CHAPTER 3

Groups of matrices: Linear algebra and symmetry in various geometries

Lecture 14

a. Orthogonal matrices and isometries of $\mathbb{R}^n$. Using the standard scalar product on $\mathbb{R}^n$, let I be an isometry of $\mathbb{R}^n$ which fixes $\mathbf{0}$; thus I is a linear map which preserves the standard scalar product. In particular, if $\mathcal{E} = \{\mathbf{e}^1, \dots, \mathbf{e}^n\}$ is an orthonormal basis, then the set $I\mathcal{E} = \{I\mathbf{e}^1, \dots, I\mathbf{e}^n\}$ is still an orthonormal basis, since $\left\langle I\mathbf{e}^i, I\mathbf{e}^j \right\rangle = \left\langle \mathbf{e}^i, \mathbf{e}^j \right\rangle$.

Let B be the $n \times n$ matrix which represents the linear transformation I in the basis $\mathcal{E}$—that is, $b_{ij} = \left\langle I\mathbf{e}^i, \mathbf{e}^j \right\rangle$, so

$$I\mathbf{e}^i = \sum_{j=1}^{n} b_{ij}\mathbf{e}^j.$$

Then the statement that $I\mathcal{E}$ is an orthonormal basis is equivalent to the statement that the row vectors of B are orthonormal, because in this case

$$\left\langle I\mathbf{e}^i, I\mathbf{e}^j \right\rangle = \sum_{k=1}^{n} b_{ik} b_{jk} = 0$$

for $i \neq j$, and

$$\left\langle I\mathbf{e}^i, I\mathbf{e}^i \right\rangle = \sum_{j=1}^{n} b_{ij}^2 = 1.$$

Recall from the rules for matrix multiplication that this is equivalent to the condition $BB^T = \mathrm{Id}$, or $B^T = B^{-1}$. This is in turn equivalent to $B^T B = \mathrm{Id}$, which is the statement that the column vectors of B are orthonormal.

Alternately, one may observe that if we let $\mathbf{x}$ denote a column vector and $\mathbf{x}^T$ a row vector, then the standard form of the scalar product (13.2) becomes $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y}$, and so we have the following general relationship:

$$\langle \mathbf{x}, A\mathbf{y} \rangle = \mathbf{x}^T A \mathbf{y} = (A^T \mathbf{x})^T \mathbf{y} = \left\langle A^T \mathbf{x}, \mathbf{y} \right\rangle. \tag{14.1}$$

Thus if B is the matrix of I, which preserves scalar products, we have

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle B\mathbf{x}, B\mathbf{y} \rangle = \left\langle B^T B \mathbf{x}, \mathbf{y} \right\rangle$$

for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, which implies $B^T B = \mathrm{Id}$, as above.

Definition 14.1. A matrix B such that $B^T = B^{-1}$ is called *orthogonal*. The group of orthogonal $n \times n$ matrices is denoted $O(n)$.

It should come as no surprise by now that the group of orthogonal matrices is identified with the group of isometries which fix the origin. Furthermore, since $\det B^T = \det B$, we see that any orthogonal matrix has

$$1 = \det \mathrm{Id} = \det(B^T B) = (\det B^T)(\det B) = (\det B)^2,$$

and hence $\det B = \pm 1$. Matrices with determinant 1 correspond to even isometries fixing the origin and compose the *special orthogonal group* that is denoted by $SO(n)$; matrices with determinant -1 correspond to odd isometries fixing the origin.

In $SO(3)$, we saw that the conjugacy class of a rotation contained all rotations through the same angle θ. Later in this lecture, we will sketch a proof of the analogous result in higher dimensions, which states that given any $B \in O(n)$, there exists $A \in O(n)$ such that the matrix $A^{-1}BA$ has the form

$$\begin{pmatrix} 1 &&&&&&&& \\ & \ddots &&&&&&& \\ && 1 &&&&&& \\ &&& -1 &&&&& \\ &&&& \ddots &&&& \\ &&&&& -1 &&& \\ &&&&&& R_{\theta_1} && \\ &&&&&&& \ddots & \\ &&&&&&&& R_{\theta_k} \end{pmatrix}, \tag{14.2}$$

where all entries not shown are taken to be 0, and where R_θ is the 2×2 rotation matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

and the θ_i may be the same or may be different but not equal to 0 or π.

One can also combine pairs of 1's into rotations by angle 0 and pairs of -1's into rotations by π. Then no more than one "loose" diagonal element 1 and -1 is left. In the even dimension n for an $SO(n)$ matrix no loose elements remain and for a matrix with determinant -1 there is one 1 and one -1. in the odd dimension exactly one loose element remains and it is 1 or -1 according to the sign of the determinant.

Geometrically, this means that $\mathbb{R}^n$ can be decomposed into the orthogonal direct sum of a number of one-dimensional subspaces X_i which are fixed by B, a number of one-dimensional subspaces Y_i on which B acts as the map $x \mapsto -x$ (that is, a reflection), and a number of two-dimensional subspaces Z_i on which B acts as a rotation by θ_i. Since the rotation is a product of two reflections this also gives a representation of isometry as the product of at most n reflections. To summarize:

The isometry determined by the matrix B can be written as the product of commuting reflections in the orthogonal complements of Y_i (reflection is *always around an affine subspace of codimension one) together with commuting rotations in the orthogonal complements of Z_i (rotation is always around an affine subspace of codimension two). The number of subspaces Y_i—that is, the number of times -1 occurs on the diagonal—determines whether the isometry given by B is even or odd.*

Notice that any isometry of $\mathbb{R}^n$ with a fixed point is conjugate to an isometry fixing the origin by a translation. Thus linear algebra gives us a complete description of isometries of $\mathbb{R}^n$ with a fixed point.

The three dimensional case is particularly easy then: there is one rotation block (possibly the identity) and either 1 on the diagonal (resulting in a rotation or the identity map) or -1 (resulting in a rotatory reflection or a pure reflection).

EXERCISE 14.1. Using the representation (14.2) give a complete classification of isometries in $\mathbb{R}^4$ with a fixed point and describe conjugacy classes in $SO(4)$ and $O(4)$.

b. Eigenvalues, eigenvectors, and diagonalizable matrices. We stated above that every orthogonal matrix $A \in O(n)$ can be put in the form (14.2) by a suitable change of coordinates—that is, a transformation of the form $A \mapsto CAC^{-1}$, where $C \in O(n)$ is the change of basis matrix. This is related to perhaps the most important result in linear algebra, *Jordan normal form*. Now we will review the relevant concepts from linear algebra and show why every orthogonal transformation can be so represented. Along the way we will learn importance of *complexification*, when objects defined over the field of real numbers (in our case, linear spaces, linear transformations and scalar products) are extended to the complex field.

Before diving into the details, we observe that our mission can be described both geometrically and algebraically. Geometrically, the story is this: we are given a linear transformation $L\colon \mathbb{R}^n \to \mathbb{R}^n$, and we wish to find a basis in which the matrix of L takes on as simple a form as possible. In algebraic terms, we are given a matrix $L \in GL(n, \mathbb{R})$, and we wish to describe the conjugacy class of L—that is, we want to characterise all matrices L' such that $L' = CLC^{-1}$ for some $C \in GL(n, \mathbb{R})$.[1] Ideally, we would like to select a good representative from each conjugacy class, which will be the *normal form* of L.

DEFINITION 14.2. Let L be an $n \times n$ matrix with real entries. An *eigenvalue* of L is a number λ such that

$$L\mathbf{v} = \lambda\mathbf{v} \tag{14.3}$$

[1]If $L \in O(n)$, then we would like to take the conjugating matrix C to be orthogonal as well. In this case there is no difference between conjugacy in the group $GL(n, \mathbb{R})$ and conjugacy in the subgroup $O(n)$, but this is not always the case; recall that rotations $R_\theta^{\mathbf{x}}$ and $R_{-\theta}^{\mathbf{x}}$ are conjugate in $\mathrm{Isom}(\mathbb{R}^2)$, but not in $\mathrm{Isom}^+(\mathbb{R}^2)$.

for some vector $\mathbf{v} \in \mathbb{R}^n$, called an *eigenvector* of L. The set of all eigenvectors of λ is a subspace of $\mathbb{R}^n$, called the *eigenspace* of λ. The *multiplicity* of λ is the dimension of this subspace.

Although this definition only allows real eigenvalues, we will soon see that complex eigenvalues can also exist, and are quite important.

Exercise 14.2. Let $\mathbf{v}_1, \dots, \mathbf{v}_k$ be eigenvectors of L, and let $\lambda_1, \dots, \lambda_k$ be the corresponding eigenvalues. Suppose that $\lambda_i \neq \lambda_j$ for $i \neq j$, and show that the eigenvectors $\mathbf{v}_i$ are linearly independent.

It follows from Exercise 14.2 that there are only finitely many eigenvalues for any matrix. But why should we be interested in eigenvalues and eigenvectors? What purpose does (14.3) serve?

One important (algebraic) reason is that the set of eigenvalues of a matrix is invariant under conjugacy.

An important geometric reason is that (14.3) shows that on the subspace containing $\mathbf{v}$, the action of the linear map $L\colon \mathbb{R}^n \to \mathbb{R}^n$ is particularly simple—multiplication by λ! If we can decompose $\mathbb{R}^n$ into a direct product of such subspaces, then we can legitimately claim to have understood the action of L.

Definition 14.3. L is *diagonalizable* (over $\mathbb{R}$) if there exists a basis $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$ such that each $\mathbf{v}_i$ is an eigenvector of L.

Suppose $\{\mathbf{v}_i\}$ is a basis of eigenvectors with eigenvalues $\{\lambda_i\}$, and let $C \in GL(n, \mathbb{R})$ be the linear map such that $C\mathbf{v}_i = \mathbf{e}_i$ for each $1 \leq i \leq n$. Observe that

$$CLC^{-1}\mathbf{e}_i = CL\mathbf{v}_i = C(\lambda_i \mathbf{v}_i) = \lambda_i \mathbf{e}_i;$$

hence the matrix of CLC^{-1} is

$$\operatorname{diag}(\lambda_1, \dots, \lambda_n) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}. \tag{14.4}$$

It follows from Exercise 14.2 that L has no more than n eigenvalues. So far, though, nothing we have said prevents it from having fewer than n eigenvalues, even if we count each eigenvalue according to its multiplicity. Indeed, one immediately sees that any rotation of the plane by an angle not equal to 0 or π is a linear map with no real eigenvalues. Thus we cannot expect to diagonalise every matrix, and must look to more general forms for our classification.

The eigenvalue equation (14.3) characterises eigenvectors (and hence eigenvalues) geometrically: $\mathbf{v}$ is an eigenvector if and only if it is parallel to its image $L\mathbf{v}$. An algebraic description of eigenvalues can be obtained by recalling that given an $n \times n$ matrix A, the existence of a vector $\mathbf{v}$ such that $A\mathbf{v} = \mathbf{0}$ is equivalent to the condition that $\det A = 0$. We can rewrite (14.3) as $(L - \lambda\,\mathrm{Id})\mathbf{v} = \mathbf{0}$, and so we see that λ is an eigenvalue of L if and only if $\det(L - \lambda\,\mathrm{Id}) = 0$.

The determinant of an $n \times n$ matrix is the sum of $n!$ terms, each of which is a product of n entries of the matrix, one from each row and column. It follows that $p(\lambda) = \det(L - \lambda\,\mathrm{Id})$ is a polynomial of degree n, called the *characteristic polynomial* of the matrix L, and that the coefficients of p are polynomial expressions in the entries of the matrix. Incidentally, we see from this that indeed there are no more than n eigenvalues, real or complex.

The upshot of all this is that the eigenvalues of a matrix are the roots of its characteristic polynomial, and now we see the price we pay for working with the real numbers—$\mathbb{R}$ is not algebraically closed, and hence the characteristic polynomial may not factor completely over $\mathbb{R}$! Indeed, it may not have any roots at all; for example the characteristic polynomial of the rotation matrix $\left(\begin{smallmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{smallmatrix}\right)$ is $p(\lambda) = \lambda^2 - 2\cos\theta + 1$.

We can resolve this difficulty and ensure that L has "enough eigenvalues" by passing to the complex numbers, over which every polynomial factors completely, and declaring any *complex* root of $p(\lambda) = 0$ to be an eigenvalue of L. Then the Fundamental Theorem of Algebra gives us

$$p(\lambda) = \det(L - \lambda\,\mathrm{Id}) = \prod_{i=1}^{n} (\lambda - \lambda_i), \tag{14.5}$$

where $\{\lambda_1, \dots, \lambda_n\} \subset \mathbb{C}$ are the eigenvalues of L.

The set of all eigenvalues of L is called the *spectrum* of L.

EXERCISE 14.3. Given an $n \times n$ matrix L and a change of coordinates $C \in GL(n, \mathbb{R})$, show that L and $L' = CLC^{-1}$ have the same spectrum, and that C takes eigenvectors of L into eigenvectors of L'.

At this point, it is not at all clear what geometric significance a complex eigenvalue has, if any. After all, if $\lambda \in \mathbb{C} \setminus \mathbb{R}$ is an eigenvalue of L and $\mathbf{v}$ is a vector in $\mathbb{R}^n$, what does the expression $\lambda\mathbf{v}$ even *mean*?

c. Complexification, complex eigenvectors and rotations. The difficulty in interpreting the expression $\lambda\mathbf{v}$ for $\lambda \in \mathbb{C}$ and $\mathbf{v} \in \mathbb{R}^n$ is that vectors in $\mathbb{R}^n$ must have real coordinates. We can solve this problem in a rather simple-minded way—just let the coordinates be complex! If we consider vectors $\mathbf{v} \in \mathbb{C}^n$, the n-dimensional *complex* vector space, then $\lambda\mathbf{v}$ makes perfect sense for any $\lambda \in \mathbb{C}$; thus (14.3) may still be used as the definition of an eigenvalue and eigenvector, and agrees with the definition in terms of the characteristic polynomial.

The same procedure can be put more formally: $\mathbb{C}^n$ is the *complexification* of the real vector space $\mathbb{R}^n$, and is equal as a real vector space to the direct sum of two copies of $\mathbb{R}^n$. We call these two copies V_R and V_I (for real and imaginary); given vectors $\mathbf{x} \in V_R$ and $\mathbf{y} \in V_I$, we intertwine the coordinates and write

$$\mathbf{z} = (x_1, y_1, x_2, y_2, \dots, x_n, y_n) \in \mathbb{R}^{2n} \tag{14.6}$$

for the vector with real part $\mathbf{x}$ and imaginary part $\mathbf{y}$. As a vector with n complex coordinates, we write $\mathbf{z}$ as

$$\mathbf{z} = (x_1 + iy_1, x_2 + iy_2, \dots, x_n + iy_n). \tag{14.7}$$

In order to go from the formulation (14.6) to the complex vector space (14.7), we must observe that multiplication by i acts on $\mathbb{R}^{2n}$ as the linear operator

$$J\colon (x_1, y_1, \dots, x_n, y_n) \mapsto (-y_1, x_1, \dots, -y_n, x_n).$$

That is, if we decompose $\mathbb{R}^{2n}$ as the direct sum of n copies of $\mathbb{R}^2$, the action of J rotates each copy of $\mathbb{R}^2$ by $\pi/2$ counterclockwise, which is exactly the effect multiplication by i has on the complex plane.[2]

Having defined $\mathbb{C}^n$, we observe that since L and J commute, L extends uniquely to a linear operator $L^{\mathbb{C}}\colon \mathbb{C}^n \to \mathbb{C}^n$. All the definitions from the previous section go through for $L^{\mathbb{C}}$, and now the fundamental theorem of algebra guarantees that (14.5) holds and the characteristic polynomial factors completely over $\mathbb{C}$. We refer to any eigenvalue of $L^{\mathbb{C}}$ as an eigenvalue of L itself, and this justifies our definition of spectrum of L as a subset of $\mathbb{C}$. But now we must ask: What do the (complex-valued) eigenvalues and eigenvectors of $L^{\mathbb{C}}$ have to do with the geometric action of L on $\mathbb{R}^n$?

To answer this, we consider an eigenvalue $\lambda \in \mathbb{C} \setminus \mathbb{R}$ and the corresponding eigenvector $\mathbf{z} \in \mathbb{C}^n$. Obviously since $\lambda \notin \mathbb{R}$ we have $\mathbf{z} \notin \mathbb{R}^n$; how do we extract a real-valued vector from $\mathbf{z}$ on which the action of L is related to λ?

Observe that since the entries of the matrix for L are real-valued, the coefficients of the characteristic polynomial $p(\lambda)$ are real-valued. It follows that (14.5) is invariant under the involution $\lambda \mapsto \bar{\lambda}$, and hence if $\lambda \in \mathbb{C} \setminus \mathbb{R}$ is an eigenvalue of $L^{\mathbb{C}}$, so is $\bar{\lambda}$. Furthermore, one may easily verify that $L^{\mathbb{C}}\bar{\mathbf{z}} = \bar{\lambda}\bar{\mathbf{z}}$, where $\bar{\mathbf{z}}$ is defined in the obvious way as

$$\bar{\mathbf{z}} = (\bar{z}_1, \bar{z}_2, \dots, \bar{z}_n) = \mathbf{x} - i\mathbf{y},$$

where $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Observe that $\mathbf{x} = (\mathbf{z} + \bar{\mathbf{z}})/2$ and $\mathbf{y} = i(\mathbf{z} - \bar{\mathbf{z}})/2$; thus the two-dimensional *complex* subspace of $\mathbb{C}^n$ spanned by $\mathbf{z}$ and $\bar{\mathbf{z}}$ intersects $V_R = \mathbb{R}^n$ in the two-dimensional *real* subspace spanned by $\mathbf{x}$ and $\mathbf{y}$.

To see how L acts on this subspace, write $\lambda = \rho e^{i\theta}$, where $\rho > 0$ and $\theta \in [0, 2\pi)$. Then we have

$$\begin{aligned} L\mathbf{x} + iL\mathbf{y} = L^{\mathbb{C}}\mathbf{z} &= \lambda\mathbf{z} \\ &= \rho(\cos\theta + i\sin\theta)(\mathbf{x} + i\mathbf{y}) \\ &= \rho(\cos\theta\mathbf{x} - \sin\theta\mathbf{y}) + i\rho(\cos\theta\mathbf{y} + \sin\theta\mathbf{x}), \end{aligned}$$

[2] It turns out that there are other settings, beyond that of linear spaces, in which one can go from a real structure to a complex structure with the help of a linear operator J with the property that $J^2 = -\operatorname{Id}$. The most accessible example (which is also one of the most important) is the theory of *Riemann surfaces*.

and so L acts on the two-dimensional subspace spanned by $\mathbf{x}$ and $\mathbf{y}$ as a spiral motion—rotation by θ scaled by ρ, with matrix

$$\rho R_\theta = \begin{pmatrix} \rho\cos\theta & -\rho\sin\theta \\ \rho\sin\theta & \rho\cos\theta \end{pmatrix}.$$

Now suppose $L^{\mathbb{C}}$ is diagonalisable over $\mathbb{C}$—that is, there exists $C \in GL(n,\mathbb{C})$ such that

$$CL^{\mathbb{C}}C^{-1} = \operatorname{diag}\left(\lambda_1, \dots, \lambda_j, \rho_1 e^{i\theta_1}, \rho_1 e^{-i\theta_1}, \dots, \rho_k e^{i\theta_k}, \rho_k e^{-i\theta_k}\right),$$

where $\lambda_i \in \mathbb{R}$, $\rho_i > 0$, $\theta_i \in (0,\pi)$, and $j + 2k = n$. Then using the above procedure, one obtains a basis for $\mathbb{R}^n$ in which the matrix of L is

$$\operatorname{diag}\left(\lambda_1, \dots, \lambda_j, \rho_1 R_{\theta_1}, \dots, \rho_k R_{\theta_k}\right). \tag{14.8}$$

Thus while L cannot be diagonalised over $\mathbb{R}$, it can at least be put into block diagonal form, provided $L^{\mathbb{C}}$ can be diagonalised over $\mathbb{C}$. But is even this much always possible?

d. Differing multiplicities and Jordan blocks. Observe that since the determinant of any upper-triangular matrix is the product of the diagonal entries, the characteristic polynomial of an upper-triangular matrix L is

$$\det(L - \lambda\,\mathrm{Id}) = \prod_{i=1}^{n}(L_{ii} - \lambda).$$

Thus the eigenvalues of L are simply the diagonal entries.

Example 14.4. Consider the matrix $L = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Its only eigenvalue is 1, and it has $(1,0)$ as an eigenvector. In fact, this is the *only* eigenvector (up to scalar multiples); this fact can be shown directly, or one can observe that if L were diagonalisable, then we would have $CLC^{-1} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ for some $C \in GL(n,\mathbb{R})$, which would then imply $L = \mathrm{Id}$, a contradiction.

This example shows that not every matrix is diagonalisable over $\mathbb{C}$, and hence not every matrix can be put in block diagonal form over $\mathbb{R}$. In general, this occurs whenever L has an eigenvalue λ for which the geometric multiplicity (the number of linearly independent eigenvectors) is strictly less than the algebraic multiplicity (the number of times λ appears as a root of the characteristic polynomial). In this case the eigenspace corresponding to λ is not as big as it "should" be. A notion of *generalised eigenspace* can be introduced, and it can be shown that every matrix can be put in *Jordan normal form*.

We shall not go through the details of this here; rather, we observe that the non-existence of a basis of eigenvectors is a result of the fact that as we select eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots$, we reach a point where there is no L-invariant subspace transverse to the subspace spanned by $\mathbf{v}_1, \dots, \mathbf{v}_k$, and thus no further eigenvectors can be found. For *orthogonal* matrices, we avoid this problem, as follows.

Let $V \subset \mathbb{R}^n$ be an invariant subspace for L—that is, $L(V) = V$—and let $V^\perp$ be the orthogonal complement of $\mathbb{R}^n$,

$$V^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \langle \mathbf{v}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in V\}.$$

Given $\mathbf{v} \in V^\perp$, we have $\langle L\mathbf{v}, L\mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle$ for all $\mathbf{w} \in V$, and hence $L\mathbf{v} \in V^\perp$. It follows that $V^\perp$ is invariant, and so there exists an eigenvector of L in $V^\perp$ (or perhaps a two-dimensional space on which L acts as ρR_θ). Continuing in this way, we can diagonalise $L^{\mathbb{C}}$, and hence put L in the form (14.8).

Finally, we observe that any eigenvalue of an orthogonal matrix must have absolute value one. This follows since the determinant of L restricted to any invariant subspace is equal to 1. It follows that (14.8) reduces to the form (14.2) and hence every orthogonal matrix can be brought to this form by an orthogonal transformation.

Lecture 15

a. Hermitian product and unitary matrices. One can extend the scalar product on $\mathbb{R}^n$ to a *Hermitian product* on $\mathbb{C}^n$ by

$$\langle \mathbf{z}, \mathbf{w} \rangle = \sum_{j=1}^{n} z_j \overline{w_j}. \tag{15.1}$$

The Hermitian product satisfies similar properties to the scalar product:

(1) $\langle \mathbf{w}, \mathbf{w} \rangle \geq 0$, with equality if and only if $\mathbf{w} = \mathbf{0}$.
(2) $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{w}, \mathbf{v} \rangle}$.
(3) *Linearity*: $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$.
(4) $\langle \lambda \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle \mathbf{v}, \lambda \mathbf{w} \rangle = \bar{\lambda} \langle \mathbf{v}, \mathbf{w} \rangle$.

This device will allow to find a natural extension of the theory of orthogonal matrices to the complex domain.

It may not be immediately apparent why we should use (15.1) instead of the more natural-looking extension $\sum_{j=1}^n z_j w_j$. One could define a scalar product on $\mathbb{C}^n$ using the latter formula; however, one would obtain a totally different sort of beast than the one we now consider. In particular, the Hermitian product defined in (15.1) has the following property: If $\mathbf{z} = \mathbf{x} + i\mathbf{y}$ and $\mathbf{w} = \mathbf{u} + i\mathbf{v}$ for real vectors $\mathbf{x}, \mathbf{y}, \mathbf{u}, \mathbf{v}$, then

$$\begin{aligned} \langle \mathbf{z}, \mathbf{w} \rangle &= \sum_{j=1}^{n} (x_j + iy_j)(u_j - iv_j) \\ &= \sum_{j=1}^{n} (x_j u_j + y_j v_j) + i(y_j u_j - x_j v_j). \end{aligned} \tag{15.2}$$

Hence the real part of $\langle \mathbf{z}, \mathbf{w} \rangle$ is the *real* scalar product of the vectors $(x_1, y_1, \dots x_n, y_n)$ and $(u_1, v_1, \dots, u_n, v_n)$ in $\mathbb{R}^{2n}$. Thus the Hermitian product is a natural generalisation of the real scalar product, and we see that the complex conjugate

$\overline{w_j}$ must be used in order to avoid a negative sign in front of the term $y_j v_j$ in (15.2).

Furthermore, the presence of the complex conjugate in (15.1) is crucial in order to guarantee that

$$\langle \mathbf{z}, \mathbf{z} \rangle = \sum_{j=1}^{n} z_j \overline{z_j} = \sum_{j=1}^{n} |z_j|^2$$

is a non-negative real number, which vanishes if and only if $\mathbf{z} = \mathbf{0}$. In particular, the Hermitian product defines a *norm* on $\mathbb{C}^n$ by $\|\mathbf{z}\|^2 = \langle \mathbf{z}, \mathbf{z} \rangle$, with the following properties.

(1) $\|\mathbf{z}\| \geq 0$, with equality if and only if $\mathbf{z} = \mathbf{0}$.
(2) $\|\lambda \mathbf{z}\| = |\lambda| \, \|\mathbf{z}\|$ for all $\lambda \in \mathbb{C}$.
(3) $\|\mathbf{z} + \mathbf{w}\| \leq \|\mathbf{z}\| + \|\mathbf{w}\|$ for all $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$.

The norm provides a notion of length, and the Hermitian product provides a notion of orthogonality: as in the real case, two vectors $\mathbf{w}, \mathbf{z} \in \mathbb{C}^n$ are orthogonal if $\langle \mathbf{w}, \mathbf{z} \rangle = 0$. Thus we once again have a notion of an *orthonormal basis*—that is, a basis $\{\mathbf{z}^1, \dots, \mathbf{z}^n\}$ of $\mathbb{C}^n$ such that

$$\left\langle \mathbf{z}^j, \mathbf{z}^k \right\rangle = \delta_{jk},$$

where δ_{jk} is the Kronecker delta, which takes the value 1 if $j = k$ and 0 otherwise.

As in $\mathbb{R}^n$, we have a standard orthonormal basis $\mathcal{E} = \{\mathbf{e}^1, \dots, \mathbf{e}^n\}$:

$$\mathbf{e}^j = (0, \dots, 0, 1, 0, \dots, 0),$$

where the 1 appears in the jth position. An orthonormal basis corresponds to a decomposition of the vector space into one-dimensional subspaces which are pairwise orthogonal. In both $\mathbb{R}^n$ and $\mathbb{C}^n$, we can generate other orthonormal bases from $\mathcal{E}$ without changing the subspaces in the decomposition: simply replace $\mathbf{e}^j$ with a parallel unit vector. In $\mathbb{R}^n$, the only parallel unit vector to $\mathbf{e}^j$ is $-\mathbf{e}^j$; in $\mathbb{C}^n$, we can replace $\mathbf{e}^j$ with $\lambda \mathbf{e}^j$, where $\lambda \in S^1$ is any complex number with $|\lambda| = 1$.

This distinction is related to a fundamental difference between $\mathbb{R}^n$ and $\mathbb{C}^n$. In the former case, replacing $\mathbf{e}^j$ with $-\mathbf{e}^j$ changes the orientation of the basis, and hence we can distinguish between even and odd orientations. In $\mathbb{C}^n$, this replacement can be done continuously by moving $\mathbf{e}^j$ to $e^{i\theta}\mathbf{e}^j$ for $0 \leq \theta \leq \pi$; consequently, there is no meaningful way to say where the "orientation" reverses. In fact, in $\mathbb{C}^n$ we must abandon the notion of orientation entirely, and can no longer speak of even and odd maps.

Definition 15.1. A linear map $A \colon \mathbb{C}^n \to \mathbb{C}^n$ is *unitary* if $\langle A\mathbf{z}, A\mathbf{w} \rangle = \langle \mathbf{z}, \mathbf{w} \rangle$ for all $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$. The group of unitary $n \times n$ complex matrices is denoted $U(n)$.

Observe that since the real part of the Hermitian product is just the usual real scalar product on $\mathbb{R}^{2n}$, every unitary map on $\mathbb{C}^n$ corresponds to an orthogonal map on $\mathbb{R}^{2n}$. The converse is not true; there are orthogonal maps on $\mathbb{R}^{2n}$ which are not unitary maps on $\mathbb{C}^n$. Indeed, such a map may not even be linear on $\mathbb{C}^n$; it must behave properly with respect to multiplication by i.

However, unitary maps *are* a generalisation of orthogonal maps in the following sense: given an orthogonal linear map $L\colon \mathbb{R}^n \to \mathbb{R}^n$, the complexification $L^{\mathbb{C}}\colon \mathbb{C}^n \to \mathbb{C}^n$ is unitary.

PROPOSITION 15.2. *If $A\colon \mathbb{C}^n \to \mathbb{C}^n$ is unitary and λ is an eigenvalue of A, then $|\lambda| = 1$.*

PROOF. Let $\mathbf{z} \in \mathbb{C}^n$ be an eigenvector for λ, and observe that

$$\langle \mathbf{z}, \mathbf{z} \rangle = \langle A\mathbf{z}, A\mathbf{z} \rangle = \langle \lambda\mathbf{z}, \lambda\mathbf{z} \rangle = \lambda\overline{\lambda} \langle \mathbf{z}, \mathbf{z} \rangle,$$

and hence

$$\lambda\overline{\lambda} = |\lambda|^2 = 1. \qquad \square$$

Because $\mathbb{C}$ is algebraically closed, the general normal form for (complex) unitary matrices is simpler than the result in the previous lectures for (real) orthogonal matrices. The proof, however, is basically the same, and relies on the fact that preservation of the (real or complex) scalar product guarantees the existence of invariant transverse subspaces.

LEMMA 15.3. *Every linear map $L\colon \mathbb{C}^k \to \mathbb{C}^k$ has an eigenvector.*

PROOF. Because $\mathbb{C}$ is algebraically complete, the characteristic polynomial $p(\lambda) = \det(L - \lambda \operatorname{Id})$ has a root λ_0. Thus $\det(L - \lambda_0 \operatorname{Id}) = 0$, and it follows that there exists $\mathbf{w} \in \mathbb{C}^k$ such that $(L - \lambda_0 \operatorname{Id})\mathbf{w} = \mathbf{0}$. This $\mathbf{w}$ is an eigenvector of L. $\square$

Recall that given a linear map $L\colon V \to V$, a subspace W is *invariant* if $L(W) \subset W$. If $W \subset \mathbb{C}^n$ is an invariant subspace of L, then we may apply Lemma 15.3 to $\mathbb{C}^k = W$ and obtain the existence of an eigenvector in W.

The relationship between eigenvectors and invariant subspaces may be made even more explicit by the observation that an eigenvector is precisely a vector which spans a one-dimensional invariant subspace.

DEFINITION 15.4. Let V be a vector space and $W \subset V$ a subspace. A subspace $W' \subset V$ is *transversal* to W if $W \cap W = \{\mathbf{0}\}$ and if $V = W + W'$. Equivalently, W and W' are transversal if for any $\mathbf{v} \in V$, there exist *unique* vectors $\mathbf{w} \in W$ and $\mathbf{w}' \in W'$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}'$.

If $\langle \cdot, \cdot \rangle$ is a Hermitian product on $\mathbb{C}^n$ and $W \subset \mathbb{C}^n$ is a subspace, then the *orthogonal complement* of W is

$$W^\perp = \{\mathbf{z} \in \mathbb{C}^n \mid \langle \mathbf{z}, \mathbf{w} \rangle = 0 \text{ for all } \mathbf{w} \in W\}.$$

PROPOSITION 15.5. *Let $A\colon \mathbb{C}^n \to \mathbb{C}^n$ be unitary and $W \subset \mathbb{C}^n$ be invariant. Then $W^\perp$ is invariant as well.*

Proof. Observe that since A is unitary, A^{-1} is as well. Thus given $\mathbf{z} \in W^\perp$ and $\mathbf{w} \in W$, we have

$$\langle A\mathbf{z}, \mathbf{w}\rangle = \left\langle A^{-1}A\mathbf{z}, A^{-1}\mathbf{w}\right\rangle = \left\langle \mathbf{z}, A^{-1}\mathbf{w}\right\rangle . \tag{15.3}$$

Furthermore, since A is invertible and W is finite-dimensional, we have $A^{-1}(W) = W$, and hence the quantity in (15.3) vanishes. Since $\mathbf{w} \in W$ was arbitrary, it follows that $A\mathbf{z} \in W^\perp$. □

Proposition 15.6. *Given a linear map $L\colon \mathbb{C}^n \to \mathbb{C}^n$, the following are equivalent:*

(1) L is unitary.
(2) If $\mathcal{U} = \{\mathbf{u}^1, \dots, \mathbf{u}^n\}$ is any *orthonormal basis for $\mathbb{C}^n$, then $L(\mathcal{U})$ is again an orthonormal basis.*
(3) There exists an orthonormal basis $\mathcal{U}$ such that $L(\mathcal{U})$ is again an orthonormal basis.

Proof. That (1) implies (2) is immediate from the definition of unitary, and (2) is *a priori* stronger than (3). Finally, if (3) holds, then for any $\mathbf{w}, \mathbf{z} \in \mathbb{C}^n$ we may decompose $\mathbf{w} = \sum_j w_j \mathbf{u}^j$ and $\mathbf{z} = \sum_k z_k \mathbf{u}^k$, obtaining

$$\begin{aligned}\langle L\mathbf{w}, L\mathbf{z}\rangle &= \sum_{j,k} w_j z_k \left\langle L\mathbf{u}^j, L\mathbf{u}^k\right\rangle \\ &= \sum_{j,k} w_j z_k \delta_{jk} = \sum_{j,k} w_j z_k \left\langle \mathbf{u}^j, \mathbf{u}^k\right\rangle = \langle \mathbf{w}, \mathbf{z}\rangle .\end{aligned}$$ □

Now we can state the fundamental theorem on classification of unitary matrices.

Theorem 15.7. *For every $A \in U(n)$ there exists $C \in U(n)$ such that*

$$CAC^{-1} = \operatorname{diag}(\lambda_1, \dots, \lambda_n), \tag{15.4}$$

where $|\lambda_j| = 1$ for $1 \le j \le n$.

Proof. We apply Lemma 15.3 and Proposition 15.5 repeatedly. First let $\mathbf{u}^1 \in \mathbb{C}^n$ be any unit eigenvector of A, and let W_1 be the subspace spanned by $\mathbf{u}^1$. Then $W_1^\perp$ is invariant, and so there exists a unit eigenvector $\mathbf{u}^2 \in W_1^\perp$. Let W_2 be the subspace spanned by $\mathbf{u}^1$ and $\mathbf{u}^2$, and continue in this manner.

Thus we obtain an orthonormal basis $\{\mathbf{u}^1, \dots, \mathbf{u}^n\}$ such that $A\mathbf{u}^j = \lambda_j \mathbf{u}^j$ for $1 \le j \le n$. By Proposition 15.2, we have $|\lambda_j| = 1$ for every j. Furthermore, if we let C be the $n \times n$ complex matrix such that $C\mathbf{u}^j = \mathbf{e}^j$, then it follows from Proposition 15.6 that C is unitary, and furthermore,

$$CAC^{-1}\mathbf{e}_j = CA\mathbf{u}^j = C(\lambda_j \mathbf{u}^j) = \lambda_j \mathbf{e}^j,$$

which is enough to establish (15.4). □

For real matrices, we considered the special orthogonal group $SO(n)$ within the orthogonal group $O(n)$. We can do the same here and consider the *special unitary group*

$$SU(n) = \{A \in U(n) \mid \det A = 1\}.$$

However, in the complex case, the relationship between $SU(n)$ and $U(n)$ is much closer to the relationship between $SL(n,\mathbb{R})$ and $GL(n,\mathbb{R})$ than it is to the relationship between $SO(n)$ and $O(n)$. In particular, observe that $SO(n)$ is a subgroup of index 2 in $O(n)$, while $SL(n,\mathbb{R})$ and $SU(n)$ both have infinite index in their respective groups.

The group $SU(2)$ deserves particular attention. One immediately sees that it consists of all matrices of the form $\left(\begin{smallmatrix} z & w \\ -\bar{w} & \bar{z} \end{smallmatrix}\right)$, where $|z^2| + |w^2| = 1$. Thus $z = x + iy$ and $w = s + it$ satisfy $x^2 + y^2 + s^2 + t^2 = 1$ so that topologically $SU(2)$ is the three-dimensional sphere.

The quaternionic group Q embeds into $SU(2)$ (and hence into $SO(4)$). Let $J = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$; notice that $J^2 = -\operatorname{Id}$. The embedding is defined by

$$\pm 1 \mapsto \pm \operatorname{Id}, \quad \pm i \mapsto \pm i \operatorname{Id}, \quad \pm j \mapsto \pm J, \quad \pm k \mapsto \pm iJ.$$

b. Normal matrices. We are interested in the class of matrices which can be diagonalised over $\mathbb{C}$, because such matrices have a simpler geometric meaning than matrices with no such diagonalisation. We have seen that this class does not include all matrices, thanks to the existence of matrices like $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Conversely, we have seen that this class *does* include all unitary matrices.

Of course, there are plenty of matrices which can be diagonalised but are not unitary; in particular, we may consider diagonal matrices $\operatorname{diag}(\lambda_1, \dots, \lambda_n)$ for which the eigenvalues λ_j do not lie on the unit circle—that is, $|\lambda| \neq 1$. Can we give a reasonable characterisation of the class of matrices which can be diagonalised over $\mathbb{C}$?

Remark. In the present setting, this question may seem somewhat academic, since any matrix can be put in Jordan normal form, which already gives us a complete understanding of its action on $\mathbb{C}^n$ (or $\mathbb{R}^n$). However, it turns out to be vital to understanding what happens in the infinite-dimensional situation, where $\mathbb{C}^n$ is replaced with the more general concept of a *Hilbert space*, and eigenvalues and eigenvectors give way to *spectral theory*. In this general setting there is no analogue of Jordan normal form, and the class of maps we examine here turns out to be very important.

Recall that given a real $n \times n$ matrix A (which may or may not be orthogonal), the *transpose* of A defined by $(A^T)_{ij} = A_{ji}$ has the property that

$$\langle \mathbf{x}, A\mathbf{y} \rangle = \left\langle A^T \mathbf{x}, \mathbf{y} \right\rangle$$

for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. For complex vectors and the Hermitian product, the analogous matrix is called the *adjoint* of A; it is denoted A^* and has the property that

$$\langle \mathbf{z}, A\mathbf{w}\rangle = \langle A^*\mathbf{z}, \mathbf{w}\rangle$$

for every $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$.

EXERCISE 15.1. Show that the matrix of A^* is the *conjugate transpose* of the matrix of A—that is, that

$$(A^*)_{ij} = \overline{A_{ji}}.$$

Recall that a matrix A is unitary if and only if $\langle A\mathbf{z}, A\mathbf{w}\rangle = \langle \mathbf{z}, \mathbf{w}\rangle$ for all $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$. This is equivalent to the condition that $\langle A^*A\mathbf{z}, \mathbf{w}\rangle = \langle \mathbf{z}, \mathbf{w}\rangle$ for all $\mathbf{z}$ and $\mathbf{w}$, which is in turn equivalent to the condition that $A^*A = \mathrm{Id}$. In particular, this implies that $A^* = A^{-1}$, and hence A and A^* commute.

DEFINITION 15.8. $A \in M(n, \mathbb{C})$ is *normal* if $A^*A = AA^*$.

Every unitary matrix is normal, but there are normal matrices which are not unitary. This follows immediately from the fact that normality places no restrictions on the eigenvalues of A; in particular, every scalar multiple of the identity matrix is normal, but $\lambda \,\mathrm{Id}$ is only unitary if $|\lambda| = 1$.

It turns out that normality is precisely the condition we need in order to make the argument from the previous section go through (modulo the statement about the absolute values of the eigenvalues). In particular, we can prove an analogue of Proposition 15.5, after first making some general observations.

First we observe that given $A \in M(n, \mathbb{C})$ and $\lambda \in \mathbb{C}$, we have

$$\langle (A - \lambda \,\mathrm{Id})\mathbf{z}, \mathbf{w}\rangle = \langle A\mathbf{z}, \mathbf{w}\rangle - \lambda \langle \mathbf{z}, \mathbf{w}\rangle = \langle \mathbf{z}, A^*\mathbf{w}\rangle - \langle \mathbf{z}, \overline{\lambda}\mathbf{w}\rangle = \langle \mathbf{z}, (A^* - \overline{\lambda}\,\mathrm{Id})\mathbf{w}\rangle$$

for every $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$, and hence

$$(A - \lambda \,\mathrm{Id})^* = A^* - \overline{\lambda}\,\mathrm{Id}\,. \tag{15.5}$$

PROPOSITION 15.9. *If $B \in M(n, \mathbb{C})$ is normal, then $\ker B = \ker B^*$.*

PROOF. Suppose $B\mathbf{w} = \mathbf{0}$. Then we have

$$\|B^*\mathbf{w}\|^2 = \langle B^*\mathbf{w}, B^*\mathbf{w}\rangle = \langle BB^*\mathbf{w}, \mathbf{w}\rangle = \langle B^*B\mathbf{w}, \mathbf{w}\rangle = 0,$$

and it follows that $\ker B \subset \ker B^*$. Equality holds since $B = (B^*)^*$. □

Applying Proposition 15.9 to $B = A - \lambda \,\mathrm{Id}$ and using (15.5), we see that if $\mathbf{w}$ is an eigenvector of A with eigenvalue λ, then it is an eigenvector of A^* with eigenvalue $\overline{\lambda}$. In particular, if W is the subspace spanned by $\mathbf{u}^1, \dots, \mathbf{u}^k$, where each $\mathbf{u}^j$ is an eigenvector of A, then each $\mathbf{u}^j$ is an eigenvector of A^* as well, and hence $A^*W \subset W$.

Now we have the following analogue of Proposition 15.5.

PROPOSITION 15.10. *Let $A \in M(n, \mathbb{C})$ be normal, and let $W \subset \mathbb{C}^n$ be an invariant subspace spanned by eigenvectors of A. Then $W^\perp$ is an invariant subspace as well.*

Proof. Given $\mathbf{z} \in W^\perp$ and $\mathbf{w} \in W$, observe that

$$\langle A\mathbf{z}, \mathbf{w}\rangle = \langle \mathbf{z}, A^*\mathbf{w}\rangle = 0,$$

where the last equality follows since $A^*\mathbf{w} \in W$ (by the above discussion). □

This lets us prove the following generalisation of Theorem 15.7.

Theorem 15.11. *An $n \times n$ complex matrix A is normal if and only if there exists $C \in U(n)$ and $\lambda_j \in \mathbb{C}$ such that*

$$CAC^{-1} = \operatorname{diag}(\lambda_1, \dots, \lambda_n). \tag{15.6}$$

Proof. One direction is easy; if (15.6) holds for some $C \in U(n)$ and $\lambda_j \in \mathbb{C}$, then we write $D = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$, and observe that $A = C^{-1}DC = C^*DC$. Thus we have

$$A^* = (C^*DC)^* = C^*D^*(C^*)^* = C^{-1}D^*C,$$

and we see that

$$AA^* = (C^{-1}DC)(C^{-1}D^*C) = C^{-1}DD^*C = C^{-1}D^*DC = A^*A,$$

where the third equality uses the fact that diagonal matrices commute.

The other direction is a word-for-word repetition of the proof of Theorem 15.7, using Proposition 15.10 in place of Proposition 15.5, and omitting the requirement that $|\lambda_j| = 1$. □

Remark. Normality characterises all matrices which can be diagonalised over $\mathbb{C}$ *with an orthonormal change of coordinates*. There are matrices that can be diagonalised with a change of coordinates which is *not* orthonormal; such matrices are not normal with respect to the standard Hermitian product. Recall that the definition of the adjoint A^* depends on the Hermitian product; if we choose a different Hermitian product on $\mathbb{C}^n$, we obtain a different adjoint, and hence a different class of normal matrices.

c. Symmetric matrices. We have settled the question of which matrices can be diagonalised over $\mathbb{C}$ via an orthonormal change of coordinates. What about the real numbers? There are plenty of matrices which can be diagonalised over $\mathbb{C}$ but which cannot be diagonalised over $\mathbb{R}$; any normal matrix with a non-real eigenvalue falls into this class.

Thus we see immediately that any matrix which can be put into the form (15.6) as a map on $\mathbb{R}^n$ must have only real eigenvalues. In particular, given $A \in M(n, \mathbb{R})$, let $A^{\mathbb{C}} \colon \mathbb{C}^n \to \mathbb{C}^n$ be the complexification of A, and observe that $(A^T)^{\mathbb{C}} = (A^{\mathbb{C}})^*$. It follows from the remarks before Proposition 15.10 that if λ is an eigenvalue of A, then $\overline{\lambda}$ is an eigenvalue of A^T, with the same eigenvectors.

Definition 15.12. A real $n \times n$ matrix such that $A^T = A$ is called *symmetric*; a complex $n \times n$ matrix such that $A^* = A$ is called *Hermitian*.

If $A \in M(n, \mathbb{R})$ is symmetric, then for every eigenvalue λ and eigenvector $\mathbf{w} \in \mathbb{C}^n$ we have

$$\lambda \mathbf{w} = A\mathbf{w} = A^T \mathbf{w} = \overline{\lambda}\mathbf{w},$$

and hence $\lambda = \overline{\lambda}$. Thus symmetric matrices have only real eigenvalues. In particular, since real symmetric matrices are normal, every real symmetric matrix is orthogonally diagonalisable *over the real numbers*. Furthermore, the converse also holds: if C is a real orthogonal matrix such that $D = CAC^{-1}$ is a diagonal matrix with real entries, then

$$A^T = (C^T DC)^T = C^T D^T (C^T)^T = C^T DC = A,$$

and hence A is symmetric.

d. Linear representations of isometries and other classes of transformations. Our discussion of linear algebra began with a quest to understand the isometries of $\mathbb{R}^n$. We have seen various classes of matrices, but have not yet completed that quest—now we are in a position to do so.

We recall the following definition from Lecture 2.

DEFINITION 15.13. A homomorphism $\varphi \colon G \to GL(n, \mathbb{R})$ is called a *linear representation* of G. If $\ker \varphi$ is trivial, we say that the representation is *faithful*.

Informally, a linear representation of a group G is a concrete realisation of the abstract group G as a set of matrices, and it is faithful if no two elements of G are represented by the same matrix. Linear representations are powerful tools, because the group of invertible matrices is general enough to allow us to embed many important abstract groups inside of it, and yet is concrete enough to put all the tools of linear algebra at our disposal in studying the group which is so embedded.

We were able to represent the group of all isometries of $\mathbb{R}^n$ *with a fixed point* as $O(n)$. In order to represent isometries with no fixed point, we must go one dimension higher and consider matrices acting on $\mathbb{R}^{n+1}$.

PROPOSITION 15.14. $\mathrm{Isom}(\mathbb{R}^n)$ *has a linear representation in* $GL(n + 1, \mathbb{R})$. *In particular,* $\mathrm{Isom}^+(\mathbb{R}^n)$ *has a linear representation in* $SL(n+1, \mathbb{R})$.

PROOF. Given $I \in \mathrm{Isom}(\mathbb{R}^n)$, let $\mathbf{b} = -I\mathbf{0}$; then $T_{-\mathbf{b}} \circ I\mathbf{0} = \mathbf{0}$, and hence $A = T_{-\mathbf{b}} \circ I \in O(n)$. Thus $I = T_{\mathbf{b}} \circ A$, and so for every $\mathbf{x} \in \mathbb{R}^n$ we have

$$I\mathbf{x} = T_{\mathbf{b}} \circ A\mathbf{x} = A\mathbf{x} + \mathbf{b}. \tag{15.7}$$

Embed $\mathbb{R}^n$ into $\mathbb{R}^{n+1}$ as the plane

$$P = \{\mathbf{x} \in \mathbb{R}^{n+1} \mid x_{n+1} = 1\}.$$

To the isometry I, associate the following block matrix:

$$\varphi(I) = \begin{pmatrix} A & \mathbf{b} \\ \mathbf{0} & 1 \end{pmatrix}. \tag{15.8}$$

Here $A \in O(n)$, $\mathbf{b}$ is an $n \times 1$ column vector, and $\mathbf{0}$ is a $1 \times n$ row vector. Observe that $\varphi(I) \in GL(n+1, \mathbb{R})$, and that $\varphi(I)$ maps P to itself; if $I \in \mathrm{Isom}^+(\mathbb{R}^n)$, then $\varphi(I) \in SL(n, \mathbb{R})$. Furthermore, the action of $\varphi(I)$ on P is exactly equal to the action of I on $\mathbb{R}^n$, and φ is a homomorphism: given $I_1, I_2 \in \mathrm{Isom}(\mathbb{R}^n)$, we have

$$\varphi(I_2)\varphi(I_1) = \begin{pmatrix} A_2 & \mathbf{b}_2 \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} A_1 & \mathbf{b}_1 \\ \mathbf{0} & 1 \end{pmatrix} = \begin{pmatrix} A_2 A_1 & A_2\mathbf{b}_1 + \mathbf{b}_2 \\ \mathbf{0} & 1 \end{pmatrix},$$

which is equal to $\varphi(I_2 \circ I_1)$ since

$$I_2 \circ I_1 \mathbf{x} = I_2(A_1\mathbf{x} + \mathbf{b}_1) = A_2(A_1\mathbf{x} + \mathbf{b}_1) + \mathbf{b}_2.$$

Finally, we observe that if I is an even isometry, then $\det \varphi(I) = 1$. $\square$

The technique exhibited in the proof of Proposition 15.14 embeds $\mathrm{Isom}(\mathbb{R}^n)$ in $GL(n+1, \mathbb{R}^n)$ as

$$\begin{pmatrix} O(n) & \mathbb{R}^n \\ \mathbf{0} & 1 \end{pmatrix}.$$

Using a the same technique, we can represent the *affine group* $\mathrm{Aff}(\mathbb{R}^n)$, which is the class of all maps which take lines to lines. As will be shown later (Theorem 17.1), every such map can again be written in the form (15.7),[3] but here A may be *any* matrix, not necessarily orthogonal. Thus we embed $\mathrm{Aff}(\mathbb{R}^n)$ into $GL(n+1, \mathbb{R}^n)$ as

$$\begin{pmatrix} GL(n, \mathbb{R}) & \mathbb{R}^n \\ \mathbf{0} & 1 \end{pmatrix}.$$

We may also do this with the group of *similarity transformations*—maps of $\mathbb{R}^n$ which take lines to lines and preserve angles. Every such map may be written as $\mathbf{x} \mapsto \lambda R\mathbf{x} + \mathbf{b}$, where $\lambda \in \mathbb{R}$ and $R \in O(n)$. Thus the group embeds into the general linear group as

$$\begin{pmatrix} \mathbb{R} \cdot O(n) & \mathbb{R}^n \\ \mathbf{0} & 1 \end{pmatrix}.$$

The common thread in all these representations is that all the tools of linear algebra are now at our disposal. For example, suppose we wish to classify isometries of $\mathbb{R}^n$, and have forgotten all the synthetic geometry we ever knew. Then we observe that every isometry can be written as $I\mathbf{x} = A\mathbf{x} + \mathbf{b}$, and note that I has a fixed point if and only if

$$A\mathbf{x} + \mathbf{b} = \mathbf{x}$$

has a solution—that is, if and only if $\mathbf{b}$ lies in the range of $A - \mathrm{Id}$. If 1 is not an eigenvalue of A, then $A - \mathrm{Id}$ is invertible, and I has a fixed point. If 1 *is* an eigenvalue of A, then $\mathbf{b}$ may not lie in the range of $A - \mathrm{Id}$, which is the orthogonal complement to the eigenspace L_1 of vectors with eigenvalue 1; in this case I has no fixed points. As before let us decompose $\mathbf{b} = \mathbf{b}_1 + \mathbf{b}_2$ where $\mathbf{b}_1 \in L_1$, i.e $A\mathbf{b}_1 = \mathbf{b}_1$, and $\mathbf{b}_2$ orthogonal to L_1, i.e. in the range of

[3]For the time being we take (15.7) as our definition of an affine map.

$A - \mathrm{Id}$. Then I is the composition of the isometry $I' \colon \mathbf{x} \to A\mathbf{x} + \mathbf{b}_2$ that has a fixed point and translation $T_{\mathbf{b}_1}$ by $\mathbf{b}_1$. I' is conjugate via a translation to the linear isometry A and I to the product of it with the translation $T_{\mathbf{b}_1}$. Notice that $AT_{\mathbf{b}_1}\mathbf{x} = A(\mathbf{x} + \mathbf{b}_1) = Ax + A\mathbf{b}_1 = A\mathbf{x} + \mathbf{b}_1 = T_{\mathbf{b}_1}A\mathbf{x}$.

Thus any isometry I without fixed points is the product of a commuting pair comprising an isometry I_0 with many fixed points and a translation along the fixed set of that isometry. Depending on the dimension of the fixed set for I_0 we obtain different geometric types of fixed point free isometries.

Similar arguments provide for the classification of similarity transformations and affine transformations without fixed points.

Lecture 16

a. The projective line. In the previous lecture we saw that certain linear groups correspond to various "geometries". Although we have spent most of our time studying the group of isometries of $\mathbb{R}^n$, we also saw that the group of affine transformations and the group of similarity transformations appear as subgroups of $GL(n+1, \mathbb{R})$. Thus we may go beyond the usual Euclidean structure of $\mathbb{R}^n$ and consider instead the affine structure of $\mathbb{R}^n$, or perhaps think about Euclidean geometry up to similarity. Each of the above examples arose from considering subgroups of the affine transformations on $\mathbb{R}^n$; that is, subgroups of $GL(n+1, \mathbb{R})$ of the form

$$\begin{pmatrix} G & \mathbb{R}^n \\ \mathbf{0} & 1 \end{pmatrix}, \tag{16.1}$$

where G is a subgroup of $GL(n, \mathbb{R})$. Such subgroups act on the n-dimensional affine subspace $P = \{\mathbf{x} \in \mathbb{R}^{n+1} \mid x_{n+1} = 1\}$.

There are other matrix groups which are of interest to us, and it turns out that they too correspond to certain geometries. The difference is that those geometries do not necessary have Euclidean space as their phase space.

In this lecture we will broaden our horizons beyond the groups of the form (16.1), examining instead the action of all of $GL(n+1, \mathbb{R})$. This will lead us in the end to *projective geometry*. We will take our time getting there, however, because the story is not quite as straightforward as it was before; for example, observe that most linear transformations of $\mathbb{R}^{n+1}$ do not preserve the subspace P, and so it is not at all obvious in what sense they are to "act" on P. We may pretend though that they do and see what comes out of it.

The fundamental fact that we *do* know about elements of $GL(n+1, \mathbb{R})$ is that they map lines to lines. Thus it makes sense to consider the action of $GL(n+1, \mathbb{R})$ on lines in $\mathbb{R}^{n+1}$; we begin in the simplest case, $n = 1$. Here we have $GL(2, \mathbb{R})$ acting on $\mathbb{R}^2$, and in particular, on the following object.

Definition 16.1. The *real projective line* $\mathbb{R}P(1)$ is the set of all lines through the origin in $\mathbb{R}^2$.