormal basis of \mathbb{R}^n . Set $\mathbf{w}_i = T(\mathbf{e}_i)$, ($1 \leq i \leq n$).

Let $\mathbf{v} = \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 + \dots + \lambda_n \mathbf{e}_n$ be a vector in \mathbb{R}^n . Then $\lambda_i = \mathbf{v} \cdot \mathbf{e}_i$. Hence we have $T(\mathbf{v}) \cdot \mathbf{w}_i = T(\mathbf{v}) \cdot T(\mathbf{e}_i) = \lambda_i$, and so since the $\{\mathbf{w}_i\}_{1 \leq i \leq n}$ are an orthonormal basis, we have

$$T(\mathbf{v}) = \sum_{i=1}^n (T(\mathbf{v}) \cdot \mathbf{w}_i) \mathbf{w}_i = \sum_{i=1}^n \lambda_i \mathbf{w}_i = \sum_{i=1}^n \lambda_i T(\mathbf{e}_i).$$

It follows that T is the linear map $\alpha: \mathbb{R}^n \rightarrow \mathbb{R}^n$ given by $\alpha(\mathbf{e}_i) = \mathbf{w}_i$ and we are done. \square

The isometries of \mathbb{R}^n form a group under composition: the composition of two isometries is clearly an isometry, but it is slightly harder to see that any isometry has an inverse. Given any $\mathbf{v} \in \mathbb{R}^n$ we can define an isometry $T_{\mathbf{v}}$ called a translation, which sends $\mathbf{w} \mapsto \mathbf{w} + \mathbf{v}$, ($\mathbf{w} \in \mathbb{R}^n$). The inverse of $T_{\mathbf{v}}$ is evidently $T_{-\mathbf{v}}$. Given

any isometry γ , we may compose it with the translation $T_{-\gamma(\mathbf{0})}$ to obtain an isometry $\sigma = T_{-\gamma(\mathbf{0})} \circ \gamma$ fixing $\mathbf{0}$. But then σ is an isometry which is also a linear map. Since any isometry must clearly be injective, and an injective linear map from \mathbb{R}^n to itself is automatically a bijection by the rank-nullity theorem, we see that σ is a bijection, and so clearly $\gamma = T_{\gamma(\mathbf{0})} \circ \sigma$ is also. Once we know that any isometry is a bijection, it is easy to see that the inverse of an isometry is automatically an isometry, and so the set of all isometries of \mathbb{R}^n forms a group, $\text{Isom}(\mathbb{R}^n)$.

Since we have shown that any isometry of \mathbb{R}^n fixing $\mathbf{0}$ is linear, it is worth explicitly singling out which linear maps are isometries:

Definition 2.2. Let $O(\mathbb{R}^n)$ be the group of isometries of \mathbb{R}^n fixing $\mathbf{0}$, so that $O(\mathbb{R}^n) \subset GL(\mathbb{R}^n)$. Since

$$(1) \quad \mathbf{v} \cdot \mathbf{w} = \frac{1}{2}(\|\mathbf{v} + \mathbf{w}\|^2 - \|\mathbf{v}\|^2 - \|\mathbf{w}\|^2),$$

a linear map α is in $O(\mathbb{R}^n)$ if and only if, for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ we have $\alpha(\mathbf{v}) \cdot \alpha(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$. That is,

$$O(\mathbb{R}^n) = \{\alpha \in GL(\mathbb{R}^n) : \alpha(\mathbf{v}) \cdot \alpha(\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}, \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n\}.$$

Now the standard basis is orthonormal (i.e. $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, $1 \leq i, j \leq n$), so that if A is the matrix of α with respect to the standard basis this condition is equivalent to the equation $AA^t = I$, (where A^t denotes the transpose of A). $O(\mathbb{R}^n)$ is called the *orthogonal group*, and a matrix A satisfying $AA^t = I$ is called an *orthogonal matrix*.

Remark 2.3. To attempt to have some way of distinguishing them from linear maps, I will denote the group of $n \times n$ invertible matrices as $GL_n(\mathbb{R})$ and the group of orthogonal matrices as $O_n(\mathbb{R})$ etc. though the distinction is usually not terribly important.

Remark 2.4. The set $\Gamma(\mathbb{R}^n)$ of all translations forms a subgroup of $\text{Isom}(\mathbb{R}^n)$. It is isomorphic to \mathbb{R}^n as an abelian group: it is straightforward to check that the map $\mathbf{v} \mapsto T_{\mathbf{v}}$ defined above is an isomorphism. Moreover, it is not hard to see that Γ is a normal subgroup of $\text{Isom}(\mathbb{R}^n)$ and that $\text{Isom}(\mathbb{R}^n)$ is the semidirect product of $\Gamma(\mathbb{R}^n)$ and $O(\mathbb{R}^n)$.

We now wish to introduce the notion of an *orientation* for a basis of \mathbb{R}^n . We start with the cases of \mathbb{R}^2 and \mathbb{R}^3 , where we may use the vector product:

Definition 2.5. Define a product $\times : \mathbb{R}^3 \times \mathbb{R}^3 \rightarrow \mathbb{R}^3$ by

$$\mathbf{v} \times \mathbf{w} = (v_2w_3 - v_3w_2, v_3w_1 - v_1w_3, v_1w_2 - w_2v_1)$$

We list some basic properties of the vector product that are easy to check directly from the definition.

Lemma 2.6. Let $\mathbf{v}, \mathbf{w}, \mathbf{u} \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$. Then the vector product satisfies:

- (1) $\mathbf{v} \times \mathbf{w} = -\mathbf{w} \times \mathbf{v}$;
- (2) $(\mathbf{v}_1 + \mathbf{v}_2) \times \mathbf{w} = \mathbf{v}_1 \times \mathbf{w} + \mathbf{v}_2 \times \mathbf{w}$;
- (3) $(\lambda \mathbf{v}) \times \mathbf{w} = \mathbf{v} \times (\lambda \mathbf{w}) = \lambda(\mathbf{v} \times \mathbf{w})$;
- (4) $\|\mathbf{v} \times \mathbf{w}\|^2 + |\mathbf{v} \cdot \mathbf{w}|^2 = \|\mathbf{v}\|^2 \|\mathbf{w}\|^2$, and so $\|\mathbf{v} \times \mathbf{w}\| = \|\mathbf{v}\| \|\mathbf{w}\| \sin(\theta)$, where θ is the (acute) angle between the vectors \mathbf{v} and \mathbf{w} .
- (5) $\mathbf{v} \cdot (\mathbf{v} \times \mathbf{w}) = 0$.
- (6) If $\alpha \in O(\mathbb{R}^3)$, and $\det(\alpha) = 1$, then $\alpha(\mathbf{v} \times \mathbf{w}) = \alpha(\mathbf{v}) \times \alpha(\mathbf{w})$.

(7) If $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ denote the standard basis vectors then we have $\mathbf{e}_i \times \mathbf{e}_{i+1} = \mathbf{e}_{i+2}$ where we read the indices modulo 3.

Proof. Parts (1), (2), (3), (5) and (7) are routine calculations, while (4) is a straightforward, though somewhat more elaborate one. For part (6), note first that if $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^3$ and A is the 3×3 matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, then

$$\det(A) = (\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3.$$

Now given $\mathbf{v}, \mathbf{w} \in \mathbb{R}^3$, the vectors $\alpha(\mathbf{v}) \times \alpha(\mathbf{w})$ and $\alpha(\mathbf{v} \times \mathbf{w})$ are equal if and only if $(\alpha(\mathbf{v}) \times \alpha(\mathbf{w})) \cdot \mathbf{e} = \alpha(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{e}$ for all $\mathbf{e} \in \mathbb{R}^3$ (as is easy to see by, say, picking an orthonormal basis). The multiplicative property of the determinant shows that if α has $\det(\alpha) = 1$, then $(\alpha(\mathbf{v}_1) \times \alpha(\mathbf{v}_2)) \cdot \alpha(\mathbf{v}_3) = (\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3$. Thus we have

$$\begin{aligned} (\alpha(\mathbf{v}) \times \alpha(\mathbf{w})) \cdot \mathbf{e} &= (\mathbf{v} \times \mathbf{w}) \cdot \alpha^{-1}(\mathbf{e}) \\ &= \alpha(\mathbf{v} \times \mathbf{w}) \cdot \mathbf{e}, \end{aligned}$$

where the second equality follows because $\alpha \in O(\mathbb{R}^3)$. Thus (6) follows. \square

Remark 2.7. Notice that (4) implies the quantity

$$|(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3|$$

is the volume of the parallelepiped spanned by the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$, that is the volume of the solid $\{\mathbf{x} \in \mathbb{R}^3 : \mathbf{x} = \sum_{i=1}^3 \lambda_i \mathbf{v}_i, 0 \leq \lambda_i \leq 1\}$.

We may view \mathbb{R}^2 as the plane inside \mathbb{R}^3 spanned by $\{\mathbf{e}_1, \mathbf{e}_2\}$. Given a basis $\mathbf{v}_1, \mathbf{v}_2$ of \mathbb{R}^2 , we say it is positively oriented if $\mathbf{v}_1 \times \mathbf{v}_2$ is a positive multiple of \mathbf{e}_3 , and negatively oriented if it is a negative multiple of \mathbf{e}_3 (that it is a multiple of \mathbf{e}_3 follows from part 6 of Lemma 2.6. It is easy to check (using part 7 of Lemma 2.6 for example) that this definition is equivalent to requiring that the angle going from \mathbf{v}_1 to \mathbf{v}_2 is acute.

Notice that condition for $\mathbf{v}_1, \mathbf{v}_2$ to be positively oriented can be rewritten as $(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{e}_3 > 0$. Given an ordered basis $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$ of \mathbb{R}^3 , we say it is positively oriented if the scalar $(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3$ is positive and negatively oriented if $(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3 < 0$. (Note that the property 5 of Lemma 2.6 shows that the product $(\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3$ cannot be equal to zero if the vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ are linearly independent). It is easy to check (again using property 6 of Lemma 2.6) that this definition agrees with the “right-hand rule” that you learn in physics.

The equality

$$\det(A) = (\mathbf{v}_1 \times \mathbf{v}_2) \cdot \mathbf{v}_3$$

gives us a general definition for \mathbb{R}^n : given an ordered basis $(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$ we say it is positively or negatively oriented according to the sign of $\det(A)$ where A is the matrix with columns $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$.

Definition 2.8. A linear map $\alpha \in GL(\mathbb{R}^n)$ is said to be *orientation preserving* if α send any triple of positively (negatively) oriented vectors to a triple of positively (negatively) oriented vectors. By the multiplicativity of the determinant, this is equivalent to requiring that $\det(\alpha) > 0$.

Remark 2.9. It is a special feature of \mathbb{R}^3 that one can define the vector product \times – notice that our general definition of an orientation for an ordered basis did *not* use such a product on \mathbb{R}^n . Note also that \times makes \mathbb{R}^3 into a noncommutative, nonunital, nonassociative algebra!

The group of orientation preserving isometries of \mathbb{R}^n is denoted $\text{SO}(\mathbb{R}^n)$. Equivalently, we may define $\text{SO}(\mathbb{R}^n)$ to be

$$\text{SO}(\mathbb{R}^n) = \{\alpha \in \text{O}(\mathbb{R}^n) : \det(\alpha) = 1\}.$$

Notice that if $\alpha \in \text{O}(\mathbb{R}^n)$, then if A is the matrix of α with respect to the standard basis, we have $AA^t = I$, and hence $\det(A)\det(A^t) = 1$. Since $\det(A) = \det(A^t)$, it follows that $\det(A) = \det(\alpha) = \pm 1$.

We now want to understand this group for small n . Consider first the case $n = 2$. Let $\alpha \in \text{SO}(\mathbb{R}^2)$. Then if A is the matrix of α with respect to the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ the columns of A have length 1 and are perpendicular. It is then easy to see that this forces A to have the form

$$A = \begin{pmatrix} \cos(\theta) & \mp \sin(\theta) \\ \sin(\theta) & \pm \cos(\theta) \end{pmatrix}$$

where $\theta \in \mathbb{R}$. Checking the condition that A preserves the orientations we see that A has the form

$$A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

which is the matrix of a rotation by θ about the origin. (*What are the linear isometries which do not preserve orientations?*). Thus in \mathbb{R}^2 , the orientation preserving isometries fixing $\mathbf{0}$ are just rotations about $\mathbf{0}$.

Now consider the case $n = 3$. A *rotation* in \mathbb{R}^3 is a map which fixes a line through the origin, and rotates \mathbb{R}^3 about that line by some angle θ . Clearly a rotation is an example of an orientation preserving isometry, and in fact in \mathbb{R}^3 this is the only example, as we will shortly see. We first make the following simple observation about the eigenvalues of a linear isometry.

Lemma 2.10. *If $\alpha \in \text{O}(\mathbb{R}^n)$ then every real eigenvalue $\lambda \in \mathbb{R}$ of α has is equal to 1 or -1 . Moreover distinct eigenspaces must be orthogonal.*

Proof. Let A be the matrix of α with respect to the standard basis. Then (thinking of \mathbf{v} as a column vector) we have

$$\begin{aligned} \lambda(\mathbf{v} \cdot \mathbf{v}) &= (\lambda\mathbf{v}) \cdot \mathbf{v} = (A\mathbf{v}) \cdot \mathbf{v} \\ &= (A\mathbf{v})^t \mathbf{v} = \mathbf{v}^t A^t \mathbf{v} \\ &= \mathbf{v}^t A^{-1} \mathbf{v} = \mathbf{v}^t (\lambda^{-1} \mathbf{v}) \\ &= \lambda^{-1}(\mathbf{v} \cdot \mathbf{v}). \end{aligned}$$

Since $\mathbf{v} \cdot \mathbf{v} > 0$ when $\mathbf{v} \neq \mathbf{0}$ it follows that $\lambda = \lambda^{-1}$ as required. The moreover part of the lemma may be proved in the same fashion. \square

Remark 2.11. This lemma can be generalized to show that the eigenvalues of an orthogonal matrix all have modulus 1.

Lemma 2.12. *The group $\text{SO}(\mathbb{R}^3)$ consists precisely of the set of rotations.*

Proof. Let $\alpha \in \text{SO}(\mathbb{R}^3)$ be an orientation preserving isometry. We must find a line in \mathbb{R}^3 which is fixed by α . The characteristic polynomial of α is $P(t) = \det(\alpha - tI)$. Since $P(0) = \det(\alpha) = 1$ and $P(t) \rightarrow -\infty$ as $t \rightarrow \infty$, the intermediate value theorem shows that P has a root in $(0, \infty)$, which by the previous lemma must be 1. Therefore α has a fixed line, spanned by a length 1 eigenvector \mathbf{e} say, (thus $\alpha(\mathbf{e}) = \mathbf{e}$).

We now claim that α is a rotation about that line. Let H be the plane perpendicular to \mathbf{e} ,

$$H = \{\mathbf{x} : \mathbf{x} \cdot \mathbf{e} = 0\}.$$

Since α is an isometry, and so preserves the dot product, α restricts to a map from H to itself, which is again an isometry. By the discussion preceding this lemma, it follows that if we take a basis $\mathbf{w}_1, \mathbf{w}_2$ of H which consists of perpendicular unit vectors, then with respect to the basis $\{\mathbf{e}, \mathbf{w}_1, \mathbf{w}_2\}$ the matrix of α is given by

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) \\ 0 & \sin(\theta) & \cos(\theta) \end{pmatrix}$$

and α is a rotation as claimed. □

Remark 2.13. It follows from the previous Lemma that the composition of two rotations is again a rotation – think about how to see this directly. One can also describe explicitly the elements of $\text{SO}(\mathbb{R}^n)$ in a similar fashion, using the orthogonality of the eigenspaces. (This is however easiest to do using the unitary group which we will define later.)

In this course we will be interested in groups analogous to $\text{SO}(\mathbb{R}^3)$ and study their structure, primarily as algebraic objects. However, first we investigate the shape or *topology* of $\text{SO}(\mathbb{R}^3)$. This makes sense to talk about, because $\text{SO}(\mathbb{R}^n)$ is a subset of $\text{Hom}(\mathbb{R}^n, \mathbb{R}^n)$ the vector space of linear maps from \mathbb{R}^n to itself, which is a normed vector space under the operator norm. Thus $\text{SO}(\mathbb{R}^n)$ is a metric space, and indeed the group operation, coming from composition of linear maps is continuous (even smooth).

Lemma 2.14. *The group $\text{SO}(\mathbb{R}^n)$ (as a subspace of $\text{Hom}(\mathbb{R}^n, \mathbb{R}^n)$) is compact.*

Proof. Because $\alpha \in \text{SO}(\mathbb{R}^n)$ is an isometry, we have $\|\alpha(\mathbf{v})\| = \|\mathbf{v}\|$ for all $\mathbf{v} \in \mathbb{R}^n$. But then it follows from the definition of the operator norm that $\|\alpha\| = 1$. Thus $\text{SO}(\mathbb{R}^n) \subset \{\alpha \in \text{GL}(\mathbb{R}^n) : \|\alpha\| = 1\}$. Moreover, the conditions defining $\text{SO}(\mathbb{R}^n)$ are clearly closed, so that since $L(\mathbb{R}^n, \mathbb{R}^n)$ is a finite dimensional vector space, it follows $\text{SO}(\mathbb{R}^n)$ is compact. □

We now give a heuristic argument for what $\text{SO}(\mathbb{R}^3)$ “looks like”. We may represent any rotation by a vector in the set $\{\mathbf{v} \in \mathbb{R}^3 : \|\mathbf{v}\| \leq \pi\}$, by letting the direction of the vector give us the line to rotate about, and the length of the vector the angle by which we rotate (rotating about the vector \mathbf{v} using the righthand rule say). Now the points in the open ball $\{\mathbf{v} \in \mathbb{R}^3 : \|\mathbf{v}\| < \pi\}$ all corresponds to distinct rotations, but antipodal points on the boundary sphere $\{\mathbf{v} \in \mathbb{R}^3 : \|\mathbf{v}\| = \pi\}$ give the same rotation (the elements of order 2 in $\text{SO}(\mathbb{R}^3)$), so that we can visualize the group $\text{SO}(\mathbb{R}^3)$ as the solid ball of radius π where we identify antipodal points on the boundary. The problem with making this rigorous is that you have to show that the map from $\text{SO}(\mathbb{R}^3)$ to this quotient of the ball is continuous, which is somewhat tedious. We shall return to the question of the shape of $\text{SO}(\mathbb{R}^3)$ after we have discussed division algebras.

Remark 2.15. There is a beautiful theory of groups which are also smooth compact spaces (where the group operations of multiplication and taking inverses are required to be smooth). The group $\text{SO}(\mathbb{R}^3)$ is the simplest nonabelian example.

We next want to show a result which is closer to the central spirit of this course, namely that the group $\text{SO}(\mathbb{R}^3)$ is simple (*i.e.* the only normal subgroups are the trivial group and $\text{SO}(\mathbb{R}^3)$ itself). To do this we need some preparatory lemma. Let S^2 denote the sphere of radius 1 around the origin $\mathbf{0} \in \mathbb{R}^3$. Given two pairs of vectors $(\mathbf{v}_1, \mathbf{v}_2)$ and $(\mathbf{w}_1, \mathbf{w}_2)$ in \mathbb{R}^3 we say they are equidistant pairs if $\|\mathbf{v}_1 - \mathbf{v}_2\| = \|\mathbf{w}_1 - \mathbf{w}_2\|$.

Lemma 2.16. *The action of $\text{SO}(\mathbb{R}^3)$ on S^2 is transitive, and moreover is transitive on equidistant pairs (whose points lie in S^2).*

Proof. To see that $\text{SO}(\mathbb{R}^3)$ acts transitively on S^2 , consider points $\mathbf{v}, \mathbf{w} \in S^2$. We may assume that \mathbf{v} and \mathbf{w} are distinct, in which case we define a plane in \mathbb{R}^3 . Let \mathbf{u} be a vector in S^2 perpendicular to this plane. Then it is clear that there is a rotation about the line through \mathbf{u} which sends \mathbf{v} to \mathbf{w} as required.

For the second part, let $(\mathbf{v}_1, \mathbf{v}_2)$ and $(\mathbf{w}_1, \mathbf{w}_2)$ in \mathbb{R}^3 be an equidistant pair. First pick $\sigma \in \text{SO}(\mathbb{R}^3)$ such that $\sigma(\mathbf{v}_1) = \mathbf{w}_1$. Then we have

$$\|\mathbf{w}_1 - \sigma(\mathbf{v}_2)\| = \|\sigma(\mathbf{v}_1) - \sigma(\mathbf{v}_2)\| = \|\mathbf{w}_1 - \mathbf{w}_2\|$$

, so that $(\mathbf{w}_1, \sigma(\mathbf{v}_2))$ and $(\mathbf{w}_1, \mathbf{w}_2)$ are an equidistant pair. Thus we need to find $\tau \in \text{Stab}_{\text{SO}(\mathbb{R}^3)}(\mathbf{w}_1)$ such that $\tau(\sigma(\mathbf{v}_2)) = \mathbf{w}_2$. The stabilizer is exactly the subgroup of rotations about the axis given by \mathbf{w}_1 . Then the equidistant criterion is exactly what is required. \square

Theorem 2.17. *The group $\text{SO}(\mathbb{R}^3)$ is simple.*

Proof. We show this using the previous lemma. The strategy is to show that a normal subgroup must contain certain kinds of elements, and then show that these elements in fact generate $\text{SO}(\mathbb{R}^3)$. Suppose that $1 \neq K \triangleleft \text{SO}(\mathbb{R}^3)$ is a normal subgroup of $\text{SO}(\mathbb{R}^3)$. Then pick $\sigma \neq 1$ in K , so that there is a vector $\mathbf{u} \in S^2$ such that σ is the rotation about \mathbf{u} by the angle θ , ($0 \leq \theta \leq \pi$). Then if we pick $\mathbf{w} \in S^2$ in the plane perpendicular to \mathbf{u} , we have $\|\mathbf{w} - \sigma(\mathbf{w})\| = 2 \sin(\theta)$. Now suppose that $k \in \mathbb{N}$ is such that $0 < \pi/k < \theta$. Then setting $b = \sin(\pi/k)/\sin(\theta)$ it follows that $\mathbf{z} = b\mathbf{w} + \sqrt{1 - b^2}\mathbf{u}$ has

$$\|\mathbf{z} - \sigma(\mathbf{z})\| = b\|\mathbf{w} - \sigma(\mathbf{w})\| = 2 \sin(\pi/k).$$

Now let $\mathbf{w}_1 \in S^2$ be the vector perpendicular to \mathbf{u} between \mathbf{w} and $\sigma(\mathbf{w})$ such that $\|\mathbf{w}_1 - \mathbf{w}\| = 2 \sin(\pi/k)$. Then using the moreover part of the previous lemma, we can find a $\tau \in \text{SO}(\mathbb{R}^3)$ such that $\tau(\mathbf{z}) = \mathbf{w}$ and $\tau(\sigma(\mathbf{z})) = \mathbf{w}_1$. Now since K is normal, we have $\sigma_1 = \tau\sigma\tau^{-1} \in K$, and $\sigma_1(\mathbf{w}) = \mathbf{w}_1$.

Repeating this argument with \mathbf{w}_1 instead of \mathbf{w} , we find $\sigma_2 \in K$ and $\mathbf{w}_2 = \sigma_2(\mathbf{w}_1)$ such that $\mathbf{w}_2 \in S^2$ is perpendicular to \mathbf{u} with $\|\mathbf{w}_1 - \mathbf{w}_2\| = 2 \sin(\pi/k)$, and continuing in this way, we get $\sigma, \sigma_1, \dots, \sigma_k \in K$ and $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k \in \mathbb{R}^3$ all perpendicular to \mathbf{u} with $\sigma_i(\mathbf{w}_{i-1}) = \mathbf{w}_i$ and $\|\mathbf{w}_i - \mathbf{w}_{i-1}\| = 2 \sin(\pi/k)$. But then $\rho = \sigma_k\sigma_{k-1}\dots\sigma_1\sigma$ has $\rho(\mathbf{w}) = -\mathbf{w}$. But then $\rho \in K$ is a rotation by π , that is, a rotation of order 2 (*why?*), and so since K is normal, it contains all rotations of order 2. But these generate $\text{SO}(\mathbb{R}^3)$, so that $K = \text{SO}(\mathbb{R}^3)$ as required. \square

Exercise 2.18. Complete the proof by showing that any rotation in $\text{SO}(\mathbb{R}^3)$ can be written as the product of two elements of order 2.

3. THE QUATERNIONS

3.1. Division algebras. We now wish to discuss division algebras – that is, fields where we drop the condition that the multiplication is commutative – and in particular the quaternions, the first division algebra to be discovered. We start with some general definitions.

Definition 3.1. Let k be a field. A k -algebra is a k -vector space A equipped with a bilinear map $m: A \times A \rightarrow A$ known as the multiplication, and a nonzero element $1 \in A$ such that $m(1, x) = m(x, 1) = x$ for all $x \in A$. (Of course, we will normally write $m(x, y)$ as xy).

A *division algebra* is a finite dimensional (as a k -vector space) algebra in which there are no zero divisors, that is, if $ab = 0$ then one of a or b is equal to 0. If A is associative, this is the same as requiring that each nonzero element x has a multiplicative inverse, *i.e.* an element x^{-1} such that $xx^{-1} = x^{-1}x = 1$, as the following lemma shows.

Lemma 3.2. *A finite dimensional associative algebra A has no zero divisors if and only if every nonzero element is invertible.*

Proof. First suppose that every nonzero element is invertible. Then if we have $a, b \in A$ such that $ab = 0$, if $a \neq 0$, we have

$$0 = a^{-1}0 = a^{-1}(ab) = (a^{-1}a)b = 1.b = b.$$

On the other hand, if A has no zero divisors, give any $a \in A$, we can define a linear map $L_a: A \rightarrow A$ by $L_a(x) = ax$. Since A has no zero divisors, L_a has zero kernel, and hence since A is finite dimensional, L_a is an isomorphism. Thus there is a $b \in A$ with $L_a(b) = ab = 1$. Similarly by considering the linear map of right multiplication by a , we find a $c \in A$ with $ca = 1$. But now

$$b = 1.b = (ca).b = c.(ab) = c.1 = c,$$

and so $b = c$ is a two-sided inverse of a as required. \square

3.2. The quaternions. Consider the algebra of 2×2 matrices, $M_2(\mathbb{C})$. This is an \mathbb{R} -algebra (since it is even a \mathbb{C} -algebra), but of course not a division algebra. However, if we define

$$\mathbb{H} = \left\{ \begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix} : z, w \in \mathbb{C} \right\},$$

then clearly \mathbb{H} is closed under addition, and moreover a simple calculation shows that it is also closed under multiplication. Since for $x \in \mathbb{H}$ we have $\det(x) = |z|^2 + |w|^2$ where $z, w \in \mathbb{C}$ are as above, it is immediate that \mathbb{H} is a division algebra, known as the quaternions. Clearly it has dimension 4 as an \mathbb{R} -vector space. Note that \mathbb{H} is not commutative.

Although $M_2(\mathbb{R})$ is an algebra over \mathbb{C} , the quaternions are only an \mathbb{R} -subspace of $M_2(\mathbb{R})$, not a \mathbb{C} -subspace, so they are not an algebra over \mathbb{C} (at least not in the obvious fashion). In fact there are no finite-dimensional associative division algebras over \mathbb{C} other than \mathbb{C} itself (see the problem set).

By the multiplicativity of the determinant, if we set the norm of an element $x \in \mathbb{H}$ to be $\|x\| = \sqrt{\det(x)} = \sqrt{|z|^2 + |w|^2}$ we see that $\|xy\| = \|x\|\|y\|$. Moreover if we identify \mathbb{H} with \mathbb{R}^4 using the real and imaginary parts of z and w , this norm is just the standard Euclidean norm. In particular, the norm comes from an inner

product on \mathbb{H} , which is also compatible with the multiplication, in the sense that $(ax, ay) = \|a\|^2(x, y)$ – one can see this by observing that in \mathbb{R}^n we have

$$(x, y) = \frac{1}{2}(\|x + y\|^2 - \|x\|^2 - \|y\|^2).$$

Set

$$i = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}; \quad j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}; \quad k = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

It is easy to check that $\{1, i, j, k\}$ is an orthonormal basis for \mathbb{H} , so that the multiplication in the quaternions is completely determined by calculating the products of the basis elements. These are

$$\begin{aligned} i^2 &= j^2 = k^2 = -1; \\ ij &= -ji = k; \\ jk &= -kj = i; \\ ki &= -ik = j. \end{aligned}$$

These are the formulas originally used by Hamilton to define \mathbb{H} .

\mathbb{H} has a conjugation, similar to complex conjugation: for $x \in \mathbb{H}$ of the form $a + bi + cj + dk$ we set $\bar{x} = a - bi - cj - dk$. Alternatively, on the level of matrices, conjugation is the map

$$\begin{pmatrix} z & w \\ -\bar{w} & \bar{z} \end{pmatrix} \mapsto \begin{pmatrix} \bar{z} & -w \\ \bar{w} & z \end{pmatrix}.$$

Since by the standard formula for the inverse of a matrix, the second form can be written as $x \mapsto \|x\|^2 x^{-1}$, it is clear that $x \mapsto \bar{x}$ is an anti-involution of \mathbb{H} , that is, $\bar{\bar{x}} = x$ and $\overline{\bar{x}y} = y\bar{x}$.

We now wish to use the quaternions to study the groups $\text{SO}(\mathbb{R}^3)$ and $\text{SO}(\mathbb{R}^4)$. The unit quaternions

$$\mathbb{U} = \{x \in \mathbb{H} : \|x\| = 1\}$$

form a group under multiplication, making the 3-sphere into a group. (This is the quaternion analogue of the unit circle in the complex plane, which is also a group under multiplication). Similarly, the non-zero quaternions \mathbb{H}^\times form a group under multiplication.

Notice that \mathbb{H}^\times acts on \mathbb{H} by conjugation:

$$x \mapsto Ad_a(x) = axa^{-1} = \|a\|^{-2}ax\bar{a}.$$

Since $\overline{Ad_a(x)} = Ad_a(\bar{x})$ it follows that the imaginary quaternions

$$\mathbb{I} = \{x \in \mathbb{H} : \bar{x} = -x\}$$

(or, alternatively, the quaternions which are orthogonal to 1), are preserved by \mathbb{H}^\times . Since \mathbb{I} is a copy of \mathbb{R}^3 , the restriction of the action of \mathbb{H}^\times to \mathbb{I} gives us a group homomorphism

$$\rho: \mathbb{H}^\times \rightarrow \text{GL}(\mathbb{R}^3).$$

But now since $\|Ad_a(x)\| = \|axa^{-1}\| = \|a\|\|x\|\|a^{-1}\| = \|x\|$, the image of this map lies in $\text{O}(\mathbb{R}^3)$. We can also calculate the kernel of ρ : suppose that $x \in \mathbb{H}^\times$ satisfies $\rho(x) = Id$. Then $xh = hx$ for all $h \in \mathbb{I}$, but since x clearly commutes with 1, it follows that x lies in the center of \mathbb{H} . Writing out x in the basis $\{1, i, j, k\}$ and letting $h = i, j, k$ in turn, it is easy to see that the center of the quaternions consists of exactly the scalars \mathbb{R} .

If we restrict to \mathbb{U} , then the map $\phi = \rho|_{\mathbb{U}}: \mathbb{U} \rightarrow \mathbf{O}(\mathbb{R}^3)$ is precisely 2-to-1: the only scalars in \mathbb{U} are $\{\pm 1\}$. In fact the image is exactly $\mathbf{SO}(\mathbb{R}^3)$, indeed the next lemma shows this directly by calculating the rotation that you obtain from a given unit quaternion, and so $\mathbf{SO}(\mathbb{R}^3) \cong \mathbb{U}/\{\pm 1\}$.

Lemma 3.3. *Let $u \in \mathbb{U}$. Then we may write $u = \cos(\theta) + \sin(\theta)v$ where v is a unit quaternion in \mathbb{I} . The transformation $\rho(u)$ is a rotation by 2θ about v .*

Proof. First suppose that $u = \cos(\theta) + i \sin(\theta)$, and let $q = xi + yj + zk$ be an element of \mathbb{I} . Then we have

$$\begin{aligned} uq\bar{u} &= xi + u(y + zi)j\bar{u} \\ &= xi + u^2(y + zi)j, \end{aligned}$$

since $ji = -ij$. Hence $\rho(u)$ preserves the i -axis, and rotates the plane perpendicular to i by 2θ . If u is now an arbitrary unit quaternion, by conjugation by elements in the planes spanned by $\{1, i\}, \{1, j\}, \{1, k\}$ we may move it to a quaternion of the above form (for example, first rotate using i so that the coefficient of k vanishes, then rotate with k so that the coefficient of j vanishes), and so the general statement follows. \square

It follows from this that the image of \mathbb{U} under ρ is the set of all rotations in $\mathbf{O}(\mathbb{R}^3)$, which we have already seen is $\mathbf{SO}(\mathbb{R}^3)$. (Since the image is automatically a group, this gives another proof that the set of rotations forms a group, and since you can check that rotations generate $\mathbf{SO}(\mathbb{R}^3)$ directly, that $\mathbf{SO}(\mathbb{R}^3)$ is exactly this group).

Remark 3.4. Here is a (sketch of a) topological proof that the image of ρ is $\mathbf{SO}(\mathbb{R}^3)$. Since \mathbb{U} is connected, the image must be contained in $\mathbf{SO}(\mathbb{R}^3)$, thus we just need to check surjectivity. Now the map ρ is clearly a smooth map from \mathbb{U} to $\mathbf{SO}(\mathbb{R}^3)$, and you can check that the derivative at the identity is invertible. Since, as we saw above, the group acts transitively on itself by conjugation, this shows that the derivative is everywhere invertible, and so by the inverse function theorem, the image of ρ is open, and hence since \mathbb{U} is compact, the image of ρ is both open and closed, and thus is the connected component of I in $\mathbf{O}(\mathbb{R}^3)$.

This gives us another way of viewing $\mathbf{SO}(\mathbb{R}^3)$ as a topological space: it is the space obtained from the 3-sphere by identifying antipodal points – this is called real projective 3-space, and we will return to it when we study projective geometry.

Finally, I want to briefly show how you can use the quaternions to show that $\mathbf{PSO}(\mathbb{R}^4) = \mathbf{SO}(\mathbb{R}^4)/\{\pm 1\}$ is isomorphic to $\mathbf{SO}(\mathbb{R}^3) \times \mathbf{SO}(\mathbb{R}^3)$ (and hence is not a simple group). Consider the action of $\mathbb{U} \times \mathbb{U}$ on \mathbb{H} via

$$(u_1, u_2)(h) = u_1 h u_2^{-1}$$

It is easy to see that this gives a map $\psi: \mathbb{U} \times \mathbb{U} \rightarrow \mathbf{SO}(\mathbb{R}^4)$. To see what the kernel of this map is, suppose that $u_1 h u_2^{-1} = h$ for all $h \in \mathbb{H}$. Letting $h = 1$ we see that $u_1 = u_2$, and moreover therefore u_1 is central in \mathbb{H} . Thus from our calculation of the center of \mathbb{H} above we see that (u_1, u_2) lies in the kernel of ψ if and only if $(u_1, u_2) = \pm(1, 1)$. Again one needs an argument to show that this homomorphism is surjective (see the problem set), but once this is known the result follows easily (the preimage of $\{\pm I\} \subset \mathbf{SO}(\mathbb{R}^4)$ in $\mathbb{U} \times \mathbb{U}$ is $\{(\pm 1, \pm 1)\} \subset \mathbb{U} \times \mathbb{U}$).

4. COMPOSITION ALGEBRAS

For the field of real numbers \mathbb{R} , division algebras are rare: they must have dimensions 1, 2, 4 or 8 as \mathbb{R} -vector spaces. To prove this fact is somewhat beyond the means of this course, however we will instead be able to prove a somewhat weaker result, first proved by Hurwitz, which shows that the only \mathbb{R} -algebras which possess a compatible inner product are \mathbb{R} , \mathbb{C} , \mathbb{H} and an 8-dimensional nonassociative algebra known as the octonions.

Definition 4.1. Let V be a vector space over a field k . A *quadratic form* on V is a function $N: V \rightarrow k$ such that

- (1) $N(\lambda v) = \lambda^2 N(v)$;
- (2) The function

$$(v, w) \mapsto v.w = N(v + w) - N(v) - N(w),$$

is a (clearly symmetric) bilinear form on V .

Notice that

$$v.v = N(2v) - N(v) - N(v) = 2N(v),$$

hence if $\text{char}(k) \neq 2$, then the quadratic form is determined by the associated bilinear form, and vice versa. In fact, in this case it is more standard to set

$$(v, w) = \frac{1}{2}(N(v + w) - N(v) - N(w)),$$

and we will speak of (\cdot, \cdot) as the bilinear form attached to the quadratic form N . \mathbb{R}^n is of course an example of a vector space with a quadratic form $N(v) = \|v\|^2$, induced by the standard inner product. Note that the definition of N makes sense for any field k , whereas for $\|\cdot\|$ we must take a square root, and so some arithmetic of the field k is necessary.

We can now define the class of algebras that we want to study.

Definition 4.2. We say that a division algebra A over k is a *composition algebra*, if there is a nonzero quadratic form on A such that

$$N(xy) = N(x)N(y).$$

Thus \mathbb{C} with $|\cdot|^2$ and \mathbb{H} with $\|\cdot\|^2$ are examples of composition algebras over \mathbb{R} . The goal of this section is to show that in fact there are only 4 such composition algebras. We begin with some lemmas about composition algebras.

Lemma 4.3. Let A be a composition algebra over a field k . Then we have

$$N(x) = (x, x) \neq 0, \quad \forall x \in A,$$

hence the bilinear form (\cdot, \cdot) attached to the quadratic form is nondegenerate, in the sense that its radical

$$J = \{x \in A : (x, y) = 0, \forall y \in A\}$$

is exactly $\{0\}$.

Proof. First note that $N(1) = N(1)N(1)$, so that either $N(1) = 0$ or $N(1) = 1$. In the former case, N is identically zero, so that we must have $N(1) = 1$. Thus if $x \in A - \{0\}$ then x has a right inverse, that is there is a $y \in A$ such that $xy = 1$.

$$1 = N(1) = N(xy) = N(x)N(y),$$

and hence $N(x) = (x, x) \neq 0$. It follows easily that $J = \{0\}$ as required. \square

We now derive some properties of the norm in the case where the composition algebra is an \mathbb{R} -algebra.

Lemma 4.4. *Let A be a composition algebra over \mathbb{R} . Then the norm N is positive definite, that is, for all $x \in A$ we have $N(x) \geq 0$, with equality if and only if $x = 0$.*

Proof. Let $x \in A$ be nonzero. Then since $N(1) = 1$ and $N(x) \neq 0$ by the proof of Lemma 4.3, we need only show that $N(x) > 0$. Suppose that $N(x) < 0$. Then consider $y = \lambda 1 + x$, so that

$$N(y) = \lambda^2 + 2(\lambda, x) + N(x).$$

Since the right-hand side is a quadratic of λ which is positive for large λ , and negative for $\lambda = 0$, it must vanish somewhere, say for $\lambda = c$. But then $N(c + x) = 0$ and we must have $x = -c$, and so $N(x) = N(-c) = c^2 \geq 0$, a contradiction. \square

A quadratic form over \mathbb{R} which is positive definite in the sense of the previous lemma is in fact (the square of) a Euclidean norm, as the next lemma shows.

Lemma 4.5. (*Gramm-Schmidt*) *Let V be a real vector space with a positive definite quadratic form N , and let $\{v_1, v_2, \dots, v_n\}$ be a basis of V . Then there exists a basis $\{w_1, w_2, \dots, w_n\}$ of A such that*

$$\text{span}(w_1, w_2, \dots, w_i) = \text{span}\{v_1, v_2, \dots, v_i\}, \quad \forall i, 1 \leq i \leq n,$$

and $(w_i, v_j) = \delta_{ij}$. The basis $\{w_1, w_2, \dots, w_n\}$ is said to be orthonormal.

Proof. We use induction on $\dim(V)$. If $\dim(V) = 0$ then there is nothing to prove. Suppose the result is known for spaces of dimension $n - 1$. Then since the first $n - 1$ vectors $\{v_1, v_2, \dots, v_{n-1}\}$ spans an $(n - 1)$ -dimensional space, by induction there exist vectors w_1, w_2, \dots, w_{n-1} satisfying $(w_i, w_j) = \delta_{ij}$ for all $i, j, 1 \leq i, j \leq n - 1$ and such that $\text{span}(w_1, w_2, \dots, w_i) = \text{span}\{v_1, v_2, \dots, v_i\}, \forall i, 1 \leq i \leq n - 1$. Now since by Lemma 4.3 we know that $(v_n, v_n) \neq 0$ we may let

$$w'_n = (v_n, v_n)^{-1} v_n - \sum_{j=1}^{n-1} \frac{(v_n, w_j)}{(v_n, v_n)} w_j.$$

Since v_n is not in the span of $\{v_1, v_2, \dots, v_{n-1}\}$, and this is the same as the span of $\{w_1, w_2, \dots, w_{n-1}\}$ by induction, we see that $w'_n \neq 0$. Moreover it is easy to check that w'_n satisfies $(w'_n, w_j) = 0$ for all $j, 1 \leq j \leq n - 1$. Since the span of $\{w_1, w_2, \dots, v_n\}$ is clearly the same as the span of $\{w_1, w_2, \dots, w'_n\}$ we can complete the inductive step by setting $w_n = N(w'_n)^{-1/2} w'_n$ (the square root is a real number by Lemma 4.4). \square

Remark 4.6. We record here a consequence of the above lemma which we will use in the proof of Proposition 4.11. Suppose that V is real vector space with a quadratic form N . For any subspace W of V , we define W^\perp to be the subspace of vectors $\{v \in V : (v, w) = 0, \forall w \in W\}$. If N is positive definite, and W is a proper subspace, then $W^\perp \neq \{0\}$. Indeed if we pick a basis of V which contains a basis of W , Lemma 4.5 shows that we can obtain an orthonormal basis of V which contains an orthonormal basis of W . Any of the vectors in this basis which are not in W clearly lie in W^\perp and so $W^\perp \neq \{0\}$ as required. (In fact, it is clear from this argument that $\dim(W^\perp) = \dim(V) - \dim(W)$).

We will use this nondegeneracy all the time in our investigation of composition algebras in the following manner: in order to show that two elements $a, b \in A$ are equal, it is enough to show that $(a, x) = (b, x)$ for all $x \in A$.

Lemma 4.7. (1) (*Scaling law*) For $x, y, z \in A$ we have

$$(xz, yz) = N(z)(x, y) = (zx, zy).$$

(2) (*Exchange law*) For $x, y, z, w \in A$ we have

$$(xy, wz) = 2(x, w)(y, z) - (xz, wy)$$

Proof. For (1), consider the expansion of $N((x + y)z)$ in two ways. First we have

$$N((x + y)z) = ((x + y)z, (x + y)z) = (xz, xz) + 2(yz, xz) + (yz, yz).$$

On the other hand, using the fact that the norm is multiplicative

$$N(x + y)N(z) = (N(x) + 2(x, y) + N(y))N(z)$$

and so comparing the two expressions we find that $(yz, xz) = (x, y)N(z)$. The other side of the equality in the statement follows similarly.

For (2) replace z by $(y + z)$ in the scaling law so as to obtain

$$(x(y + z), w(y + z)) = N(y + z)(x, w) = (N(y) + 2(y, z) + N(z))(x, w)$$

While expanding directly we get

$$(x(y + z), w(y + z)) = (xy, wy) + (xy, wz) + (xz, wy) + (xz, wz).$$

Hence comparing the two expressions, and using the scaling law we get the result. \square

We now introduce a “conjugation” operation on A : the idea is to mimic the conjugation of complex numbers. Thinking of complex conjugation as reflection in the real axis, we define for $x \in A$

$$\bar{x} = 2(1, x) - x$$

We can show that this operation behaves much as complex conjugation does.

Lemma 4.8. We have

- (1) $(xy, z) = (y, \bar{x}z)$ and $(xy, z) = (x, z\bar{y})$;
- (2) $\overline{\bar{y}} = y$;
- (3) $\bar{\bar{x}} = x$.

Proof. For the first part we note that by the exchange law

$$\begin{aligned} (xy, z) &= (xy, 1.z) = 2(x, 1)(y, z) - (xz, y) \\ &= (y, 2(x, 1)z) - (y, xz) \\ &= (y, \bar{x}z). \end{aligned}$$

The other case is proved similarly. Now consider for $x, z \in A$

$$(x, z) = (x.1, z) = (1, \bar{x}z) = (\bar{x}z, 1) = (z, \bar{\bar{x}}) = (\bar{\bar{x}}, z)$$

Since the inner product is nondegenerate, this implies that $\bar{\bar{x}} = x$. Finally, to see that $\overline{\bar{y}} = y$ note that

$$(z, \overline{\bar{y}}) = (xyz, 1) = (yz, \bar{x}) = (z, \bar{y}\bar{x})$$

Thus again by nondegeneracy, we must have $\overline{\bar{y}} = y$ as required. \square

Remark 4.9. Not that $\bar{x}x = x\bar{x} = N(x)1$, because $\overline{x\bar{x}} = \bar{\bar{x}\bar{x}} = x\bar{x}$, so that $x\bar{x} \in \mathfrak{k}1$. Moreover

$$N(x) = (x, x) = (x\bar{x}, 1),$$

so that $x\bar{x} = N(x)1$. The same argument shows that $\bar{x}x = N(x)1$.

Now we want to show that given these properties, in fact A is one of just four possible real composition algebras. To do this we consider a “doubling” process known as the Cayley-Dickson double.

Definition 4.10. Suppose that A is a composition algebra. Define a new algebra D to be $A \oplus A$, where the new multiplication is given by

$$[a, b][c, d] = [ac - d\bar{b}, cb + \bar{a}d].$$

If N is the quadratic form attached to A , we equip D with a quadratic form N_D by setting $N_D([a, b]) = N(a) + N(b)$, so that the two copies of A are orthogonal to each other. Note that the conjugation in D is given by

$$\overline{[a, b]} = [\bar{a}, -b]$$

Thus starting with any composition algebra, we may double it to obtain an algebra with a quadratic form of twice the dimension (it is *not* necessarily a composition algebra – we will see shortly precisely when it is).

Remarkably, we can use this doubling construction to show that $\mathbb{R}, \mathbb{C}, \mathbb{H}$ and \mathbb{O} are the only real composition algebras. The key is the following observation.

Proposition 4.11. *Suppose A is a real composition algebra, and B is a proper subalgebra of A . Then A contains the double of B .*

Proof. Since B is a proper subalgebra of A , Remark 4.6 shows that the subspace of A

$$B^\perp = \{x \in A : (x, b) = 0, \forall b \in B\}$$

is nonzero. Choose $e \in B^\perp$ with $N(e) = 1$ (to find such an e , we first pick some nonzero $f \in B^\perp$, since $N(f) > 0$ by Lemma 4.4 we may define $e = N(f)^{-1/2}f$). We claim that $B + iB$ is isomorphic to the double of B . First note that for $a, b, c, d \in B$ we have

$$\begin{aligned} (a + ib, c + id) &= (a, c) + (a, id) + (ib, c) + (ib, id) \\ &= (a, c) + (a\bar{d}, i) + (i, c\bar{b}) + N(i)(b, d) \\ &= (a, c) + (b, d), \end{aligned}$$

since B is closed under multiplication and conjugation, and $N(i) = 1$. Thus $B + iB$ is an orthogonal direct sum. Moreover

$$\overline{a + ib} = \bar{a} + \bar{i}\bar{b} = \bar{a} - ib,$$

since $\bar{i}\bar{b} = 2(1, ib) - ib = 2(\bar{b}, i) - ib = -ib$. Note that this implies in particular that for all $b \in B$ we have $ib = \bar{b}i$, since $\bar{i}\bar{b} = \bar{b}i = -\bar{b}i$. Finally we have

$$(2) \quad (a + ib)(c + id) = (ac + ib.c + a.id + ib.id)$$

we examine the last three terms of this expression separately, using the nondegeneracy of the inner product: let $z \in A$ be arbitrary. Then

$$(ib.c, z) = (ib, z\bar{c}) = (\bar{b}i, z\bar{c}) = 0 - (\bar{b}\bar{c}.zi) = (\bar{b}\bar{c}.i, z) = (i.cb, z).$$

Similarly we find that

$$(a.id, z) = (id, \bar{a}z) = 0 - (iz, \bar{a}d) = (z, i\bar{a}d).$$

Also

$$\begin{aligned} (ib.id, z) &= (ib, z(-id)) = -2(i, z)(b, id) + (i.id, zb) \\ &= (i.id, zb) = (id, -i.zb) \\ &= N(i)(d, -zb) = (-\bar{d}b, z) \end{aligned}$$

Combining these three equations, we find that Equation (2) becomes

$$(a + ib)(c + id) = (ac - \bar{d}b) + i(\bar{a}d + cb),$$

hence $B + iB$ is indeed isomorphic to the double of B . □

It is now quite easy for us to conclude that the only normed division algebras over \mathbb{R} are $\mathbb{R}, \mathbb{C}, \mathbb{H}$ and \mathbb{O} . It is easy to check that the double of \mathbb{R} is \mathbb{C} , and that the double of \mathbb{C} is \mathbb{H} . Let \mathbb{O} denote the double of \mathbb{H} . Now suppose that A is a normed division algebra. Since it contains a copy of \mathbb{R} , it is either equal to \mathbb{R} by the previous proposition it contains the double of \mathbb{R} . Then either A is equal to \mathbb{C} or it contains the double of \mathbb{C} . Since A is finite dimensional, it follows that continuing in this way we find A is obtained from \mathbb{R} by successive doubling.

But why does the list of division algebras stop, as opposed to there being one for each power of 2? The answer is that the doubling procedure produces a new algebra equipped with a quadratic form, but it is not necessarily a composition algebra. Our next proposition shows that the double of a composition algebra A is a division algebra if and only if the algebra A is associative.

Proposition 4.12. *If A is the double of a composition algebra B , then A is a composition algebra if and only if B is associative.*

Proof. Let B be a subalgebra such that $A = B + i_B A$. Then we must have

$$N(a + i_B b)N(c + i_B d) = N(ac - \bar{d}b) + i_B(cb + \bar{a}d).$$

Expanding we find that this is equivalent to requiring that

$$N(a)N(c) + N(a)N(d) + N(b)N(c) + N(b)N(d) = (ac - \bar{d}b, ac - \bar{d}b) + (cb + \bar{a}d, cb + \bar{a}d)$$

and using the multiplicativity of the norm again and the fact that $N(\bar{a}) = N(a)$, we find that the right hand side is

$$N(a)N(c) - 2(ac, \bar{d}b) + N(d)N(b) + N(c)N(b) + 2(cb, \bar{a}d) + N(a)N(d)$$

Hence the norm on A is compatible with the multiplication if and only if

$$(ac, \bar{d}b) = (cb, \bar{a}d)$$

for all $a, b, c, d \in B$. But this holds if and only if

$$((ac).b, d) = (a.(cb), d), \quad \forall a, b, c, d \in B.$$

By nondegeneracy of the bilinear form, this last equation is equivalent to the associativity of B , as required. Finally, to see that A is a division algebra, note that since B is a division algebra, Lemma 4.4 shows that the norm on B is positive definite. It follows directly from the definition of the double that the norm on A is

positive definite. We use this to show that A has no zero divisors. Indeed suppose that $xy = 0$. Then since the multiplication is compatible with the norm, we have

$$0 = N(0) = N(xy) = N(x)N(y),$$

and so $N(x) = 0$ or $N(y) = 0$, which is equivalent to requiring that x or y is zero. \square

In precisely the same fashion we can show the following:

Lemma 4.13. *Let D be the double of composition algebra A , and assume that D is a composition algebra. Then D is associative if and only if A is commutative and associative.*

Proof. Write $D = A + i_A A$. Clearly the composition algebra A must be associative. Moreover, since the formula for multiplication in D shows that $i_A.(ab) = (i_A b).a$, we see that if D is associative, then A must be commutative. To see that this condition is sufficient, we must check the equation

$$(a + i_A b)(c + i_A d).(e + i_A f) = (a + i_A b).(c + i_A d)(e + i_A f),$$

holds. Expanding the left side we get:

$$\begin{aligned} ((ac - d\bar{b}) + i_A(cb + \bar{a}d))(e + i_A f) &= ((ac - d\bar{b})e - f(\bar{b}\bar{c} + \bar{d}a)) \\ &\quad + i_A(e(cb + \bar{a}d) + (\bar{c}\bar{a} - b\bar{d})f) \end{aligned}$$

while expanding the right hand side we get

$$\begin{aligned} (a + i_A b)((ce - f\bar{d}) + i_A(ed - \bar{c}f)) &= (a(ce - f\bar{d}) - (ed - \bar{c}f)\bar{b}) \\ &\quad + i_A((ce - f\bar{d})b + \bar{a}(ed - \bar{c}f)) \end{aligned}$$

Comparing the two expressions we find that we must have

$$ac.e - d\bar{b}.e - f.\bar{b}\bar{c} - f.\bar{d}a = a.ce - a.fd - ed.\bar{b} + \bar{c}f.\bar{b}$$

and

$$e.cb + e.\bar{a}d + \bar{c}\bar{a}.f - b\bar{d}.f = ce.b - f\bar{d}.b + \bar{a}.ed - \bar{a}.\bar{c}f$$

But it follows immediately from the fact A is commutative that these equations hold, and so we are done. \square

Finally, note the following:

Lemma 4.14. *If A is the double of a composition algebra B , then A is commutative and associative if and only if B is commutative, associative, and has trivial conjugation.*

Proof. Clearly, if $A = B + i_B B$, then B must be commutative and associative. Now since $i_B b = \bar{b}i_B$, it follows that the conjugation on B must be trivial if A is commutative. For the converse note that for A to be commutative we must have

$$(a + i_B b)(c + i_B d) = (c + i_B d)(a + i_B b),$$

that is,

$$(ac - d\bar{b}) + i_B(cb + \bar{a}d) = (ca - b\bar{d}) + i_B(ad + \bar{c}b)$$

but this clearly holds if B is commutative with trivial conjugation. \square

Applying the lemmas to the case of a real composition algebra, we find that \mathbb{C} is commutative, associative, but has a nontrivial conjugation. Thus its double, \mathbb{H} , is associative but not commutative. Therefore doubling again, we get a composition algebra, known as the octonions \mathbb{O} which is not associative. Since the double of \mathbb{O} will not be a composition algebra, there can be no other composition algebras over \mathbb{R} .

Remark 4.15. If we take the double of \mathbb{O} , we obtain an algebra \mathbb{S} known as the sedenions. It is *not* a composition algebra, since \mathbb{O} is not associative, thus it may have zero divisors. On the other hand, every element has an inverse, since you can check that if $x \in \mathbb{S}$ then $x\bar{x} = N(x) \in \mathbb{R}$ so that $\bar{x}/N(x)$ is a two-sided inverse of x (since $\overline{\bar{x}} = x$). In fact it turns out that \mathbb{S} does possess zero divisors.

5. THE CLASSICAL GROUPS

5.1. Compact classical groups. .

In the first section, we defined the groups $O(\mathbb{R}^n)$ by considering the linear transformations which preserved distance. We now want to define analogous groups for the other real division algebras we have constructed. This is straight-forward provided the algebras are associative, so for the moment we exclude the octonions. Let \mathbb{D} be one of \mathbb{R}, \mathbb{C} or \mathbb{H} .

The first thing to observe is that one can do linear algebra over \mathbb{D} just as we do for a field, provided we are careful about left and right. (Indeed if you know about rings and modules, a left \mathbb{D} -vector space is just a left \mathbb{D} -module, and similarly a right \mathbb{D} -vector space is a right \mathbb{D} -module. The important point is that all the results about the existence of bases *etc.* continue to hold in the case where \mathbb{D} is noncommutative. From now on when we say vector space, we tacitly assume it to be a left vector space.

As an example, we show that any finitely generated vector space V over \mathbb{D} has a basis. Recall that a basis is a generating set which is linearly independent (recall that this means that whenever we have $\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k = 0$ then $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$). Suppose that $\{v_1, v_2, \dots, v_k\}$ is a generating set. Then if they are linearly independent we are done. Otherwise there is a linear dependence,

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k = 0,$$

where not all the λ_i are zero. But then suppose j is such that $\lambda_j \neq 0$. Then we have $v_j = \sum_{i \neq j} (\lambda_i^{-1} \lambda_j) v_i$, and it is clear that the vectors $\{v_1, v_2, \dots, v_{j-1}, v_{j+1}, \dots, v_k\}$ already span V . Continuing in this way we must find a basis as required.

A linear transformation of a \mathbb{D} -vector space is just a map of \mathbb{D} -modules (that is, it is compatible with the addition of vectors, and multiplication by the scalars in D). An isomorphism of \mathbb{D} -vector spaces is a bijective linear transformation. Just as for fields, it is automatic that the inverse is a linear map. Thus if V is a \mathbb{D} -module we obtain a group $GL(V)$ of invertible \mathbb{D} -linear maps from V to itself. Let \mathbb{D}^n be the vector space of n -tuples of elements of \mathbb{D} thought of as a left D -vector space. Just as for a field, we may attach a matrix to a linear map. The subtlety here is that, in order for the matrix to correspond to a linear map of (left) vector spaces, we need to multiply on the right, that is, the linear map associated to a matrix A is given by $x \mapsto xA$ (thus we are thinking of \mathbb{D}^n as the space of row vectors). Using this correspondence, we may identify the group $GL(\mathbb{D}^n)$ with the (opposite of the) group $GL_n(\mathbb{D})$ of invertible $n \times n$ matrices with entries in D , where the group operation is given by matrix multiplication. Note that it is here that there is a problem with the octonions – matrix multiplication over the octonions is not associative.

For the real vector spaces \mathbb{R}^n we also had a notion of distance. Since D has a norm, we can also obtain a distance on D^n in an obvious way: let $N: D^n \rightarrow \mathbb{R}$ be given by

$$N(x) = \sum_{i=1}^n N(x_i) = \sum_{i=1}^n \bar{x}_i x_i,$$

where $x = (x_1, x_2, \dots, x_n) \in D^n$ (and then the distance is given by $N(x)^{1/2}$). Unlike in the case $D = \mathbb{R}$, this distance is not associated to a quadratic form over D , but rather over a form twisted by the conjugation on D : for $z = (z_1, z_2, \dots, z_n) \in$

D^n , let $\bar{z} = (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_n)$, and let, for $x, y \in D^n$

$$(x, y) = \sum_{i=1}^n x_i \bar{y}_i$$

then $N(x) = (x, x) \geq 0$, but now whereas (\cdot, \cdot) is compatible with the addition in both variables, its behavior with respect to scalar multiplication is not symmetric in the variables: instead we have

$$\lambda(x, y) = (\lambda x, y); \quad (x, \bar{\lambda}y) = (x, y)\lambda, \quad x, y \in \mathbb{D}^n.$$

The form (\cdot, \cdot) is called a *skew* or Hermitian form on \mathbb{D}^n . We can then set

$$I(\mathbb{D}^n) = \{\phi \in GL(\mathbb{D}^n) : (\phi(x), \phi(y)) = (x, y), \forall x, y \in \mathbb{D}^n\}.$$

(in fact, one can show that this is equivalent to requiring that $N(\phi(x)) = N(x)$ for all $x \in \mathbb{D}^n$.) In the case $\mathbb{D} = \mathbb{R}$, we recover the orthogonal groups $O(\mathbb{R}^n)$. In the case $\mathbb{D} = \mathbb{C}$ we obtain the *unitary groups* $U(\mathbb{C}^n)$, while in the case $\mathbb{D} = \mathbb{H}$ we obtain the (compact) *symplectic groups* $Sp(\mathbb{H}^n)$. Collectively these groups are known as the *classical groups*. In the first two cases, the determinant gives us a natural normal subgroup, $SO(\mathbb{R}^n)$, and $SU(\mathbb{C}^n)$, the special orthogonal and special unitary groups respectively (in the case of $Sp(\mathbb{H}^n)$, since \mathbb{H} is noncommutative, we do not have a well-defined determinant).

Just as for the orthogonal groups, we may also give a more explicit description of these group in terms of their image in $GL_n(\mathbb{D})$. For this, we use an involution on matrices. Let $M_{n,m}(\mathbb{D})$ denote the space of $n \times m$ matrices with entries in \mathbb{D} . Then $t: M_{n,m}(\mathbb{D}) \rightarrow M_{m,n}(\mathbb{D})$ is the linear map given by $(a_{ij}) \mapsto (a_{ji})$. Similarly we have a map $\bar{\cdot}: M_{n,m}(\mathbb{D}) \rightarrow M_{n,m}(\mathbb{D})$ given by $(a_{ij}) \mapsto (\bar{a}_{ij})$. Then $\bar{\cdot}$ and t both square to give the identity, and commute with each other. Their composition is therefore also of order 2, and we denote it $A \mapsto A^*$. We have the following lemma:

Lemma 5.1. *For $A \in M_{n,m}(\mathbb{H})$ and $B \in M_{m,p}(\mathbb{D})$ we have*

$$(AB)^* = B^* A^* \in M_{n,p}(\mathbb{D}).$$

Proof. Let $A = (a_{ij})$ and $B = (b_{jk})$. The ik -th entries of (AB) is then $\sum_{j=1}^m a_{ij} b_{jk}$, and hence the ik -th entry of $(AB)^*$ is

$$\sum_{j=1}^m \overline{a_{kj} b_{ji}} = \sum_{j=1}^m \bar{b}_{ji} \bar{a}_{kj} = \sum_{j=1}^m (B^*)_{ij} (A^*)_{jk}.$$

which is just the ik -th entry of $B^* A^*$ as required. \square

It follows immediately from this lemma that the map $A \mapsto A^*$ restricts to a map on $GL_n(\mathbb{D})$. Its connection with the group $I(\mathbb{D}^n)$ is given by the following lemma.

Lemma 5.2. *Let $A \in GL_n(\mathbb{D})$. Then we have*

$$(x, yA) = (xA^*, y)$$

and

$$(xA, y) = (x, yA^*).$$

Moreover the image of $I(\mathbb{D}^n)$ in $GL_n(D)$ is exactly the set

$$\{A \in GL_n(D) : A.A^* = I\}$$

Proof. Since it is clear that $(A^*)^* = A$, it is enough to show the first of the equalities. It is easy to check that $(\overline{xA})^t = A^* \bar{x}^t$. Hence we have

$$\begin{aligned} (x, yA) &= x.(\overline{yA})^t \\ &= xA^* \bar{y}^t \\ &= (xA^*) \bar{y} \\ &= (xA, y). \end{aligned}$$

To see the moreover part, note that,

$$(x, y) = (xA, yA) = (xA.A^*, y), \quad \forall x, y \in \mathbb{D}^n.$$

For $1 \leq i \leq n$, let $e_i = (0, \dots, 1, 0, \dots, 0)$ (where the 1 is in the i -th place). Then $\{e_i\}_{1 \leq i \leq n}$ is a \mathbb{D} -basis of \mathbb{D}^n and $(e_i, e_j) = \delta_{ij}$. Hence the entries of the matrix $A.A^*$ are exactly the numbers $(e_i.A.A^*, e_j) = \delta_{ij}$ so it follows that A preserves the form if and only if $A.A^* = I$. □

Remark 5.3. If you know what a manifold is (if not, then very briefly it is a smooth space which locally looks like \mathbb{R}^n for some n) each of the classical groups is an example of a compact manifold which is a group (and the group operations are smooth). It turns out that if you abstractly define a smooth compact group (known as a compact Lie group) then the families $\mathrm{SO}_n(\mathbb{R})$, $\mathrm{SU}_n(\mathbb{C})$ and $\mathrm{Sp}_n(\mathbb{H})$ and their product are almost the only possibilities – there are exactly five “exceptional” simple compact Lie groups which are not of this type. These can all be related in some way to the octonions, though we cannot give all the details here.

Since the matrices over the octonions are nonassociative, we cannot use them to construct groups. Nevertheless there other ways of constructing groups from an algebra. For example, we have seen that the automorphisms of the Hamiltonians are isomorphic to the group $\mathrm{SO}(\mathbb{R}^3)$, and one can similarly ask about the group of automorphisms of the octonions. This turns out to be the smallest example of an exceptional compact Lie group (known as G_2).

We now consider the octonions more explicitly: Suppose that we wish to find a multiplication table for them similiar to Hamilton’s original rules for multiplying $\{1, i, j, k\}$. One approach is to recall how the octonions are constructed by doubling. We start by picking e_1 orthogonal to 1. This yields a copy of the double of \mathbb{R} (that is, \mathbb{C}) in \mathbb{O} . We then pick $e_2 \in \mathbb{O}$ which is perpendicular to each of $\{1, e_1\}$. Then $\{1, e_1, e_2\}$ generate a copy of the quaternions, with \mathbb{R} -basis $\{1, e_1, e_2, e_1e_2\}$. Finally if we pick e_3 perpendicular to each of $\{1, e_1, e_2, e_1e_2\}$ we find that \mathbb{O} has an \mathbb{R} -basis given by $\{1, e_1, e_2, e_1e_2, e_3, e_1e_3, e_2e_3, e_1e_2e_3\}$ and the multiplication of any pair of elements of this basis is easy to determine from the rule for multiplication in a Cayley-Dickson double. We call a triple of unit vectors (e_1, e_2, e_3) of this form a *basic triple*. Note also that this multiplication table is independent of the choices made for e_1, e_2, e_3 , just as for the quaternions, the doubling construction shows that any two unit vectors perpendicular to 1 can be used as “ i ” and “ j ”.

We now use this to study the group of automorphisms of \mathbb{O} . Pick a particular choice of basic triple (e_1, e_2, e_3) . It is clear from our discussion that if (e_1, e_2, e_3) is any other basic triple, then the map $c_i \mapsto e_i$ extends to a unique automorphism of \mathbb{O} , and conversely, any automorphism sends (c_1, c_2, c_3) to a basic triple. It follows that, once the choice of basic triple (c_1, c_2, c_3) is made, the automorphisms of \mathbb{O}

are in bijection with the space of basic triples in \mathbb{O} , indeed the bijection sends $\alpha \in \text{Aut}(\mathbb{O})$ to $(\alpha(c_1), \alpha(c_2), \alpha(c_3))$. Indeed the map is even a homeomorphism.

But what does the space of basic triples look like? Well, e_1 can be chosen to be any point on the unit sphere in the imaginary octonions, which is a 6-sphere. Similarly, e_2 can be any point in the unit sphere of the 6-dimensional \mathbb{R} -vector space perpendicular to 1 and e_1 , and thus e_2 is any point in a 5-sphere. Finally e_3 is a unit vector in the 4-dimensional vector space perpendicular to 1, e_1, e_2 and e_1e_2 , hence e_3 lies on a 3-sphere. This picture of the set of basic triples suggests that they are a space of dimension $6 + 5 + 3 = 14$, and thus we expect that $\text{Aut}(\mathbb{O})$ is 14-dimensional. (Any reasonable rigorous definition of dimension will declare a space like the space of basic triples to be 14-dimensional, so what is missing here is just a precise definition).

In the problem set you are asked to calculate a dimension for the classical groups. It is straightforward to check from this that none of the classical group have dimension 14, and so the automorphism group of \mathbb{O} must indeed be a different group (once you are prepared to believe that it is simple).

5.2. Classical groups over arbitrary fields. We now wish to construct similar families of groups over an arbitrary field k . The strategy of considering division algebras over k is doomed to failure, because, as is shown in the problem sets, there are no nontrivial division algebras over an algebraically closed field. Indeed, the case where k is a finite field, which will concern us for some time, also does not permit any interesting associative division algebras either, as a beautiful theorem of Wedderburn shows. We need some preliminary definitions. Given any ring R , and $x \in R$ we write

$$C_R(x) = \{y \in R : xy = yx\}$$

for the *centralizer* of x . It is a subring of R . The *center* $Z(R)$ of R is the intersection of all the subrings $C_R(x)$ as x runs over the elements of R , thus

$$Z(R) = \{x \in R : xy = yx, \forall y \in R\}.$$

A *skew field* is a ring in which every nonzero element is invertible (that is, a field where the multiplication is not necessarily commutative). Note that given a skew field D , its center $Z(D)$ is a field, and D is an associative division algebra over its center. Thus as discussed in the last section, much of the theory of fields extends to skew-fields: in particular, we have a notion of vector space over a skew-field, and such spaces have a well defined dimension. If $k \subset K$ then we can define $[K : k] = \dim_k(K)$, the dimension of K as a k -vector space. Thus if $k_1 \subset k_2 \subset k_3$ then we may consider k_2 and k_3 as vector spaces over k_1 , and k_3 as a vector space over k_2 . It is easy to check that, just as for fields we have

$$(3) \quad [k_3 : k_1] = [k_2 : k_1][k_3 : k_2].$$

The final ingredients we need before we can prove Wedderburn's theorem are the cyclotomic polynomials. The polynomial $t^n - 1 \in \mathbb{C}[t]$ can be factored as

$$t^n - 1 = \prod_{\varepsilon} (t - \varepsilon)$$

where ε runs over the n -th roots of unity $\{e^{2\pi ik/n}\}_{0 \leq k < n}$. We say ε is a *primitive* n -th root of unity if $\varepsilon^n = 1$ and $\varepsilon^k \neq 1$ for all $k < n$, and define the n -th cyclotomic

polynomial to be

$$\Phi_n(t) = \prod_{\varepsilon \text{ primitive}} (t - \varepsilon).$$

Thus $\Phi_n(t) \in \mathbb{C}[t]$ and has leading coefficient equal to 1. We show $\Phi_n(t) \in \mathbb{Z}[t]$ using induction. For $n = 1$, clearly $\Phi_1(t) = t - 1$. But now since

$$t^n - 1 = \Phi_n(t)f(t)$$

where $f(t)$ is a product of $\Phi_d(t)$ over the positive integers d dividing n , we see that by induction $f(t) \in \mathbb{Z}[t]$ and $f(t)$ has leading coefficient 1. But the expression $\Phi_n(t) = (t^n - 1)/f(t)$ immediately implies that $\Phi_n(t) \in \mathbb{Z}[t]$ also. Finally, observe that $\Phi_n(t)$ divides $(t^n - 1)/(t^d - 1)$ for any proper divisor d of n , by the same argument.

Theorem 5.4. (Wedderburn): *A finite skew field is a field.*

Proof. Let D be a division ring with finite many elements. For each $x \in D$ we have $C_D(x)$ is a subring of D , and $Z(D)$ is a field. Since D is finite, $Z(D)$ is finite, say with q elements. Since each $C_D(x)$ is a vector space over $Z(D)$, we may write $|C_D(x)| = q^{d_x}$ for some integer $d_x \in \mathbb{N}$. Let $n = d_1$. Using Equation 3 we see that d_x divides n for all $x \in D$.

Now consider the group of nonzero elements of D under multiplication. It is a finite group, and is the disjoint union of its conjugacy classes. Since the conjugacy class of an element has size one if and only if that element lies in the center of D , we see that

$$|D^\times| = q^n - 1 = (q - 1) + \sum_{d|n} (q^n - 1)/(q^d - 1),$$

where d runs over the dimensions of the centralizers of elements in distinct conjugacy classes of size greater than 1. We claim that such an equation can only hold if $n = 1$, and hence $D = Z(D)$ as required. To see this, we use the remark on cyclotomic polynomials above. Since $\Phi_n(t)$ divides $(t^n - 1)/(t^d - 1)$ for all d a proper divisor of n , setting $t = q$ implies that $\Phi_n(q)$ divides $q - 1$, since it divides all the other terms in the above equality. But now since $q \in \mathbb{N}$ is at least 2, and any $\varepsilon \in \mathbb{C}$ a primitive n -th root of unity lies on the unit circle in the complex plane, clearly $|q - \varepsilon| > q - 1$, unless $n = 1$. But then $|\Phi_n(q)| > q - 1$ contradicting the fact that $\Phi_n(q)$ divides $q - 1$, and so we must have $n = 1$ as required. \square

To find analogues of our classical groups which will make sense over an arbitrary field, we start by expressing the classical groups as subgroups of $\text{GL}_n(\mathbb{C})$. We may define two involutions on $\text{GL}_n(\mathbb{C})$. Let $\theta: \text{GL}_n(\mathbb{C}) \rightarrow \text{GL}_n(\mathbb{C})$ be given by setting $\theta(A)$ for $A = (a_{ij})$ to be (\bar{a}_{ij}) , the matrix whose entries are the complex conjugates of the entries of A , and let $\kappa: \text{GL}_n(\mathbb{C}) \rightarrow \text{GL}_n(\mathbb{C})$ be given by $A \mapsto \kappa(A) = (A^t)^{-1} = (A^{-1})^t$. It follows immediately that the orthogonal groups may be written as

$$\text{O}(\mathbb{R}^n) = \{A \in \text{GL}_n(\mathbb{C}) : \kappa(A) = A, \theta(A) = A\}.$$

We also have

$$\text{U}(\mathbb{C}) = \{A \in \text{GL}_n(\mathbb{C}) : \kappa(A) = \theta(A)\}$$

Finally, and most interestingly, we have the symplectic groups. If $\{e_1, e_2, \dots, e_n\}$ denotes the standard basis of \mathbb{H}^n , and we view \mathbb{C} as a subring of \mathbb{H} via $z \mapsto \text{Re}(z) +$

$\text{Im}(z)i$, then H^n becomes a complex vector space with basis

$$\{e_1, e_2, \dots, e_n, je_1, je_2, \dots, je_n\}.$$

Let $\phi: \mathbb{C}^{2n} \rightarrow \mathbb{H}^n$ be the unique \mathbb{C} -linear map sending the standard basis of \mathbb{C}^{2n} to this basis. Then ϕ induces an algebra map $\phi^*: \text{End}(\mathbb{H}^n) \rightarrow \text{End}(\mathbb{C}^{2n})$ from the \mathbb{H} -linear maps of \mathbb{H}^n to the \mathbb{C} -linear maps of \mathbb{C}^{2n} given by $\alpha \mapsto \phi^{-1} \circ \alpha \circ \phi$ (here End denotes the space of linear maps from a vector space to itself). Clearly ϕ^* is an embedding. We want to compute the image of $\text{Sp}(\mathbb{H}^n)$ under this map. To do this we first calculate the way our identification of \mathbb{C}^{2n} and \mathbb{H}^n interacts with the structure of these spaces. Let $J \in \text{GL}(\mathbb{C}^{2n})$ be the matrix given by

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

and let $-: \mathbb{C}^n \rightarrow \mathbb{C}^n$ be the map $(z_1, z_2, \dots, z_n) \mapsto (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_n)$.

Lemma 5.5. *Under the identification $\phi: \mathbb{C}^n \rightarrow \mathbb{H}^n$ given by*

$$(z_1, z_2, \dots, z_n, w_1, w_2, \dots, w_n) \mapsto (z_1 + w_1j, z_2 + w_2j, \dots, z_n + w_nj)$$

multiplication by j becomes the map

$$(z_1, z_2, \dots, z_n, w_1, w_2, \dots, w_n) \mapsto (-\bar{w}_1, -\bar{w}_2, \dots, -\bar{w}_n, \bar{z}_1, \bar{z}_2, \dots, \bar{z}_n).$$

or more compactly, we may write this as $v \mapsto \overline{Jv}$. Moreover, if $\langle \cdot, \cdot \rangle_{\mathbb{C}}$ denotes the complex Hermitian form on \mathbb{C}^{2n} and $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ denotes the quaternionic Hermitian form on \mathbb{H}^n , we have

$$\langle v, w \rangle_{\mathbb{C}} = C(\langle \phi(v), \phi(w) \rangle_{\mathbb{H}}),$$

where $C: \mathbb{H} \rightarrow \mathbb{C}$ is the \mathbb{R} -linear map given $x + yi + zj + wk \mapsto x + yi$.

Proof. The first part follows from the fact that if $w \in \mathbb{C} \subset \mathbb{H}$ then $kw = \bar{w}j$. For the second part, let $v = (v_1, v_2)$ and $w = (w_1, w_2)$ where $v_1, v_2, w_1, w_2 \in \mathbb{C}^n$. Then observe that

$$\langle v_1 + v_2j, w_1 + w_2j \rangle_{\mathbb{H}} = \langle v_1, w_1 \rangle_{\mathbb{H}} + j \langle v_2, w_1 \rangle_{\mathbb{H}} + \langle v_1, w_2 \rangle_{\mathbb{H}}j + \langle v_2, w_2 \rangle_{\mathbb{H}},$$

and hence

$$\begin{aligned} C(\langle v_1 + v_2j, w_1 + w_2j \rangle_{\mathbb{H}}) &= \langle v_1, w_1 \rangle_{\mathbb{H}} + \langle v_2, w_2 \rangle_{\mathbb{H}} \\ &= \langle v_1, w_1 \rangle_{\mathbb{C}} + \langle v_2, w_2 \rangle_{\mathbb{C}} \\ &= \langle v, w \rangle_{\mathbb{C}} \end{aligned}$$

as required. \square

Now the image of $\text{GL}(\mathbb{H}^n)$ in $\text{GL}_{2n}(\mathbb{C})$ under the map induced by ϕ consists of the invertible complex matrices A which are compatible with multiplication by elements of \mathbb{H} . But since $\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j$, this is equivalent to requiring that A commutes with multiplication by $j \in \mathbb{H}$. The previous lemma shows that this is just the condition that $JAv = A(\overline{Jv})$ for all $v \in \mathbb{C}^{2n}$, that is, $AJ = J\theta(A)$.

Next we wish to calculate the image of $\text{Sp}(\mathbb{H}^n)$ in $\text{GL}_{2n}(\mathbb{C})$. For this we use the second part of the above lemma. Indeed if α preserves $\langle \cdot, \cdot \rangle_{\mathbb{H}}$, then clearly it preserves $\langle \cdot, \cdot \rangle_{\mathbb{C}}$, so that $\phi^*(\text{Sp}(\mathbb{H}^n)) \subset \text{U}(\mathbb{C}^{2n})$. However, even more is true: the fact that α is \mathbb{H} -linear implies that α preserves $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ if and only if $\phi^*(\alpha)$ preserves $\langle \cdot, \cdot \rangle_{\mathbb{C}}$. To see this, we just have to note that α preserves $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ if and only if it preserves the norm N (see the problem set), but $N(x) = C(N(x))$, so if $\phi^*(\alpha)$ preserves $\langle \cdot, \cdot \rangle_{\mathbb{C}} = C(\langle \cdot, \cdot \rangle_{\mathbb{H}})$ then α preserves N .

Thus we have shown that the image of $\mathrm{Sp}(\mathbb{H}^n)$ under ϕ^* (identifying $\mathrm{GL}(\mathbb{C}^{2n})$ with $\mathrm{GL}_{2n}(\mathbb{C})$), is

$$\{A \in \mathrm{GL}_{2n}(\mathbb{C}) : \kappa(A) = \theta(A), AJ = J\theta(A)\}$$

Now we may rewrite this as

$$\mathrm{Sp}(\mathbb{H}^n) = \{A \in \mathrm{GL}_{2n}(\mathbb{C}) : AJA^t = J, \kappa(A) = \theta(A)\}$$

In order to obtain groups which are defined any field (not using any properties of the field \mathbb{C} like complex conjugation) we simply loosen our conditions, dropping the conditions involving θ . Thus we obtain three families of groups:

- The general linear group: $\mathrm{GL}_n(\mathbb{C}) = \{A \in M_n(\mathbb{C}) : A \text{ is invertible}\}$.
- The orthogonal groups: $\mathrm{O}_n(\mathbb{C}) = \{A \in \mathrm{GL}_n(\mathbb{C}) : A.A^t = I_n\}$.
- The symplectic groups: $\mathrm{Sp}_{2n}(\mathbb{C}) = \{A \in \mathrm{GL}_{2n}(\mathbb{C}) : AJA^t = J\}$.

These groups make sense over any field k , and we call the groups that arise the *classical groups* over k , denoted $\mathrm{GL}_n(k)$, $\mathrm{O}_n(k)$ and $\mathrm{Sp}_{2n}(k)$. Note that in the case of the first two families the map given by the determinant $A \mapsto \det(A)$ has as kernel a subgroup denoted $SL_n(k)$ and $SO_n(k)$ respectively (the *special linear* and *special orthogonal* groups respectively).

Remark 5.6. Notice that the compact groups are all simply the intersection of the above groups with the unitary group. One can use this observation to show that over \mathbb{C} we do not really lose anything by removing the condition involving θ : each compact classical group we started with is a maximal compact subgroup of the classical groups over \mathbb{C} . In the case of $\mathrm{GL}_n(\mathbb{C})$, for example, any compact subgroup of $\mathrm{GL}_n(\mathbb{C})$ is conjugate to a subgroup of $U_n(\mathbb{C})$. The proof of this statement is essentially a glorification of the proof of Lemma 4.5.

Let us consider the three families of groups we have constructed more closely. The general linear groups are clearly just the full group of automorphisms of a vector space of dimension n over the field. The orthogonal groups are, similarly to the case of \mathbb{R}^n , the group of linear transformations which preserve the bilinear form on k^n given by

$$(x, y) = \sum_{j=1}^n x_j y_j,$$

where $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$. The last family however is somewhat new: it corresponds to the linear automorphisms of a $2n$ -dimensional vector space k^{2n} which are compatible with a bilinear form given by

$$(x, y) = x^t J y$$

where J is the $(2n) \times (2n)$ matrix given above. Since $J^t = -J$ we see that this form is *skew-symmetric* in the sense that

$$(x, y) = -(y, x).$$

Thus these families of groups lead us to consider vector spaces over a field k which come equipped with the additional structure of a bilinear form that is either symmetric or skew-symmetric.

6. PROJECTIVE GEOMETRY

We now switch gears somewhat and study the geometry of lines in a vector space. This will give us an object that the first of our classical groups, the general linear group $GL_n(k)$, acts upon, thus it will allow us to better understand the structure of this group.

Definition 6.1. Let V be a vector space of dimension $n + 1$. Then $\mathbb{P}(V)$ the projective space of V is the set of lines in V (i.e. one-dimensional subspaces of V). We say that such a projective space has dimension n .

Example 6.2. Over the real numbers, the n -dimensional projective space is denoted $\mathbb{R}P^n$. Since any line in \mathbb{R}^{n+1} intersects the unit sphere $S^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$ in exactly two points (which are antipodal to each other), we may picture $\mathbb{R}P^n$ as the space obtained from S^n by identifying opposite points. In the case $n = 1$, it is clear identifying opposite points on a circle yields another circle, while in the case $n = 3$, our discussion of $SO(\mathbb{R}^3)$ showed that, as a space, $SO(\mathbb{R}^3) = \mathbb{R}P^3$. To picture $\mathbb{R}P^2$, one can also imagine the upper hemisphere with antipodal points on the equator identified. This perspective reveals that a large piece of $\mathbb{R}P^2$ looks like \mathbb{R}^2 (since an open ball is homeomorphic to \mathbb{R}^n). We will see below that this is a general phenomenon.

A one dimensional projective space is called a projective line, while a two-dimensional projective space is called a projective plane. Given a vector in $V \setminus \{0\}$, it spans a line in V , and hence there is a natural map from $V \setminus \{0\} \rightarrow \mathbb{P}(V)$, which we denote $v \mapsto [v]$. Clearly $[v] = [w]$ if and only if $v = \lambda w$ for some $\lambda \in k \setminus \{0\}$. It is easy to check that $v \sim w$ if $v = \lambda w$, for some $\lambda \in k \setminus \{0\}$ gives an equivalence relation on $V \setminus \{0\}$, and thus $\mathbb{P}(V) = V \setminus \{0\} / \sim$. Given a point p in $\mathbb{P}(V)$ (i.e. a line in V) we say that v is a *representative vector* for p if $[v] = p$.

We use this idea to introduce coordinates in projective space. Let $\{v_0, v_1, \dots, v_n\}$ be a basis of V . Suppose that $p \in \mathbb{P}(V)$ is a point in projective space and take a representative vector v . Then v can be written uniquely as

$$v = \sum_{j=0}^n x_j v_j.$$

By convention, if the basis $\{v_0, v_1, \dots, v_n\}$ is understood, we write $p = [x_0 : x_1 : \dots : x_n]$. Since $v \neq 0$ there is some x_j which is nonzero (and moreover, the set $\{k : x_k \neq 0\}$ is independent of the choice of representative). Let $\mathbb{A}_l = \{[x_0 : x_1 : \dots : x_n] \in \mathbb{P}(V) : x_l \neq 0\}$, a subset of $\mathbb{P}(V)$, the l -th *affine chart* for $\mathbb{P}(V)$. Then we may identify \mathbb{A}_l with k^n via

$$[x_0 : x_1 : \dots : x_n] \mapsto (x_0/x_k, \dots, x_{k-1}/x_k, x_{k+1}/x_k, \dots, x_n/x_k).$$

Notice that $\mathbb{P}(V) = \bigcup_{l=0}^n \mathbb{A}_l$. Moreover, if $p \notin \mathbb{A}_0$, then p is a line in the n -dimensional vector space W spanned by $\{v_1, v_2, \dots, v_n\}$, and hence we see that

$$\mathbb{P}(V) = \mathbb{A}_0 \cup \mathbb{P}(W),$$

that is, $\mathbb{P}(V)$ is the union of k^n and a projective space of one smaller dimension.

Exercise 6.3. Many points lie in more than one \mathbb{A}_k , and hence have different coordinates depending on which affine chart we look at them in. What is the relation between the different coordinates?

Example 6.4. We use the quaternions to identify $\mathbb{C}P^1$ with the two-sphere S^2 . We may think of \mathbb{H} as a copy of $\mathbb{C}^2 = \mathbb{C} \oplus \mathbb{C}j$. Then the nonzero vectors are just the nonzero elements of \mathbb{H} , which we denote by \mathbb{H}^\times , moreover, the equivalence classes of the relation \sim above are now exactly the left cosets of $\mathbb{C}^\times \subset \mathbb{H}^\times$, and so $\mathbb{C}P^1$ is just $\mathbb{C}^\times \backslash \mathbb{H}^\times$.

But why does this help us see that $\mathbb{C}P^1$ is a 2-sphere? Let \mathbb{I} be the purely imaginary quaternions, so that since $\mathbb{I} \cong \mathbb{R}^3$ as a real vector space, the purely imaginary quaternions of unit length are a 2-sphere, S^2 . Now \mathbb{H}^\times acts on S^2 by conjugation, moreover, the action of \mathbb{H}^\times is transitive, (since we know that the image of \mathbb{H}^\times in $\text{GL}(\mathbb{R}^3)$ is all of $\text{SO}(\mathbb{R}^3)$) hence $S^2 \cong \mathbb{H}^\times / \text{Stab}_{\mathbb{H}^\times}(i)$, the quotient of \mathbb{H}^\times by the stabilizer of $i \in S^2$. But now this stabilizer is exactly \mathbb{C}^\times , the nonzero complex numbers (viewing $\mathbb{C} \subset \mathbb{H}$ in the standard way). Thus we have identified S^2 with the quotient $\mathbb{H}^\times / \mathbb{C}^\times$, and hence $\mathbb{C}P^1$

Definition 6.5. A *line* in projective space $\mathbb{P}(V)$ is the set of lines in V contained in a 2-plane. A *k-dimensional subspace* of a projective space is a subset consisting of the lines in V which are contained in a $k + 1$ -dimensional vector subspace of V . The set of k -dimensional subspaces of V , the *Grassmannians* of k -subspaces, is denoted $\text{Gr}_k(V)$ or $\mathbb{G}r_{k-1}(\mathbb{P}(V))$ (thinking of them as $(k - 1)$ -dimensional subspaces of $\mathbb{P}(V)$). A *projective geometry* is a projective space equipped with collection of all subspaces. We call a vector subspace of V a *hyperplane* if it has dimension $\dim(V) - 1$. Notice that subspaces are partially ordered by inclusion. We say that a *flag* of length k in $\mathbb{P}(V)$ is a sequence $F_0 \subsetneq F_1 \subsetneq \dots \subsetneq F_k$ of subspaces. Notice that $\dim(\mathbb{P}(V))$ is the maximal length of a flag in $\mathbb{P}(V)$.

Lemma 6.6. *If $\mathbb{P}(V)$ is a projective space, then any two distinct points determine a unique line. Moreover, in a projective plane, any two lines intersect in a point.*

Proof. Let p, q be distinct points in V , and let v and w be representative vectors in V . Then since p and q are distinct, v, w are linearly independent. The 2-plane they span gives the unique line through p and q . If $\mathbb{P}(V)$ is two dimensional, then a line in $\mathbb{P}(V)$ corresponds to a 2-dimensional subspace in V . For any two lines $\ell_1, \ell_2 \subset \mathbb{P}(V)$, let W_1 and W_2 be the planes in V they correspond to. Then

$$\dim(W_1 \cap W_2) = \dim(W_1) + \dim(W_2) - \dim(W_1 + W_2) = 4 - \dim(W_1 + W_2) = 1$$

unless $W_1 = W_2$, in which case the lines ℓ_1 and ℓ_2 are equal. \square

We now wish to define maps between projective spaces. In fact it turns out that it is easier to only define the notion of an isomorphism between projective spaces: Suppose that V and W are vector spaces, and T is any invertible linear map from V to W . Then clearly T induces a map between the subspaces of V and the subspaces of W given by $U \mapsto T(U)$. Since T has zero kernel, $\dim(T(U)) = \dim(U)$, and so in particular, T gives a map from $\mathbb{P}(V)$ to $\mathbb{P}(W)$.

Definition 6.7. Given vector spaces V, W , a *projective transformation* from $\mathbb{P}(V)$ to $\mathbb{P}(W)$ is the map τ induced by a linear isomorphism $T: V \rightarrow W$. We write $\tau = [T]$.

Remark 6.8. One could also make a more abstract definition by letting a projective transformation be a bijection $\phi: \mathbb{P}(V) \rightarrow \mathbb{P}(W)$ which takes collinear points in $\mathbb{P}(V)$ to collinear points in $\mathbb{P}(W)$. This gives a slightly larger class of transformations, which can nevertheless be completely described, by what is known as the *Fundamental Theorem of Projective Geometry*. We will not use this larger class of transformations.

Notice that if $S = \lambda T$ for some $\lambda \in k$, then T and S induce the same projective transformation. In fact this is the only time that two linear isomorphism induce the same projective transformation.

Lemma 6.9. (*Vandermonde determinant*): Suppose that $\lambda_1, \lambda_2, \dots, \lambda_m \in k$, and that A is the matrix

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_m \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{m-1} & \lambda_2^{m-1} & \dots & \lambda_m^{m-1} \end{pmatrix}$$

Then $\det(A) = \prod_{i>j}(\lambda_i - \lambda_j)$. In particular, A is invertible if all the λ_i are distinct.

Proof. Let $P(\lambda_1, \lambda_2, \dots, \lambda_m)$ be this determinant thought of as a function of the λ_i s. Then from the standard formula for the determinant we see that P is a sum of homogeneous terms of degree $1 + 2 + \dots + (m-1) = m(m-1)/2$ (each term in the expression for the determinant contains one entry from each row, and the entries in row i have degree i), i.e. P is a homogeneous polynomial of degree $m(m-1)/2$. Now since the determinant vanishes whenever two columns of a matrix are equal, it follows that P is divisible by $(\lambda_i - \lambda_j)$ for every pair i, j with $1 \leq j < i \leq m$. But now since a polynomial ring in m variables is a unique factorization domain, it follows that the product $\prod_{i>j}(\lambda_i - \lambda_j)$ divides P . Since this product clearly has degree $\binom{m}{2}$ it follows that the two expressions are equal up to a constant. Finally comparing the coefficient of $\lambda_2 \lambda_3^2 \dots \lambda_m^{m-1}$ we see that they are in fact equal. \square

Lemma 6.10. Let V be a vector space over k , and let $\alpha: V \rightarrow V$ be a linear map. Let V_λ be

$$\{v \in V : \alpha(v) = \lambda v\},$$

then the sum $\sum_{\lambda \in k} V_\lambda$ is direct.

Proof. Suppose that v_1, v_2, \dots, v_k are k vectors in eigenspaces $V_{\lambda_1}, V_{\lambda_2}, \dots, V_{\lambda_k}$ respectively, where all the scalars $\lambda_1, \lambda_2, \dots, \lambda_k$ are distinct. But then if we have $v_1 + v_2 + \dots + v_k = 0$, it follows that

$$\lambda_1^s v_1 + \lambda_2^s v_2 + \dots + \lambda_k^s v_k = 0,$$

for all $s > 0$. But since the λ_i s are all distinct, the previous lemma shows that the matrix $(\lambda_i^{j-1})_{1 \leq i, j \leq k}$ is invertible, and so the above equations hold only if $v_i = 0$ for all i . Thus the sum is direct as claimed. \square

Lemma 6.11. If S and T are isomorphism from V to W , then S and T induce the same projective transformation if and only if $S = \lambda T$ for some $\lambda \neq 0$.

Proof. By considering $A = T^{-1}S$ we are reduced to showing that if $A: V \rightarrow V$ is a linear map such that $[A]$ is the identity map, then A is a scalar multiple of the identity. Since $[A(v)] = [v]$ for each vector $v \in V$, it follows that every vector is an eigenvector. Thus V is the union of its eigenspaces, $V = \bigcup_{\lambda \in k} V_\lambda$. But then the previous lemma implies that $V = \bigoplus_{\lambda \in k} V_\lambda$, and these both happen only if $V = V_\lambda$ for some $\lambda \in k$, that is, if A is a scalar multiple of the identity as claimed. \square

Example 6.12. A geometric example of a projective transformation is given as follows. Take two lines $\mathbb{P}(U_1), \mathbb{P}(U_2)$ in a projective plane $\mathbb{P}(V)$, and let p be a point not on either lines. Then the map which takes a point $q \in \mathbb{P}(U_1)$ to the intersection

of $\mathbb{P}(U_2)$ with the line through p and q is a projective transformation from $\mathbb{P}(U_1)$ to $\mathbb{P}(U_2)$.

Definition 6.13. Let $\mathbb{P}(V)$ be an n -dimensional projective space. We say that $n + 2$ points in $\mathbb{P}(V)$ are in *general position* if each subset of $n + 1$ points has representative vectors which are linearly independent.

Lemma 6.14. If a_0, a_2, \dots, a_{n+1} are in general position in $\mathbb{P}(V)$ and b_0, b_2, \dots, b_{n+1} are in general position in $\mathbb{P}(W)$, then there is a unique projective transformation from $\tau: \mathbb{P}(V) \rightarrow \mathbb{P}(W)$ such that $\tau(a_i) = b_i$.

Proof. Choose arbitrary representative vectors $v_0, v_2, \dots, v_{n+1} \in V$ for the points a_0, a_2, \dots, a_{n+1} . By the general position assumption, the first $n + 1$ of these vectors form a basis of V . Thus we may write

$$v_{n+1} = \sum_{i=0}^n \lambda_i v_i.$$

Now if some $\lambda_i = 0$, then the vectors $\{v_0, \dots, v_{i-1}, v_{i+1}, \dots, v_{n+1}\}$ are not linearly independent, contradicting the general position assumption, hence we must have $\lambda_i \neq 0$ for all i . ($0 \leq i \leq n$). Replacing v_i by $\lambda_i v_i$, it follows that we may choose our representative vectors v_0, v_1, \dots, v_n so that

$$v_{n+1} = \sum_{i=0}^n v_i.$$

Notice moreover, that the choice of such representatives is unique up to simultaneous scaling. We may similarly find representative vectors w_0, w_1, \dots, w_n for the points b_0, b_1, \dots, b_n in $\mathbb{P}(W)$ such that w_0, w_1, \dots, w_{n+1} satisfy $w_{n+1} = \sum_{i=0}^n w_i$. Hence the linear map $T: V \rightarrow W$ defined by $T(v_i) = w_i$ for $i = 0, 1, \dots, n$ satisfies $[T](a_i) = b_i$ for all i ($0 \leq i \leq n + 1$), and $[T]$ is unique as required. \square

Remark 6.15. We call the group of projective transformations $\text{PGL}(V)$, the *projective general linear group*. The previous lemma shows that we may identify it with $\text{GL}(\mathbb{V})/Z(\text{GL}(V))$ where $Z(\text{GL}(V))$ is the center of $\text{GL}(V)$ – the group of scalar matrices (*check this*). Suppose now that \mathbb{P} is a projective line, then $n + 2 = 3$ and so there is a projective transformation τ taking any three distinct points to any other three points. Fixing any triple of points we see that we can realize any permutation of these three points as a projective linear transformation. In the case where $k = \mathbb{F}_2$, then a projective line has exactly three points, and so this is the entire group of projective linear transformations.

Theorem 6.16. (*Desargues*) Let A, B, C, A', B', C' be points in a projective space $\mathbb{P}(V)$ such that the lines AA', BB' and CC' are distinct and concurrent. Then the three points of intersection $AB \cap A'B', BC \cap B'C'$ and $AC \cap A'C'$ are all collinear.

Proof. Let P be the point on all the lines AA', BB' and CC' . Since P, A, A' are distinct points on a projective line, they are in general position, and so using the proof of the previous lemma, we may find representative vectors a, a', p such that $p = a + a'$. Similarly we can find representative vectors b, b', c, c' for B, B', C, C' such that

$$p = b + b'; \quad p = c + c'.$$

It follows that $a + a' = b + b'$ and so $a - b = b' - a' = c''$, and similarly

$$b - c = c' - b' = a''; \quad c - a = a' - c' = b''.$$

Thus it follows that $a'' + b'' + c'' = b - c + c - a + a - b = 0$, and so the three vectors a'', b'', c'' are linearly dependent, and thus lie in a two-dimensional subspace of V . It follows that the three points A'', B'', C'' represented by these vectors lie on the corresponding line in $\mathbb{P}(V)$. (It is easy to check that these points are the intersections described in the statement of the theorem. \square)

Remark 6.17. It is possible to define a projective space abstractly, using axioms similar to Euclid's: A projective space is a set (consisting of "points") with a collection of subsets (called "lines") such that:

- Any two distinct points lie on unique line.
- There is at least one line, and not all points lie on that line.
- A line contains at least three points.
- If A, B, C, D and E are five points such that B, C and D are collinear, A, C and E are collinear, but A, B and C are not collinear, then there is a point F collinear with D and E and with A and B .

(The last of these axioms is attempting to say when points lie on the same plane without having to use the word "plane"). We say that a subset of a projective space is a subspace if whenever it contains distinct points A and B , it contains the line through them (thus it is another projective space). A chain of nested subspaces ($F_0 \subset F_1 \subset \dots \subset F_k$) in a projective space is called a flag of length k . The dimension of a projective space is then the maximal length of a flag. It can be shown that if the projective space has dimension at least 3, then it is $\mathbb{P}(V)$ for some finite dimensional vector space over a skew-field.

Finally we discuss duality in the context of projective geometry. Given a finite dimensional vector space V over a field k recall that its *dual space* is the vector space $V^* = \text{Hom}(V, k)$. Given any basis v_1, v_2, \dots, v_n of V , we may define a *dual basis* $v_1^*, v_2^*, \dots, v_n^*$ of V^* by setting $v_i^*(v_j) = \delta_{ij}$. It is clear from this that V and V^* are vector spaces of the same dimension, however there is no natural isomorphism between them. We can however give a natural identification of subspaces.

Definition 6.18. Let $U \subset V$ be a subspace of V . Then set $U^\circ \subset V^*$, the *annihilator* of U to be the set

$$\{\phi \in V^* : \phi(u) = 0, \forall u \in U\}.$$

It is easy to check that U° is a subspace of V^* . Notice that taking annihilators reverses containment on subspaces: if $U_1 \subset U_2$ then $U_1^\circ \supset U_2^\circ$. Moreover we have the following easy result.

Lemma 6.19. Let V be a finite dimensional vector space over k and let $U \subset V$ be a subspace of V . Then

$$\dim(U) + \dim(U^\circ) = \dim(V).$$

Proof. Pick a basis $\{u_1, u_2, \dots, u_k\}$ of U , and extend it to a basis $\{u_1, u_2, \dots, u_n\}$ of V . Let $\{u_1^*, u_2^*, \dots, u_n^*\}$ be the dual basis. It is easy to check that $\phi \in V^*$ lies in U° if and only if $\phi \in \text{span}\{u_{k+1}^*, u_{k+2}^*, \dots, u_n^*\}$. The result follows. \square

Given a linear map $\alpha: V \rightarrow W$ between vector spaces, there is a natural map $\alpha^*: W^* \rightarrow V^*$ given by

$$\alpha^*(\phi)(v) = \phi(\alpha(v)), \quad \forall v \in V.$$

If α is an isomorphism, it is easy to check that if U is a subspace of V and $\alpha(U) = S \subset W$, then $\alpha^*(S^\circ) = U^\circ$.

Although there is not a natural isomorphism between V and V^* , if we dualize again and consider $V^{**} = (V^*)^*$, then in fact there is a natural map $S: V \rightarrow V^{**}$ given by

$$S(v)(\phi) = \phi(v).$$

It is immediate that S is linear, and moreover it is injective. Hence if V is finite dimensional, S is an isomorphism.

Given a point in $\mathbb{P}(V)$, that is, a line in V , the annihilator of this line is a subspace of dimension $\dim(V) - 1$ in V^* , *i.e.* a hyperplane in V^* . Similarly, given a point $p \in \mathbb{P}(V^*)$, its annihilator in $V^{**} \cong V$ is a hyperplane in V , thus we see that the set of hyperplanes in V can naturally be identified with a projective space.

The most concrete consequence of duality in projective geometry is seen in the case of a projective plane $\mathbb{P}(V)$. Then V has dimension 3, and so duality gives a correspondence between points of $\mathbb{P}(V)$ and lines in $\mathbb{P}(V^*)$, and similarly a correspondence between lines in $\mathbb{P}(V)$ and points in $\mathbb{P}(V^*)$. Thus we see that the statement that any two points lie on a unique line by duality implies that any two lines intersect in a point, and so the two statements of Lemma 6.6 are the duals of each other – we need only have proved one of them for the case of a projective plane.

As a more elaborate example, consider Desargues' theorem. The dual theorem says that if we have six lines $l_1, l_2, l_3, l'_1, l'_2, l'_3$ such that the intersections $p_1 = l_1 \cap l'_1, p_2 = l_2 \cap l'_2$ and $p_3 = l_3 \cap l'_3$ are distinct collinear points, then if $p_{ij} = l_i \cap l_j$ and $p'_{ij} = l'_i \cap l'_j$, the lines l_{ij} given by p_{ij}, p'_{ij} for $1 \leq i < j \leq 3$ are concurrent. This can be thought of as the converse to Desargues' theorem.

Finally we use duality to understand a very natural space: the space of all lines (not necessarily through the origin) in a real plane. We know from our discussions above that this space corresponds to the lines in $\mathbb{R}P^2$, with the "line at infinity" removed. By duality, this is the same space as $\mathbb{R}P^2$ with a point removed. What does this space look like? If we use the model of the 2-sphere S^2 with antipodal point identified, we see that the space is just the sphere minus the poles with antipodal points identified. To make this more explicit, we can use spherical coordinates:

$$(\theta, \phi) \mapsto (\cos(\theta) \cos(\phi), \sin(\theta) \cos(\phi), \sin(\phi)), \quad \theta \in [0, 2\pi] \times (0, \pi).$$

When we remove the poles, the range of values for θ and ϕ are $0 \leq \theta < 2\pi$, $0 < \phi < \pi$, and the antipodal map corresponds to the map

$$(\theta, \phi) \mapsto (\theta + \pi, \pi - \phi).$$

(where the first component must be read modulo 2π). But then we may identify the space of lines in \mathbb{R}^2 with pairs $(\theta, \phi) \in [0, \pi] \times (0, \pi)$ where we identify $(0, \phi)$ with $(\pi, \pi - \phi)$. Drawing a picture of a square with the appropriate identifications, we immediately see that this space is a Möbius band.

7. THE GENERAL LINEAR GROUP

Recall that given a vector space V over a field k the group $GL(V)$ is the set of invertible linear transformations of V . By picking a basis of V , we may identify $GL(V)$ with the group of invertible $n \times n$ matrices, where $n = \dim(V)$.

The determinant function gives a homomorphism $\det: GL(V) \rightarrow k^\times$. The kernel of \det is the *special linear group* (over \mathbb{R} these are the volume preserving linear transformations). The centers of these groups are easy to compute:

$$Z(GL(V)) = \{\lambda I : \lambda \in k^\times\}; \quad Z(SL(V)) = \{\lambda I : \lambda^n = 1\}$$

(where $n = \dim(V)$ as above). Thus the quotient $PGL(V)$ of $GL(V)$ by its center is the group of projective transformations, which contains the group $PSL(V)$ of projective special linear transformations. We want to study these groups over a finite field, $k = \mathbb{F}_q$, where $q = |\mathbb{F}_q|$.

Lemma 7.1. (1) $|GL(\mathbb{F}_q^n)| = q^{n(n-1)/2} \prod_{k=1}^n (q^k - 1)$.
 (2) $|SL(\mathbb{F}_q^n)| = q^{n(n-1)/2} \prod_{k=2}^n (q^k - 1)$.
 (3) $|PSL(\mathbb{F}_q^n)| = |SL(\mathbb{F}_q^n)| / g.c.d(n, q - 1)$.

Proof. The last two parts are immediate from the first. To obtain the first part, we count the number of invertible $n \times n$ matrices over \mathbb{F}_q . Note that such a matrix is invertible if and only if its columns are linearly independent. The first column can thus be any nonzero vector, that is one of $q^n - 1$ vectors. The second column can be any vector not on the line spanned by the first column, and therefore is one of $q^n - q$ vectors. Continuing in this way, we see that the k -th vector can be any vector not lying in the span of the first $k - 1$ vectors, which are linearly independent and so span a $(k - 1)$ dimensional subspace. Hence there are $q^n - q^{k-1}$ choices. Thus we see that in total there are

$$\prod_{i=1}^n (q^n - q^{i-1}) = q^{n(n-1)/2} \prod_{i=1}^n (q^i - 1)$$

invertible $n \times n$ matrices as required. \square

Notice that $GL(V)$ has two natural subgroups – $Z(GL(V))$ and $SL(V)$. In order to obtain a group which does not obviously possess a nontrivial normal subgroup, we must pass to the quotient of the group $SL(V)$ by its center, $PSL(V)$. We will now show that this group is in fact simple in almost all cases. To do this we need to gather a collection of elements of $GL(V)$ which we can understand explicitly.

Definition 7.2. Let $\rho \in GL(V)$. Let $\chi(t)$ be the minimal polynomial of ρ , that is the polynomial of smallest degree in $k[t]$ such that $\chi(\rho) = 0$. We say that ρ is *semisimple* if the polynomials $\chi(t)$ and $\chi'(t)$ (the derivative of χ) are relatively prime. If k is algebraically closed then this is equivalent to requiring that there is a basis of V consisting of eigenvalues of ρ . ρ is said to be *unipotent* if there is some $m > 0$ such that $(\rho - 1)^m = 0$.

Definition 7.3. Consider a transformation $\tau \in GL(V)$ which fixes every element of a hyperplane $H \subset V$. If we pick $v \in V \setminus H$ and define $\phi: V \rightarrow k$ by

$$\phi(\lambda v + h) = \lambda, \quad \forall \lambda \in k, h \in H.$$

Thus $H = \ker(\phi)$ and for all $w \in V$ we have $w - \phi(w)v \in H$. Thus we have $\tau(w - \phi(w)v) = w - \phi(w)v$, and hence

$$\tau(w) = w + \phi(w)u.$$

where $u = \tau(v) - v$. Thus we may specify any invertible linear map which fixes a hyperplane by giving $\phi \in V^*$ and $u \in V$, such that $\phi(u) \neq -1$. Write $\tau_{u,\phi}$ for the map obtained by using $u \in V, \phi \in V^*$ in the above formula. If $\phi(u) = 0$ then we say that ϕ is a *transvection*. If $\phi(u) \notin \{0, -1\}$ then we say that $\tau_{u,\phi}$ is a *dilatation*. Thus a transvection is a linear map τ fixing a hyperplane H such that $\tau(v) - v \in H$ for all $v \in V$. Note that $(\tau - 1)^2 = 0$, and so in particular, τ is unipotent. It follows that a linear map fixing a hyperplane is either semisimple, in which case it is a dilatation, or unipotent, in which case it is a transvection.

We record the following obvious properties of transvections.

Lemma 7.4. *Let $\tau_{u,\phi}$ be a transvection. Then we have the following.*

- (1) $\tau_{au,\phi} = \tau_{u,a\phi}$;
- (2) $\tau_{u,\phi_1+\phi_2} = \tau_{u,\phi_1} \circ \tau_{u,\phi_2}$;
- (3) $\tau_{u_1+u_2,\phi} = \tau_{u_1,\phi} \circ \tau_{u_2,\phi}$;
- (4) $g\tau_{u,\phi}g^{-1} = \tau_{g(u),g^*(\phi)}$, for all $g \in GL(V)$.

Proof. The first property holds for any $\tau_{u,\phi}$. For the second, observe that

$$\begin{aligned} \tau_{\phi_1,u} \circ \tau_{\phi_2,u}(w) &= \tau_{\phi_1,u}(w + \phi_2(w)u) \\ &= w + \phi_1(w + \phi_2(w)u)u \\ &= w + (\phi_1 + \phi_2)(w)u, \end{aligned}$$

since $\phi_2(u) = 0$. The third property follows in a similar way. Finally we have

$$\begin{aligned} g\tau_{u,\phi}g^{-1}(w) &= g\tau_{u,\phi}(g^{-1}(w)) \\ &= g(g^{-1}(w) + \phi(g^{-1}(w))u) \\ &= w + g^*(\phi)(w)g(u) \\ &= \tau_{g^*(\phi),g(u)}(w). \end{aligned}$$

□

Notice that if $\tau = \tau_{\phi,u}$ is a transvection, then we have a homomorphism of groups $e_\tau: \mathbf{k} \rightarrow SL(V)$ given by $a \mapsto \tau_{u,a\phi}$. We call the images of these homomorphisms *unipotent one-parameter subgroups*. It should be pointed out that if u and ϕ are put into bases of V and V^* which are dual to each other, then the matrix of $t_{u,\phi}$ is very simple: it has 1s on the main diagonal and all but one of the off-diagonal entries equal to zero. It is easy to see from this that any transvection has determinant 1, and so $\tau_{u,\phi} \in SL(V)$ for all $u \in V, \phi \in V^*$ such that $\phi(u) \notin \{0, -1\}$.

Lemma 7.5. *If $\dim(V) = n \geq 2$ then all transvections are conjugate in $GL(V)$. If $\dim(V) \geq 3$, they are all conjugate in $SL(V)$. When $n = 2$, the conjugacy classes of transvections in $SL(V)$ are in bijection with $\mathbf{k}^\times / (\mathbf{k}^\times)^2$.*

Proof. Let τ_1 and τ_2 be two transvections in $SL(V)$, and let W_1 and W_2 be the hyperplanes fixed by τ_1 and τ_2 respectively. Choose $v_i \in V \setminus W_i$ and let $w_i = \tau_i(v_i) - v_i \in W_i$, for $i = 1, 2$. Then define $g \in SL(V)$ by picking a basis $\{w_1, u_1, u_2, \dots, u_{n-2}\}$ of W_1 and a basis $\{w_2, r_1, r_2, \dots, r_{n-2}\}$ of W_2 and letting $g(w_1) = w_2, g(v_1) = v_2$, and

$g(u_i) = r_i$ for $1 \leq i \leq n-3$, and $g(u_{n-2}) = ar_{n-2}$ where a is chosen so $\det(g) = 1$. It then readily follows that $g\tau_1g^{-1} = \tau_2$.

If $n = 2$ then we may still use this procedure to obtain a $g \in GL(V)$ conjugating τ_1 to τ_2 , but we cannot guarantee that $g \in SL(V)$. Thus we need to think about the case of $SL(V)$ where $\dim(V) = 2$ separately. Fix a basis $\{e_1, e_2\}$ of V with dual basis $\{e_1^*, e_2^*\}$, and let $\tau_a = \tau_{e_1, ae_2^*}$ for $a \in k^\times$. Suppose that $\tau_{u, \phi}$ is any transvection. Pick $g \in SL(V)$ with $g(u) = e_1$. Then $g\tau_{u, \phi}g^{-1} = \tau_{e_1, g^*(\phi)}$ (by Lemma 7.4), and so, since $g^*(\phi)(e_1) = \phi(u) = 0$, we see that $g^*(\phi) = \lambda e_2^*$ for some $\lambda \in k^\times$, that is $g\tau_{u, \phi}g^{-1} = \tau_\lambda$. Hence we are reduced to finding which of the τ_a are conjugate. But if $g \in SL(V)$ is such that $g\tau_a g^{-1} = \tau_b$, it follows that the matrix of g with respect to the basis $\{e_1, e_2\}$ has the form

$$\begin{pmatrix} \lambda & \mu \\ 0 & \lambda^{-1} \end{pmatrix}$$

and hence $b = \lambda^2 a$. It follows that τ_a and τ_b are conjugate if and only if a and b differ by a square in k^\times . Thus the conjugacy classes of transvections in $SL(V)$ are in bijection with the set $k^\times / (k^\times)^2$. In the case where k is finite, it is easy to see that $k^\times / (k^\times)^2$ has two elements, so that there are two conjugacy classes of transvections. \square

- Proposition 7.6.** (1) *Given any two linearly independent vectors $v, w \in V$ there is a transvection τ taking v to w .*
(2) *Given two distinct hyperplanes W_1, W_2 and a vector v such that $v \notin W_1 \cup W_2$, there is a transvection τ fixing v taking W_1 to W_2 .*
(3) *The set of transvections generates $SL(V)$.*

Proof. For (1), since $\{v, w\}$ are linearly independent, so are $\{u, v\}$ where $u = v - w$, and thus we can extend this set to a basis $\{u, v, w_1, w_2, \dots, w_k\}$ of V . Then define $\tau: V \rightarrow V$ by setting $\tau(u) = u, \tau(w_i) = w_i$ for all $i, (1 \leq i \leq k)$, and $\tau(v) = w$. Then clearly τ fixes a hyperplane $H = \text{span}\{u, w_1, w_2, \dots, w_k\}$ and $\tau(v) - v = u \in H$ so that τ is a transvection.

For the second part, since $W_1 + W_2$ must be all of V (unless $\dim(V) = 1$, in which case the statement is trivial), we may find $w_1 \in W_1$ and $w_2 \in W_2$ such that $v = w_2 - w_1$. Since $v \notin W_1, W_2$ we cannot have w_1 or $w_2 \in W_1 \cap W_2$. It follows that we may define $\tau: V \rightarrow V$ by setting $\tau|_{W_1 \cap W_2}$ to be the identity, $\tau(v) = v$, and $\tau(w_1) = w_2$. Since τ fixes the hyperplane spanned by $W_1 \cap W_2$ and v , and $\tau(w_1) - w_1 = v$, it follows that τ is a transvection as required.

Finally, to show that the set of transvections generates $SL(V)$ we use induction on $\dim(V)$. Let $\rho \in SL(V)$, let $\{e_1, e_2, \dots, e_n\}$ be a basis of V , and let $v_i = \rho(e_i)$ ($1 \leq i \leq n$). If $\rho(e_1) = e_1$, let $\rho_1 = \rho$. Otherwise, if $\rho(e_1)$ and e_1 are linearly dependent, apply a transvection τ_1 so that $\rho(e_1)$ and $\tau_1(\rho(e_1))$ are linearly independent, and set $\rho' = \tau_1 \circ \rho$. Then since $\rho'(e_1)$ and e_1 are linearly independent, we may find a transvection τ_2 such that $\tau_2 \rho'(e_1) = e_1$, using part (1). In this case, set $\rho_1 = \tau_2 \rho'$, so that $\rho_1(e_1) = e_1$.

Let $W_1 = \text{span}\{e_2, e_3, \dots, e_n\}$, and let $W_2 = \text{span}\{\rho_1(e_2), \rho_1(e_3), \dots, \rho_1(e_n)\}$. If $W_1 \neq W_2$, then by part (2) we may find a transvection τ_3 such that $\tau_3(W_1) = W_2$, and $\tau_3(e_1) = e_1$. Replacing ρ_1 by $\rho_2 = \tau_3 \rho_1$ we see that ρ_2 preserves the hyperplane W_1 , and $\rho_2(e_1) = e_1$. It follows that $(\rho_2)|_{W_1} \in SL(W_1)$, and thus by induction $(\rho_1)|_{W_1}$ is a product of transvections, $(\rho_1)|_{W_1} = \sigma_1 \sigma_2 \dots \sigma_k$. Defining $\sigma_i(e_1) = e_1$,

we can extend the σ_i to transvections on V , so that $\rho_2 = \sigma_k^{-1} \sigma_{k-1}^{-1} \dots t_1^{-1} \rho_2$ is the identity on V since it fixes e_1 and W_1 , which together span V . Hence ρ_2 , and therefore ρ is a product of transvections, and we see that $\text{SL}(V)$ is generated by transvections as required. \square

Remark 7.7. Since multiplication by a transvection corresponds to an elementary row operation on matrices (choosing an appropriate basis), the previous proposition is essentially a “geometric” way of proving that every invertible matrix can be reduced to the identity matrix by row operations.

Corollary 7.8. *Suppose that $\dim(V) > 2$ or $\dim(V) = 2$ and $|\mathbf{k}| > 3$, then the group $\text{SL}(V)$ is equal to its own derived subgroup.*

Proof. Let $G = \text{SL}(V)$, and let G' be its derived group. Let $\tau_{u,\phi}$ be a transvection. Since all transvections are conjugate, we can find a $g \in \text{SL}(V)$ so that $g\tau_{u,\phi}g^{-1} = \tau_{u,\psi}$, where $\psi(u) = 0$ but $\psi \neq \phi$ (such a ψ exists and is nonzero provided $\dim(V) \geq 3$ or $|\mathbf{k}| > 2$). But then using Lemma 7.4 we see that

$$g\tau_{u,\phi}g^{-1}\tau_{u,\phi}^{-1} = \tau_{u,\psi-\phi} \in G'$$

Since G' is normal, and $\tau_{u,\psi-\phi} \in G'$, the conjugacy class of all transvections lies in G' . But since these generate G , it follows that $G' = G$ as required.

Again the case when $\dim(V) = 2$ needs separate attention. We claim that $\text{SL}_2(\mathbf{k})$ is equal to its own derived subgroup whenever $|\mathbf{k}| > 3$. We use the notation of the proof of Lemma 7.5 above. Notice that if

$$g = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda^{-1} \end{pmatrix},$$

then $g\tau_a g^{-1} = \tau_b$ where $b = \lambda^2 a$, and so for $a \in \mathbf{k}$, we have $g\tau_a g^{-1}\tau_a^{-1} = \tau_{\lambda^2 a - a}$. Now the equation $a(\lambda^2 - 1) = b$ can be solved for any $b \in \mathbf{k}^\times$ with $a, \lambda \in \mathbf{k}^\times$ provided $|\mathbf{k}| > 3$. Therefore if \mathbf{k} has more than 3 elements, the derived subgroup of $\text{SL}(V)$ contains every τ_b , and hence since it is a normal subgroup, all transvections. Thus $\text{SL}(\mathbf{k}^2)$ is also its own derived group provided $|\mathbf{k}| > 3$. \square

Finally, we prove that $\text{SL}(V)$ is a simple group. We use a preparatory lemma on the action of $\text{PSL}(V)$ on $\mathbb{P}(V)$.

Lemma 7.9. *The group $\text{PSL}(V)$ acts 2-transitively on $\mathbb{P}(V)$. Hence the stabilizer of a point in $\mathbb{P}(V)$ is a maximal subgroup of $\text{PSL}(V)$.*

Proof. We showed before that the group of projective transformations acts transitively on $(n+1)$ -tuples of points of $\mathbb{P}(V)$ in general position. Any two points p_1, p_2 can be completed to a set of points in general position $\{p_1, p_2, \dots, p_{n+1}\}$, and so given two pairs of points (p_1, p_2) and (q_1, q_2) we may complete them to sets of points $\{p_1, p_2, \dots, p_{n+1}\}$ and $\{q_1, q_2, \dots, q_{n+1}\}$ in general position. Thus there is a unique projective transformation τ taking p_i to q_i . Let T be a linear transformation such that $[T] = \tau$, and suppose that $\det(T) = a \in \mathbf{k}$. Pick a representative vector for p_1 say v_1 , we may define $S \in \text{GL}(V)$ to be the identity on the lines p_2, p_3, \dots, p_n , and let $S(v_1) = a^{-1}v_n$. Then $\det(S) = a^{-1}$, and hence ST has $\det(ST) = 1$, while $[TS](p_i) = q_i$ for each i , $(1 \leq i \leq n)$. This shows that the action is 2-transitive as required.

To see that this implies the stabilizers are maximal subgroups, suppose that $P = \text{Stab}_{\text{PSL}(V)}(\ell)$ for some $\ell \in \mathbb{P}(V)$, and let $P < K \leq \text{PSL}(V)$. Since $P \subsetneq K$, we may

take $k \notin P$, so that $k(\ell) \neq \ell$. Then if $g \in \text{PSL}(V) \setminus P$, by 2-transitivity there is a $g' \in \text{PSL}(V)$ such that $g'(\ell) = \ell$ and $g'(g(\ell)) = k(\ell)$. But then $g' \in P$, and since $k^{-1}g'g \in P$ we have $g \in K$, and so since g was arbitrary, $K = G$ as claimed. \square

Theorem 7.10. (L.E.Dickson). *Suppose that $n = \dim(V) \geq 2$, and $|k| > 3$ if $n = 2$. Then the group $\text{PSL}(V)$ is simple.*

Proof. Let $v \in V - \{0\}$, and set $P = \text{Stab}([v])$, the stabilizer of the line through v . Suppose that $K \triangleleft \text{PSL}(V)$ is a normal subgroup. We consider two cases: either $K \subseteq P$ or $K \not\subseteq P$.

In the first case, we have $K[v] = [v]$, and so since K is normal, $K([v]) = [v]$ for all $[v] \in \mathbb{P}(V)$, (since $\text{PSL}(V)$ acts 2-transitively and so certainly transitively, on $\mathbb{P}(V)$). As the action of $\text{PSL}(V)$ on $\mathbb{P}(V)$ is faithful, it follows that $K = \{1\}$. In the second case, since the previous lemma shows that P is maximal, we must have $PK = \text{PSL}(V)$. But then if π denotes the quotient map $\pi: \text{PSL}(V) \rightarrow \text{PSL}(V)/K$, we must have $\pi(P) = \text{PSL}(V)/K$. Now let

$$N = \{\tau \in \text{PSL}(V) : \tau = [\tau_{e_1, \phi}], \phi \in V^*, \phi(e_1) = 0\}.$$

Then by Lemma 7.4, N is an abelian normal subgroup of P , and moreover every projective transvection (*i.e.* image of a transvection in $\text{PSL}(V)$) is conjugate to one in N . Thus the conjugates of N generate $\text{PSL}(V)$ by Proposition 7.6, and hence the conjugates of $\pi(N)$ generate $\pi(\text{PSL}(V))$. But since $\pi(\text{PSL}(V)) = \pi(P)$ and $N \triangleleft P$, $\pi(N) \triangleleft \pi(P)$, and so in fact $\pi(N) = \pi(P)$. Hence $\text{PSL}(V) = KN$. But then the derived group of $\text{PSL}(V)$ is contained in KN' , and since N is abelian, it follows that N' is trivial, and so the derived group is contained in K (indeed you can check that in general if $K \triangleleft KN$ then $(KN)'\subseteq KN'$). But by Corollary 7.8 we know that $\text{PSL}(V)$ is its own derived group, hence $K = \text{PSL}(V)$, and $\text{PSL}(V)$ is simple as claimed. \square

In fact the groups $\text{PSL}(\mathbb{F}_2^2)$ and $\text{PSL}(\mathbb{F}_3^2)$ are genuine exceptions – they are not simple groups. We have already seen that $\text{PSL}(\mathbb{F}_2^2) \cong S_3$ a solvable group, so it only remains to show that $\text{PSL}(\mathbb{F}_3^2)$ is not simple. But $\mathbb{P}(\mathbb{F}_3^2)$ has 4 elements, so the action of $\text{PSL}(\mathbb{F}_3^2)$ on $\mathbb{P}(\mathbb{F}_3^2)$ gives an injective homomorphism from $\text{PSL}(\mathbb{F}_3^2)$ to S_4 , the symmetric group on 4 letters. By our formulas for the orders of linear groups, we know $|\text{PSL}(\mathbb{F}_3^2)| = (3 \cdot 8/2) = 12$, and so we must have $\text{PSL}(\mathbb{F}_3^2) \cong A_4$, which is solvable. Notice that $\text{PSL}(\mathbb{F}_2^3) = \text{PGL}(\mathbb{F}_2^3)$ is the group of projective transformations of the projective plane over \mathbb{F}_2 . This is the smallest possible projective plane, having $7 = 2^3 - 1$ points (this is true even in the sense of abstract projective planes). This group has 168 elements, and it is isomorphic to $\text{PSL}(\mathbb{F}_7^2)$.

Definition 7.11. Let \mathcal{B} be the set of all complete flags in k^n , that is

$$\mathcal{B} = \{(0 = F_0 \subsetneq F_1 \subsetneq \dots \subsetneq F_n = V) : F_i \text{ subspaces of } V\}$$

(it follows automatically that $\dim(F_i) = i$).

Given a complete flag F_\bullet let $B = B(F_\bullet)$ be the stabilizer of F_\bullet in $GL(V)$. For any basis $\{v_1, v_2, \dots, v_n\}$ we may associate a flag F_\bullet by setting $F_i = \text{span}\{v_1, v_2, \dots, v_i\}$. It is easy to check that the stabilizer of the flag associated to the standard basis in this way (sometimes called the standard flag) is exactly the set of linear maps α whose matrices are upper triangular, that is, if $A = (a_{ij})$ is the matrix associated to α , then $a_{ij} = 0$ for all $j > i$. The groups which arise in this way are called

Borel subgroups of $GL(k^n)$. Since $GL(k^n)$ acts transitively on bases of k^n , the Borel subgroups are all conjugate.

Given a basis $\{v_1, v_2, \dots, v_n\}$ we may also associate a smaller subgroup of $GL(k^n)$. Let $T = \{g \in GL(k^n) : g(v_i) = \lambda v_i, \text{ some } \lambda \in \mathbb{R}\}$, that is, T is the subgroup of matrices which preserve the lines spanned by the basis vectors. Clearly, T is an Abelian subgroup of $GL(k^n)$. The group T is called a *torus* in $GL(k^n)$, and the torus attached to the standard basis is known as the standard torus. As for Borel subgroups, the tori in $GL(k^n)$ are all conjugate.

Lemma 7.12. *Let N be the normalizer of T the standard torus in $GL_n(k)$, and assume that $|k| > 2$. Then N is subgroup of monomial matrices, that is, if $A \in N$, then each row and each column of A contain exactly one nonzero entry. Moreover, N/T is isomorphic to S_n , the symmetric group on n letters.*

Proof. Let $n \in N$. Fix $i, 1 \leq i \leq n$, and choose $t \in T$ such that $t(e_i) = \lambda e_i$ where $\lambda \neq 1$, and $t(e_j) = e_j$ for all $j \neq i$. Then $t' = ntn^{-1}$ fixes a hyperplane and has $n(e_i)$ as an eigenvector with eigenvalue λ . Since $t' \in T$, each e_j is an eigenvector of t' , hence we must have $n(e_i) = \mu_i e_j$ for some $\mu_i \in k$, and some $j, 1 \leq j \leq n$. Let $\sigma: \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ be the map given by $n(e_i) = \mu_i e_{\sigma(i)}$. Then since $n \in GL_n(k)$ σ must be an injection, it follows that σ is a bijection, that is, $\sigma \in S_n$. Hence N consists of monomial matrices as claimed. For $\rho \in S_n$, let $w_\rho \in GL(k^n)$ be the linear map given by $w_\rho(e_i) = e_{\rho(i)}$. Thus if we let $s \in T$ be the diagonal matrix with diagonal entries μ_i , we have shown $n = w_\rho s$.

Let \bar{W} be the image of the homomorphism $\rho \mapsto w_\rho$. We have shown that $N = \bar{W}T$, and clearly T is normal in N so that $N/T \cong \bar{W} \cong S_n$ as required. \square

The group N/T is called the *Weyl group* of $GL(k^n)$, and is denoted W . We now use W to analyze the double cosets of B the standard Borel in $GL(k^n)$. Let U be the subgroup of B consisting of the unipotent elements, in other words it is the subgroup of element of B with 1s on the main diagonal. Let U^- be the subgroup of $GL(k^n)$ consisting on *lower triangular* matrices with 1s on the main diagonal. Finally, let $U_w = U \cap wU^-w^{-1}$.

Proposition 7.13. (*Bruhat Decomposition*): *Any $g \in GL(k^n)$ can be written as a product nwb where $u \in U$, $w \in W$ and $b \in B$. Moreover w is uniquely determined, and if we insist that $u \in U_w$ then b and u are also.*

Proof. Given a matrix $g \in GL(k^n)$, right multiplication by elements of B correspond to column operations on the matrix of g , where we can add to a column multiples of any column to its right. We may therefore use these column operations to transform a matrix into the form

$$\begin{pmatrix} * & * & * & 1 \\ * & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

that is, the matrix has column vectors c^i ($1 \leq i \leq n$) such that:

- (1) Each c^i ends in a 1, say in row $\pi(i)$, followed by zeros.
- (2) The entries to the right of the 1 in a column are all zero, i.e. $c_{\pi(i)}^j = 0$ for $j > i$.

It is clear that we can obtain a unique matrix of this form (known as *reduced echelon form*) starting with any matrix, recursively working from left to right through the columns. Since the column operations used define an element of B , we may write our reduced echelon form matrix as gb . Moreover the function $\pi: \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ $\pi(i)$ is clearly a bijection. If we multiply gb on the right by w_π^{-1} , we obtain a matrix u of the form

$$\begin{pmatrix} 1 & * & * & * \\ 0 & 1 & * & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

Thus we obtain the product $g = uw_\pi b^{-1}$, where $u \in U_w = U \cap w_\pi(U^-)w_\pi^{-1}$. This last condition simply translates the conditions of reduced echelon form, which can be expressed in the form $(gb)w_\pi^{-1} \in U$ and $w_\pi^{-1}(gb) \in U^-$. \square

Remark 7.14. This last result gives us two ways of computing the number of elements in \mathcal{B} : There are clearly $|\mathrm{GL}_n(\mathbb{F}_q)|/|B(\mathbb{F}_q)|$ complete flags, since $\mathrm{GL}_n(\mathbb{F}_q)$ acts transitively on \mathcal{B} . Since $B(\mathbb{F}_q)$ has $(q-1)^n q^{n(n-1)/2}$ elements, we find that \mathcal{B} has $\prod_{i=1}^n (q^n - 1)/(q - 1)$ elements. On the other hand, by the previous proposition, each coset gB has a unique representative of the form uw where $w \in W$ and $u \in U \cap wU^-w^{-1}$, thus another way to count the number of complete flags is to count how many of these representatives there are. But $U \cap wU^-w^{-1}$ clearly has $q^{l(w)}$ elements, where

$$l(w_\pi) = l(\pi) = |\{(i, j) : 1 \leq i < j \leq n, \pi(i) > \pi(j)\}|,$$

is the number of *descents* of π . It follows that

$$\prod_{i=1}^n \frac{q^i - 1}{q - 1} = \sum_{\pi \in \mathcal{S}_n} q^{l(\pi)},$$

as both expression count the number of elements of $\mathcal{B}(\mathbb{F}_q)$. Since this equality holds for any prime-power, it is an identity of polynomials in an indeterminate q .

Finally, the decomposition $\mathcal{B} = \bigsqcup_{w \in W} U_w w B$ expresses \mathcal{B} as a disjoint union of pieces isomorphic to $k^{l(w)}$ for any field k , not just a finite field. If k is \mathbb{R} or \mathbb{C} this can be used to compute topological invariants of \mathcal{B} , such as its cohomology.

Corollary 7.15. *Let B be the standard Borel subgroup, W is the Weyl group, and T is the standard torus.*

- (1) *We have $GL(V) = BWB$. The double cosets are indexed by W , and hence there are $n!$ of them.*
- (2) *$B = TU$, where U , the subgroup of unipotent elements of B , is the derived group of B , a normal subgroup of B .*
- (3) *$N = T\tilde{W}$, where T is a normal subgroup of N .*

Proof. The first part is immediate from the previous proposition. The second part follows by an easy computation, while the third has already been observed. \square

We now define the notion of a BN -pair, first introduced by Jacques Tits.

Definition 7.16. Let G be a group, and B, N subgroups such that

- (1) B, N generate G .
- (2) $T = B \cap N$ is a normal subgroup of N .

- (3) $W = N/T$ is a finite group generated by a finite set S of involutions $\{s_i : i \in I\}$.
- (4) For each $w \in W$ we may pick representatives $\dot{w} \in N$. Then we must have
- (a) $\dot{s}B\dot{w} \subseteq B\dot{w}B \cup B\dot{s}\dot{w}B$ for any $s \in S$ and $w \in W$.
 - (b) $\dot{s}B\dot{s} \neq B$ for any $s \in S$.

Proposition 7.17. *Let $G = GL(V)$, B the standard Borel subgroup, and N the normalizer of T the standard torus. The (B, N) is a BN -pair for $GL(V)$.*

Proof. The first condition follows from the fact the identity $GL(V) = BWB$, since W is a quotient of N . Clearly the intersection of B and N is T , the standard torus, which is by definition normal in N , so it remains to check the third and fourth condition. For the third, we have seen that W is isomorphic to the symmetric group on n letters. If we set $s_i = (i, i+1)$, the transposition which interchanges i and $i+1$ fixing everything else, then it is easy to check that $S = \{s_i : 1 \leq i \leq n-1\}$ generates W (identified with S_n) and clearly each s_i is an involution. Let I denote the set $\{1, 2, \dots, n-1\}$, so that I indexes the generators in the set S .

Thus it remains to check the last axiom for a BN -pair. For this we define *root subgroups* X_{ij} where $1 \leq i \neq j \leq n$. These are the unipotent one-parameter subgroups given by pairs of elements of the standard basis:

$$X_{ij} = \{\tau_{e_i, ae_j^*} : a \in \mathfrak{k}\}$$

Thus each X_{ij} is a subgroup of $GL(V)$ which is isomorphic to \mathfrak{k} , and moreover $X_{ij} \subset U$ if $i < j$ and $X_{ij} \subset U^-$ if $i > j$. Set $X_i = X_{i, i+1}$ and $X_{-i} = X_{i+1, i}$, and write $X_{ij}(a)$ (or $X_i(a)$ if $j = i+1$, etc.) for τ_{e_i, ae_j^*} . For each $i \in I$ let $U_i = \{u \in U : e_i^*(u(e_{i+1})) = 0\}$. We claim first that $U = U_i X_i$. To see this, note that if $u \in U$, then if $v \in F_i$ (the i -th term in the standard flag) we have $u(v) - v \in F_{i-1}$, thus in particular $u(e_{i+1}) = e_{i+1} + \sum_{j \leq i} \lambda_j e_j$ for some $\lambda_j \in \mathfrak{k}$. Then consider $u' = uX_i(-\lambda_i)$. Then we have

$$\begin{aligned} u'(e_{i+1}) &= u(e_{i+1} - \lambda_i e_i) \\ &= e_{i+1} + \sum_{j \leq i} \lambda_j e_j - \lambda_i u(e_i) \\ &= e_{i+1} + \sum_{j < i} \lambda_j e_j + \lambda_i (u(e_i) - e_i), \end{aligned}$$

which clearly lies in $e_{i+1} + F_{i-1}$, we have $e_i^*(u'(e_{i+1})) = 0$, and so $u' \in U_i$. Since $u = u'X_i(\lambda_i)$ we have established $U = U_i X_i$.

We are now ready to verify the last of the BN -pair axioms. Clearly (a) and (b) are independent of the choice of representatives \dot{w} , we are free to chose them as we please. For $w \in W \cong S_n$, let \dot{w} be the corresponding element of $\tilde{W} \subset N$, so that $\dot{w}T = w$. Then we have $U_i = U \cap \dot{s}_i U \dot{s}_i$. Since $\dot{s}_i u \dot{s}_i = \dot{s}_i u \dot{s}_i^{-1}$ is automatically unipotent, it is enough to find when it lies in B . To see this, note that if $j \neq i, i+1$, then $\dot{s}_i u \dot{s}_i(e_j) = \dot{s}_i(u(e_j)) \in F_j$. Similarly $\dot{s}_i u \dot{s}_i(u_{i+1}) = \dot{s}_i u(e_i) \in F_{i+1}$. Finally,

$$\dot{s}_i u \dot{s}_i(e_i) = \dot{s}_i(u(e_{i+1})) \in e_i^*(u(e_{i+1}))e_{i+1} + F_i,$$

and so $\dot{s}_i u \dot{s}_i \in U$ if and only if $u \in U_i$ as claimed. Note that since \dot{s}_i is an involution this also implies that $U_i = \dot{s}_i U_i \dot{s}_i$.

Next note that $\dot{w}X_{ij}\dot{w}^{-1} = X_{w(i)w(j)}$ by the formula for the conjugate of a transvection in Lemma 7.4, and so in particular $\dot{s}_i X_i \dot{s}_i = X_{-i}$. It now follows

that $\dot{s}_i U = \dot{s}_i U_i X_i = (\dot{s}_i U_i \dot{s}_i) \dot{s}_i X_i = U_i \dot{s}_i X_i$. Hence since T is normalized by N , and $B = TU$ we have

$$(4) \quad \dot{s}_i B = \dot{s}_i TU = T \dot{s}_i U = T U_i \dot{s}_i X_i \subset B \dot{s}_i X_i.$$

It follows that if $\dot{w} \in \tilde{W}$, then we have

$$\dot{s}_i B \dot{w} \subset B \dot{s}_i X_i \dot{w} = B \dot{s}_i \dot{w} X_{w^{-1}(i), w^{-1}(i+1)}.$$

Therefore, if $w \in W$ has $w^{-1}(i) < w^{-1}(i+1)$ we have $X_{w^{-1}(i), w^{-1}(i+1)} \subset B$, and so $\dot{s}_i B \dot{w} \subset B \dot{s}_i \dot{w} B$. In other words, if $l(w^{-1} s_i) > l(w^{-1})$ (or, taking inverses, $l(s_i w) > l(w)$, since $l(w) = l(w^{-1})$), then

$$(5) \quad \dot{s}_i B \dot{w} \subset B \dot{s}_i \dot{w} B$$

On the other hand, it is easy to calculate directly that $X_{-i} \subset B \cup B \dot{s}_i B$ – indeed in $\text{GL}_2(k)$ this reduces to the fact that, if $\lambda \neq 0$, then

$$\begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} = \begin{pmatrix} -\lambda^{-1} & 1 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & \lambda^{-1} \\ 0 & 1 \end{pmatrix}.$$

Hence by (4) we have

$$(6) \quad \dot{s}_i B \dot{s}_i \subset B \dot{s}_i X_i \dot{s}_i = B X_{-i} \subset B \cup B \dot{s}_i B,$$

and clearly $X_{-i} \subset \dot{s}_i B \dot{s}_i$, so that $\dot{s}_i B \dot{s}_i$ is not contained in B .

Finally, if $w^{-1}(i) > w^{-1}(i+1)$, then $w' = \dot{s}_i w$ has $w'^{-1} = w^{-1} s_i$ and so $w'^{-1}(i) < w'^{-1}(i+1)$. Thus using (5) and (6) we see that

$$\dot{s}_i B \dot{w} = \dot{s}_i B \dot{s}_i \dot{w}' \subset B \dot{w}' \cup B \dot{s}_i B \dot{w}' \subset B \dot{w}' \cup B \dot{s}_i \dot{w}' B \subset B \dot{s}_i \dot{w} B \cup B \dot{w} B$$

Thus the final BN -pair axiom is established. \square

8. BILINEAR FORMS

We now study bilinear forms on a finite dimensional vector space – our description of classical groups over an arbitrary field showed that they were all the isomorphisms of a vector space which respect a certain bilinear pairing on the space, thus a better understanding of such pairings will be useful in understanding these groups.

Definition 8.1. Let V be a vector space over a field k . A *bilinear form* on V is a function $B: V \times V \rightarrow k$ such that for all $v, v_1, v_2, w, w_1, w_2 \in V, \lambda_1, \lambda_2 \in k$ we have

- (1) $B(\lambda_1 v_1 + \lambda_2 v_2, w) = \lambda_1 B(v_1, w) + \lambda_2 B(v_2, w);$
- (2) $B(v, \lambda_1 w_1 + \lambda_2 w_2) = \lambda_1 B(v, w_1) + \lambda_2 B(v, w_2).$

Thus B is a linear function of each of its variables. We say that B is *symmetric* if $B(v, w) = B(w, v)$, and *skew-symmetric* if $B(v, w) = -B(w, v)$, ($\forall v, w \in V$). Denote the space of bilinear forms by $\text{Bil}(V)$.

Example 8.2. Euclidean space, \mathbb{R}^n equipped with the dot product, is the basic example of a vector space equipped with a bilinear form. However, we will see that this example is quite special due to the fact that the real numbers are an ordered field. Nevertheless, it can be useful to think of B as giving some notion of “distance” in V , and thus allowing us to speak of isometries of V .

Example 8.3. In multivariable calculus, the derivative of a function $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a function $Df: \mathbb{R}^n \rightarrow \text{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R})$, from \mathbb{R}^n to the space of linear maps from \mathbb{R}^n to \mathbb{R} (i.e. at each point in \mathbb{R}^n , the derivative gives a linear map $\mathbb{R}^n \rightarrow \mathbb{R}$). It follows that the second derivative, if it exists, should be a map from

$$D^2(f): \mathbb{R}^n \rightarrow \text{Hom}_{\mathbb{R}}(\mathbb{R}^n, \text{Hom}(\mathbb{R}^n, \mathbb{R})).$$

In other words, for each $x \in \mathbb{R}^n$ we get a linear map $D^2(f)(x)$ from \mathbb{R}^n to $\text{Hom}_{\mathbb{R}}(\mathbb{R}^n, \mathbb{R})$. Thus for each $x \in \mathbb{R}^n$, $D^2(f)(x)$ gives us a function $B_x: \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, from pairs of vectors $v_1, v_2 \in \mathbb{R}^n$ to \mathbb{R} , given by setting

$$B_x(v_1, v_2) = D^2(f)(x)(v_1)(v_2).$$

You can then check that B is a bilinear form on \mathbb{R}^n . Moreover, symmetry of mixed partial derivatives can be neatly expressed in this form as the fact that $B_x(v_1, v_2) = B_x(v_2, v_1)$.

Just as with linear maps, a bilinear form is determined by its action on a basis of V .

Lemma 8.4. Let $\{v_1, v_2, \dots, v_n\}$ be a basis of V , and suppose that $v, w \in V$. If we write $v = \sum_{j=1}^n \lambda_j v_j$ and $w = \sum_{j=1}^n \mu_j v_j$, then

$$B(v, w) = \sum_{1 \leq j_1, j_2 \leq n} \lambda_{j_1} \mu_{j_2} B(v_{j_1}, v_{j_2}).$$

In matrix notation, if we let $D = (B(v_i, v_j))_{i,j=1}^n$, and v, w have coordinates $\lambda, \mu \in k^n$, then $B(v, w) = \lambda^t D \mu$.

Proof. The proof is immediate: We have

$$\begin{aligned} B(v, w) &= B\left(\sum_{j_1=1}^n \lambda_{j_1} v_{j_1}, w\right) = \sum_{j_1=1}^n \lambda_{j_1} B(v_{j_1}, w) \\ &= \sum_{j_1=1}^n \lambda_{j_1} B(v_{j_1}, \sum_{j_2=1}^n \mu_{j_2} v_{j_2}) = \sum_{j_1=1}^n \lambda_{j_1} \sum_{j_2=1}^n \mu_{j_2} B(v_{j_1}, v_{j_2}) \\ &= \sum_{j_1, j_2=1}^n \lambda_{j_1} \mu_{j_2} B(v_{j_1}, v_{j_2}). \end{aligned}$$

as required. \square

We call D the matrix of B with respect to the basis $\{v_1, v_2, \dots, v_n\}$. If A is the change of basis matrix from the basis $\{v_1, v_2, \dots, v_n\}$ to another basis $\{w_1, w_2, \dots, w_n\}$, and D is the matrix of the bilinear form B with respect to $\{v_1, v_2, \dots, v_n\}$, it is easy to see that the matrix of B with respect to the basis $\{w_1, w_2, \dots, w_n\}$ is given by $A^t D A$. We say that two matrices D_1 and D_2 which are related by $D_2 = A^t D_1 A$ for some invertible matrix A are *congruent* (it is easy to check directly that this gives an equivalence relation).

Lemma 8.5. *Suppose that $\text{char}(\mathbf{k}) \neq 2$. Then any bilinear form can be written as the sum of a symmetric and a skew-symmetric form.*

Proof. Let $B_s(v, w) = \frac{1}{2}(B(v, w) + B(w, v))$ and $B_a = \frac{1}{2}(B(v, w) - B(w, v))$, then clearly $B = B_s + B_a$, and B_s is symmetric, while B_a is skew-symmetric. \square

Remark 8.6. We say that a bilinear form B is *alternating* if for all $v \in V$ we have $B(v, v) = 0$. Considering $B(v + w, v + w)$ we see that an alternating form is skew-symmetric, and conversely if $\text{char}(\mathbf{k}) \neq 2$, then a skew-symmetric form is alternating (since $B(v, v) = -B(v, v)$). In characteristic 2 the alternating bilinear forms are the ones we will need.

Formalizing the idea in the above example on the second derivative, we make the following definition. Given $B \in \text{Bil}(V)$ we may define linear maps $L_B: V \rightarrow V^*$ and $R_B: V \rightarrow V^*$, by setting

$$L_B(v)(w) = B(v, w), \quad R_B(v)(w) = B(w, v).$$

The kernel of L_B is called the *left radical* of B , while the kernel of R_B is called the *right radical* of B . A bilinear form is said to be *nondegenerate* if its left and right radicals are zero. (In the case of a symmetric or skew-symmetric form notice that the left and right radical are equal, hence we may refer just to the *radical* of such a form). We may use duality to see that in fact the left radical is zero if and only if the right radical is zero.

Lemma 8.7. *Let $B \in \text{Bil}(V)$. Then $\ker(R_B) = \{0\}$ if and only if $\ker(L_B) = \{0\}$.*

Proof. Recall that $S: V \rightarrow V^{**}$ is the natural isomorphism between V and V^* . Then $L_B: V \rightarrow V^*$ has dual map $L_B^*: V^{**} \rightarrow V^*$. We claim that $R_B = L_B^* \circ S$. To see this, let $v \in V$, then

$$L_B^*(S(v))(w) = S(v)(L_B(w)) = B(w, v) = R_B(v)(w)$$

Since S is an isomorphism, it follows that $\ker(R_B) = \{0\}$ if and only if $\ker(L_B^*) = \{0\}$. To finish the proof, we need only observe that if $\alpha: V \rightarrow W$ is a linear map, and α^* is its dual, then $\ker(\alpha^*) = \text{im}(\alpha)^\circ$. \square

Clearly B is nondegenerate if and only if $\det(D) \neq 0$, where D is the matrix of B with respect to some (and hence any) basis. Since $\det(A^t D A) = \det(A)^2 \det(D)$, we see that while we cannot assign a determinant to B , we can assign a *discriminant* to B . This is simply $\det(D) \in k^\times / (k^\times)^2$ if B is nondegenerate, and 0 otherwise.

Given two vector spaces V_1, V_2 equipped with bilinear forms B_1, B_2 , there is a natural notion of isomorphism $\phi: V_1 \rightarrow V_2$: we require that ϕ is a linear isomorphism such that $B_2(\phi(v), \phi(w)) = B_1(v, w)$. We say that such a ϕ is an *isometry*. Given a vector space with a bilinear form, we may then consider the group of isometries from the vector space to itself. The classical groups we have defined can all be expressed as groups of this form, (rather degenerately in the case of $\text{GL}(V)$, where one can take the bilinear form to be zero).

It is natural therefore to try and classify bilinear forms on a vector space up to isometry. Since any such form can be written as the sum of a symmetric and an alternating form (except in characteristic 2), we may first try and classify these forms. It turns out that the classification of symmetric forms depends on the nature of the field k in a way that the classification of alternating forms does not.

From now on we will restrict ourselves to the case of symmetric and alternating bilinear forms. In this case we see that the maps L_B and R_B are either equal, or $L_B = -R_B$. Recall that a subspace W of V has a subspace of V^* , its annihilator W° , naturally attached to it. When we equip V with a bilinear form, we may use the map L_B or R_B to associate a subspace W^\perp in V to W° , simply by taking the preimage of W° in V under the map L_B : Concretely we have:

$$W^\perp = \{v \in V : B(v, w) = 0, \forall w \in W\}.$$

In the case of $W \subset \mathbb{R}^3$ with the bilinear form given by the dot product, this is the set of vectors perpendicular to W . It follows immediately from our dimension formula for annihilators that, if B is nondegenerate we have

$$\dim(W) + \dim(W^\perp) = \dim(V).$$

In general $(W^\perp)^\perp \supset W$, and hence in the case of a nondegenerate form (using the above dimension formula for W^\perp instead of W) we find that $(W^\perp)^\perp = W$. Given subspaces U_1, U_2 of V we say they are *orthogonal* if $B(u_1, u_2) = 0$ for all $u_1 \in U_1$ and $u_2 \in U_2$ (equivalently, U_1, U_2 are orthogonal if $U_1 \subset U_2^\perp$). We say that a sum of subspaces $\sum_{i \in I} V_i$ is orthogonal if the V_i are pairwise orthogonal.

In the setting of a general (symmetric or skew-symmetric) bilinear form, it is not necessarily the case that $W \cap W^\perp = \{0\}$. Indeed, if B is skew-symmetric, and $v \in V$, then we have

$$B(v, v) = -B(v, v),$$

and so if $\text{char}(k) \neq 2$ then $B(v, v) = 0$ for all $v \in V$. Hence if $L \subset V$ is a one-dimensional vector subspace, then L is always contained in L^\perp .

We now classify all skew-symmetric bilinear forms on a vector space. Consider first the case of a two dimensional vector space H with an alternating form B . Either we have $B = 0$, or there are vectors v_1, v_2 such that $B(v_1, v_2) \neq 0$. Replacing v_2 by $v_2/B(v_1, v_2)$ we may assume that $B(v_1, v_2) = 1$. Since $B(v_1, v_1) = B(v_2, v_2) =$

0 it follows that $\{v_1, v_2\}$ is a basis of H , and B is completely determined by its values on $\{v_1, v_2\}$. Thus we see that there is exactly one nonzero skew-symmetric bilinear form on a 2-dimensional vector space, up to isomorphism. The general case is very similar, as the next proposition shows.

Proposition 8.8. *Let V be a vector space over k , with an alternating bilinear form B . Then there are 2-planes $H_1, H_2, \dots, H_k \subset V$ such that $B|_{H_i}$ is nondegenerate for each i , ($1 \leq i \leq k$) and*

$$V = H_1 \oplus H_2 \oplus \dots \oplus H_k \oplus R(B),$$

an orthogonal direct sum, where $R(B)$ is the radical of B .

Proof. We use induction on $\dim(V)$. When $V = \{0\}$ there is nothing to prove, so suppose $V \neq 0$. If $B = 0$ then $V = R$ and we are done. Otherwise we may find $v_1 \in V$ such that there is a $v_2 \in V$ with $B(v_1, v_2) = 1$ (since B is nondegenerate, there is a $w \in V$ such that $B(v_1, w) \neq 0$, thus we may set $v_2 = w/B(v_1, w)$). Since $B(v_1, v_1) = 0$, we see that $\{v_1, v_2\}$ are linearly independent. Let $H_1 = \text{span}\{v_1, v_2\}$. Then we have $V = V' \oplus H_1$, a direct sum of orthogonal subspaces, where $V' = H_1^\perp = \{v \in V : B(v, v_1) = B(v, v_2) = 0\}$. To see this take $v \in V$, and set

$$v' = v + B(v, v_1)v_2 - B(v, v_2)v_1$$

then v' clearly lies in V' , and so $V = H_1 + V'$. Since $B|_{H_1}$ is nondegenerate, we also have $V' \cap H_1 = \{0\}$, and so the sum is direct as claimed. Now V' has dimension $\dim(V) - 2$, and so by induction we have

$$V' = H_1 \oplus H_2 \oplus \dots \oplus H_k \oplus R(B|_{V'})$$

where $B|_{H_i}$ is nondegenerate, and $R(B|_{V'})$ is the radical of $B|_{V'}$. Since $V = V' \oplus H_1$ is an orthogonal direct sum, and B is nondegenerate on H_1 , it follows that $R(B|_{V'})$ is in fact also the radical of B on all of V , and so we have

$$V = H \oplus H_1 \oplus H_2 \oplus \dots \oplus H_k \oplus R(B).$$

as required. \square

Since we have already observed that nonzero alternating forms on a two dimensional vector space are unique up to isomorphism, it follows there is, up to isomorphism, a unique nondegenerate alternating form on an even dimensional vector space (and none on an odd dimensional space). We call a nondegenerate alternating bilinear form a *symplectic form*, and a vector space with such a form a *symplectic vector space*. A *symplectic basis* of a symplectic vector space V is given by choosing a decomposition of V into 2-planes H_i ($1 \leq i \leq n$) as in the previous proposition, and then picking a basis $\{e_i, f_i\}$ for each 2-plane such that $B(e_i, f_i) = 1$. Thus a symplectic basis is characterized by the properties:

$$B(e_i, f_j) = \delta_{ij}; \quad B(e_i, e_j) = 0; \quad B(f_i, f_j) = 0.$$

A 2-plane $H \subset V$ such that $V = H \oplus H^\perp$ as in the proposition is called a *hyperbolic 2-plane*, and a basis of such a plane $\{e, f\}$ which satisfies $B(e, f) = 1$ is called a *hyperbolic pair*.

Remark 8.9. We end this section with a few remarks about symmetric forms. Here the nature of the field k must enter into the classification of forms up to isometry. If k is algebraically closed, then there is a unique nondegenerate symmetric bilinear form in each dimension – this is proved by finding an “orthonormal basis” for

the form. However if the field is arbitrary the answer is more complicated – for example over the real numbers, the nondegenerate forms on an n -dimensional vector space are classified, up to isometry, by the integer

$$k = \max\{\dim(W) : W \text{ a subspace of } V, B|_W = 0\}.$$

For example, all positive definite symmetric bilinear forms are isometric, and these are the ones for which $k = 0$.

Our classical groups over arbitrary fields were the isometries of a vector space equipped with a bilinear form – symmetric in the case of the orthogonal groups, and alternating in the case of the symplectic groups. We have now seen that in the alternating case, this was the only nondegenerate alternating form (up to isometry), but in the symmetric case, depending on the field we may have other “orthogonal” groups to consider. We will not be able to examine this in the present course.

9. THE SYMPLECTIC GROUP

In this section we analyze the symplectic group in the spirit of our analysis of the general linear group. Let V be a vector space over k equipped with a nondegenerate alternating bilinear form, $B: V \times V \rightarrow k$. Thus the dimension of V is $2n$ for some integer n . The *symplectic group* is the group of linear isomorphisms preserving the form B :

$$\mathrm{Sp}(V) = \{\alpha \in \mathrm{GL}(V) : B(\alpha(v), \alpha(w)) = B(v, w), \text{ for all } v, w \in V\}.$$

Picking a symplectic basis $\{e_i, f_i : 1 \leq i \leq n\}$, we may identify B with the matrix

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

and then $\mathrm{Sp}(V)$ becomes identified with the matrix group

$$\mathrm{Sp}_{2n}(k) = \{A \in \mathrm{GL}_{2n}(k) : A^t J A = J\}$$

The existence of symplectic bases also allows us to compute the order of $\mathrm{Sp}(V)$ over a finite field. Indeed the set of symplectic bases of V is a set on which $\mathrm{Sp}(V)$ clearly acts transitively, and the stabilizer of an element is trivial (such an action is called a *free* action). Hence the set of symplectic bases and the elements of $\mathrm{Sp}(V)$ are in bijection. When $k = \mathbb{F}_q$ a finite field, we may count the number of symplectic bases as follows: To construct a symplectic basis one first chooses a hyperbolic pair in V . There are clearly

$$(q^{2n} - 1)(q^{2n} - q^{2n-1})/(q - 1) = q^{2n-1}(q^{2n} - 1)$$

such pairs. To extend this pair to a symplectic basis, we must choose another hyperbolic pair in the subspace orthogonal to the hyperbolic pair, and continue until we have exhausted V . Thus we see that

$$|\mathrm{Sp}(V)| = q^{(2n-1)+(2n-3)+\dots+1} \prod_{i=1}^n (q^{2i} - 1) = q^{n^2} \prod_{i=1}^n (q^{2i} - 1).$$

Notice that in the case of $\mathrm{GL}(V)$ (and also in the case of the orthogonal group) one can obtain a natural normal subgroup using the determinant. It turns out however, that in the case of the symplectic group, $\mathrm{Sp}(V)$ is a subgroup of $\mathrm{SL}(V)$. The simplest case of this is the following.

Lemma 9.1. *Let $\dim(V) = 2$. Then $\mathrm{Sp}(V) = \mathrm{SL}(V)$.*

Proof. This can easily be checked if you pick a basis of V and use matrices. However we give also a coordinate free proof. Recall that for any vector space V there is a unique (up to scaling) alternating multilinear function

$$D: V \times V \times \dots \times V \rightarrow k$$

(where there are $n = \dim(V)$ copies of V), and the determinant of a matrix is the value of D on the columns of this matrix. More naturally, given any linear map α we can form a new alternating multilinear map $\alpha^*(D)$ from D and α by setting

$$\alpha^*(D)(v_1, v_2, \dots, v_n) = D(\alpha(v_1), \alpha(v_2), \dots, \alpha(v_n)).$$

Since D is unique up to scalar, we must have $\alpha^*(D) = \lambda D$ for some $\lambda \in k$. Then $\det(\alpha)$ is exactly this scalar. The lemma follows by observing that in the case $\dim(V) = 2$, the form B must be a nonzero multiple of D . \square

Our first strategy is to find a set of elements of $\mathrm{Sp}(V)$ which will play the role of transvections. Conveniently we may simply use the transvections which lie in $\mathrm{Sp}(V)$!

Lemma 9.2. *A transvection in $\mathrm{GL}(V)$ lies in $\mathrm{Sp}(V)$ if and only if it has the form $\tau_{u,\phi}$ where $\phi(v) = aB(v, u)$ for some $a \in \mathbf{k}$.*

Proof. We need to calculate when $\tau_{u,\phi}$ preserves B . But

$$\begin{aligned} B(\tau_{u,\phi}(v), \tau_{u,\phi}(w)) &= B(v + \phi(v)u, w + \phi(w)u) \\ &= B(v, w) + \phi(w)B(v, u) + \phi(v)B(u, w). \end{aligned}$$

Therefore $\tau_{u,\phi} \in \mathrm{Sp}(V)$ if and only if

$$\phi(w)B(v, u) = \phi(v)B(w, u), \quad \forall v, w \in V.$$

Now pick $w \in V$ such that $B(u, w) = 1$. Then we find that $\phi(v) = \phi(w)B(v, u)$, and so setting $a = \phi(w)$ we are done. \square

We use the notation $\tau_{u,a}$ for the transvection $v \mapsto v + aB(v, u)u$. Let \mathcal{T} denote the subgroup of $\mathrm{Sp}(V)$ generated by transvections. Our goal, of course, is to show that in fact this group is all of $\mathrm{Sp}(V)$.

Lemma 9.3. *The group \mathcal{T} acts transitively on $V - \{0\}$. Indeed \mathcal{T} acts transitively on the set of hyperbolic pairs.*

Proof. Suppose that $v, w \in V$. If $B(v, w) \neq 0$, then set $u = w - v$ and $a = B(v, w)^{-1}$ so that

$$\tau_{u,a}(v) = v + aB(v, u)u = v + B(v, w)^{-1}B(v, w)(w - v) = w.$$

If $B(v, w) = 0$, then we may find $\phi \in V^*$ such that $\phi(v)$ and $\phi(w) \neq 0$. Then setting $u = R_B^{-1}(\phi)$ we have $B(v, u) \neq 0$ and $B(w, u) \neq 0$. Then by what has already been established, there are symplectic transvections τ_1, τ_2 taking v to u and w to u respectively. Then $\tau_2^{-1} \circ \tau_1$ takes v to w as required.

For the second part, suppose that $\{e_1, f_1\}$ and $\{e_2, f_2\}$ are hyperbolic pairs. Using the first part, we may assume that $e_1 = e_2 = e$ say, and so we need only find an element of \mathcal{T} fixing e taking f_1 to f_2 . Now $B(f_1, f_2) \neq 0$ we may take $u = f_2 - f_1$ and $a = B(f_1, f_2)^{-1}$ so that

$$\tau_{u,a}(f_1) = f_1 + aB(f_1, u)u = f_2$$

and $\tau_{u,a}(e) = e$ since $B(e, f_1) = B(e, f_2) = 0$. If $B(f_1, f_2) = 0$, then consider the pair $(e, e + f_1)$. It is also a hyperbolic pair, and $B(f_1, e + f_1) = -1$, so that by the above there is a symplectic transvection τ_3 taking the pair (e, f_1) to the pair $(e, e + f_1)$. Similarly $B(f_2, e + f_1) = -1$ so that there is a transvection τ_4 taking $(e, e + f_1)$ to (e, f_2) . Then $\tau_4 \circ \tau_3$ takes (e, f_1) to (e, f_2) . \square

This is enough for us to be able to show that in fact the symplectic transvections generate $\mathrm{Sp}(V)$.

Lemma 9.4. *The group \mathcal{T} coincides with $\mathrm{Sp}(V)$.*

Proof. We show this by induction on $\dim(V)$. If $\dim(V) = 2$, then $\mathrm{SL}(V) = \mathrm{Sp}(V)$, and since we know transvections generate $\mathrm{SL}(V)$, we are done. Now suppose that $\dim(V) > 2$. Let $g \in \mathrm{Sp}(V)$, and pick a symplectic basis $\{e_i, f_i : 1 \leq i \leq n\}$ for V . Thus $\{g(e_1), g(f_1)\}$ are a hyperbolic pair. By the previous lemma we know

that there is an element h of \mathcal{T} such that $h(e_1) = g(e_1)$ and $h(f_1) = g(f_1)$. Thus $g' = h^{-1}g$ fixes the hyperbolic pair $\{e_1, f_1\}$. Thus g' fixes the subspace orthogonal to $H_1 = \text{span}\{e_1, f_1\}$, that is the span of the vectors $\{e_i, f_i : i \geq 2\}$, and is the identity on H_1 . But then $B|_{H_1^\perp}$ is a nondegenerate alternating form, and $g'|_{H_1^\perp}$ preserves B , so by induction, $g'|_{H_1^\perp}$ is a product of symplectic transvections. Extending these by letting them act trivially on H_1 , we see that g' is a product of symplectic transvections, and hence g is also. Thus $\text{Sp}(V) = \mathcal{T}$ as required. \square

Corollary 9.5. *The group $\text{Sp}(V)$ is a subgroup of $\text{SL}(V)$.*

Proof. This follows immediately from the fact that any symplectic transvection lies in $\text{SL}(V)$. \square

Remark 9.6. Morally, this is the “wrong” proof. A better proof is given in the exercises. The corollary is not, however, completely obvious – if you translate the statement into one about matrices, it says that a matrix A satisfying $A^t J A = J$ must have $\det(A) = 1$. It is easy to see that $\det(A) = \pm 1$ from the multiplicativity of the determinant and the fact that $\det(A^t) = \det(A)$, but it is not clear from this why A must have determinant 1.

Lemma 9.7. *The center of $\text{Sp}(V)$ is precisely the group $\{\pm 1\}$.*

Proof. Notice that if $g \in \text{Sp}(V)$, then $g\tau_{u,a}g^{-1}(v) = \tau_{g(u),a}$. It follows that if $g \in Z(\text{Sp}(V))$ we must have $g(u) = \lambda u$, where $B(v, \lambda u)u = B(v, u)u$ for all $v \in V$, which holds only if $\lambda^2 = 1$, that is, if and only if $\lambda = \pm 1$. \square

We set $\text{PSp}(V) = \text{Sp}(V)/\{\pm 1\}$. Our goal is to show that this group is simple in almost all cases, just as $\text{PSL}(V)$ was simple in almost all cases.

Lemma 9.8. *Suppose that $|k| > 3$, then the group $\text{Sp}(V)$ is its own derived subgroup.*

Proof. The $n = 2$ case follows from the case of $\text{SL}(V)$, so we may assume that $\dim(V) \geq 4$. As for $\text{SL}(V)$, we need only express a symplectic transvection as a commutator, since the derived group will then contain all the conjugates of that transvection, that is, all symplectic transvections. Now fix $u \in V - \{0\}$ and $a \in k^\times$. Since $\text{Sp}(V)$ is transitive on $V - \{0\}$, we may pick $g \in \text{Sp}(V)$ such that $g(u) = \lambda u$, for any $\lambda \in k^\times$. Then $g\tau_{u,a}g^{-1} = \tau_{\lambda u,a} = \tau_{u,\lambda^2 a}$. Thus

$$[g, \tau_{u,a}] = \tau_{u,(\lambda^2-1)a}.$$

Then the equation $b = (\lambda^2 - 1)a$ can be solved for any $b \in k^\times$ whenever $|k| > 3$, and hence the derived group of $\text{Sp}(V)$ contains all symplectic transvections are required. \square

It remains to understand the cases when $|k| = 2$ or 3 . For $\dim(V) = 2$, we already know $\text{Sp}(V)$ is not its own derived group in these cases. For $\dim(V) \geq 4$ however, one can show that $\text{Sp}(V)$ is its own derived group if $|k| = 3$, and if $\dim(V) \geq 6$ then $\text{Sp}(V)$ is its own derived group for all k . To see these claims requires specific computations to produce a transvection as a commutator. The conclusion is that $\text{Sp}(V)$ is its own derived group unless $\dim(V) = 2$ and $k = \mathbb{F}_2$ or \mathbb{F}_3 , or $\dim(V) = 4$ and $k = \mathbb{F}_2$. This last case is in fact an exception, as we will see later.

In order to establish the simplicity of the group $\text{PSp}(V)$ we follow the strategy we used for $\text{PSL}(V)$. We already know that $\text{Sp}(V)$ acts transitively on $V - \{0\}$, so

that clearly $\mathrm{PSp}(V)$ acts faithfully and transitively on $\mathbb{P}(V)$. If P is the stabilizer in $\mathrm{PSp}(V)$ of a point $p \in \mathbb{P}(V)$, then we would like to know that for any normal subgroup $K \triangleleft \mathrm{PSp}(V)$ we have either $K \subset P$ or $KP = \mathrm{PSp}(V)$. For the projective special linear group, we used the fact that $\mathrm{PSL}(V)$ acts 2-transitively on $\mathbb{P}(V)$ to conclude that the corresponding stabilizer is a maximal subgroup of $\mathrm{PSL}(V)$. The action of $\mathrm{PSp}(V)$ is unfortunately not 2-transitive (*why?*) so we need some more subtle argument.

There are a number of possibilities – the stabilizer P is in fact a maximal subgroup, as can be seen by developing the machinery of BN -pairs for $\mathrm{Sp}(V)$. We choose a shorter root however, for which we need some definitions. The idea is to find a weaker notion than 2-transitivity which will suffice to show that $KP = \mathrm{PSp}(V)$ in the case where $K \triangleleft \mathrm{PSp}(V)$ is not contained in $P = \mathrm{Stab}_{\mathrm{PSp}(V)}(p)$.

Definition 9.9. Let X be a transitive G -set. A *block* in X is a proper subset $B \subset X$ such that $|B| \geq 2$ and for each $g \in G$ either $g.B = B$ or $g.B \cap B = \emptyset$. If X has no blocks we say the G -action is *primitive*.

Remark 9.10. It is not hard to check that if the action of G on X is 2-transitive, then it is primitive.

Lemma 9.11. *Suppose that G acts primitively on X , and that $K \triangleleft G$ is a normal subgroup not contained in the kernel of the action of G on X . Then K acts transitively on X . Moreover, if $a \in X$, then $G = K.\mathrm{Stab}_G(a)$.*

Proof. We show the contrapositive. Suppose that K does not act transitively. Then let B be an K -orbit in X of size greater than 1 (such an orbit exists by the assumption that K is not in the kernel of the action). We claim that B is a block. Indeed suppose that $g \in G$. Then if $x \in B$, we have $B = K.x$, and so $g.B = g.(K.x) = (gKg^{-1}).g.x$. Since K is normal this shows that $g.B = K.(gx)$, and hence $g.B$ is an K -orbit in X . Since distinct orbits are disjoint, it follows that B is a block, and hence the action of G is not primitive. For the moreover part, fix $a \in X$, and let $g \in G$. Then there is an $k \in K$ such that $k.a = g.a$. It follows that $g = k(k^{-1}g) \in K.\mathrm{Stab}_G(a)$ as required. \square

We are now able to prove the simplicity of $\mathrm{PSp}(V)$ in the cases where it is equal to its own derived group. What we need to do is to show that $\mathrm{PSp}(V)$ acts primitively on $\mathbb{P}(V)$.

Lemma 9.12. *$\mathrm{PSp}(V)$ acts primitively on $\mathbb{P}(V)$.*

Proof. Let B be a block for the action of $\mathrm{PSp}(V)$, and suppose that $[v], [w]$ are two distinct elements of B . Suppose that $B(v, w) = 0$. Then pick $u \in V$ such that $B(u, v) \neq 0$ and $B(u, w) \neq 0$ (see the proof of Lemma 9.3). Then $\tau_{v,1}(u) = u + B(u, v)v$ and $\tau_{v,1}(w) = w$, hence since B is a block, $[u + B(u, v)v] \in B$. Thus setting $u_1 = u + B(u, v)v$, we have $[v], [u_1] \in B$ such that $B(v, u_1) = B(v, u) \neq 0$. We may rescale so that $\{v, u_1\}$ are a hyperbolic pair, and since $\mathrm{Sp}(V)$ acts transitively on such pairs, it follows that B contains all $[p] \in \mathbb{P}(V)$ such that $B(v, p) \neq 0$. Now if $u \in V$ has $B(v, u) = 0$, as before we may find a u' such that $B(u, u') \neq 0$ and $B(v, u) \neq 0$. Then we have $u' \in B$ and hence (using what we have done for u' instead of v) it follows $u \in B$. Thus $B = \mathbb{P}(V)$ as required. \square

Theorem 9.13. *Let V be a symplectic vector space over k such that either $\dim(V) \geq 6$, or $\dim(V) = 4$ and $|k| > 2$, or $k > 3$ and $\dim(V) = 2$. Then $\mathrm{PSp}(V)$ is a simple group.*

Proof. Let $K \triangleleft \mathrm{PSp}(V)$ be a normal subgroup of $\mathrm{PSp}(V)$. Then pick $[v] \in \mathbb{P}(V)$, and let $P = \mathrm{Stab}_{\mathrm{PSp}(V)}([v])$. Then if $K \subset P$, it follows that K fixes $[v]$. Since K is normal and the action of $\mathrm{PSp}(V)$ is transitive on $\mathbb{P}(V)$ it follows that K fixes $\mathbb{P}(V)$ pointwise. Since the action of $\mathrm{PSp}(V)$ is faithful, this implies that $K = \{1\}$.

On the other hand, if $K \not\subset P$ then since K is normal and the action of $\mathrm{PSp}(V)$ on $\mathbb{P}(V)$ is primitive, we have $\mathrm{PSp}(V) = KP$. But then if $\pi: \mathrm{PSp}(V) \rightarrow \mathrm{PSp}(V)/K$ is the quotient map, we have $\pi(P) = \pi(\mathrm{PSp}(V))$. But if we let $N = \{\tau_{v,a} : a \in k\}$, then $N \triangleleft P$, and by the proof of Lemma 9.7 the union of the conjugates of N , being (the image in $\mathrm{PSp}(V)$ of) the set of transvections in $\mathrm{Sp}(V)$ generate $\mathrm{PSp}(V)$. It follows that $\pi(P) = \pi(N)$, and so $KN = \mathrm{PSp}(V)$. But N is abelian, so taking the derived subgroups of both sides (and using the assumptions of the theorem so that $\mathrm{PSp}(V)$ is its own derived subgroup) we find $K = \mathrm{PSp}(V)$ as claimed. \square

The cases excluded by the hypotheses of the theorem are in fact not simple: When $\dim(V) = 2$ these are $\mathrm{PSL}_2(\mathbb{F}_2)$ and $\mathrm{PSL}_2(\mathbb{F}_3)$, which have already established are S_3 and A_4 respectively. The only group that remains to examine is $\mathrm{Sp}_4(\mathbb{F}_2)$. It turns out that this group is just S_6 , the symmetric group on 6 letters.

To see this we show a more general result.

Lemma 9.14. *Let X be a set and let $\mathcal{P}(X)$ be the power set of X , i.e. the set of subsets of X . For $P, Q \in \mathcal{P}(X)$ we define*

$$P + Q = (P \cup Q) - (P \cap Q), \quad P \cdot Q = P \cap Q.$$

Under these operation, $\mathcal{P}(X)$ is a ring, and indeed an \mathbb{F}_2 -algebra. Moreover if we set

$$B(P, Q) = |P \cdot Q| \pmod{2},$$

then B is a nondegenerate symmetric form on $\mathcal{P}(X)$.

Proof. We must check the axioms for a ring. To see that $\mathcal{P}(X)$ is an abelian group, note that \emptyset is an identity for $+$, and $P + P = \emptyset$, so every element has an inverse. Clearly, $+$ is commutative, and for associativity of $+$ one should draw a Venn diagram. It follows that $\mathcal{P}(X)$ is an abelian group and so a \mathbb{Z} -module. Since $2P = 0$ for each $P \in \mathcal{P}(X)$, it is in fact a $\mathbb{Z}/2\mathbb{Z}$ -module, i.e. an \mathbb{F}_2 -vector space. The multiplication is clearly commutative and associative, and it is easy to check that it distributes over addition. This also readily establishes that B is bilinear. That B is nondegenerate is also immediate. \square

Remark 9.15. Another way to think of this lemma is to order the elements of X and assign a binary string to each subset – with a 1 if the element of X is in the subset and a 0 if it isn't. This gives a bijection between $\mathcal{P}(X)$ and \mathbb{F}_2^n , and the vector space structures match up. The bilinear form is just the "dot product" in this setting.

Now suppose that X is a set with $2m$ elements. Let L be the span of the set X itself, so that $L \subset L^\perp$, which is just the set \mathcal{E} of subsets of X with an even number of elements. Then the space L^\perp/L inherits a nondegenerate form, which is alternating, since $B(P, P) = 0$ for any set with an even number of elements. Since it is clearly nondegenerate, this gives a realization of the symplectic vector space of dimension $2m - 2$ over \mathbb{F}_2 . The symmetric group S_{2m} acts on X , and this induces an action of S_{2m} on $\mathcal{P}(X)$ as an \mathbb{F}_2 -algebra, hence in particular the

action preserves the form B . It follows that S_{2m} acts by symplectic isometries on \mathcal{E}/L , and so we obtain an embedding $S_{2m} \rightarrow \mathrm{Sp}(\mathcal{E})$. In the case when $m = 6$, it is easy to compute the orders of these groups and hence obtain the isomorphism $\mathrm{Sp}_4(\mathbb{F}_2) \cong S_6$ as required.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CHICAGO.