# WOMP Talk 1, Part 1: Algebra I

## Vector spaces and linear transformations

Travis Schedler

**Abstract**

We review the notion of a vector space, basis and dimension, linear transformations between vector spaces, dual vector spaces and transformations, spectral decomposition for normal operators (which includes symmetric, Hermitian, orthogonal, and unitary operators), and determinants. Along the way we review direct-sum decompositions, bilinear forms and inner product spaces, adjoints, characteristic polynomials, and generalized-eigenspace decompositions for algebraically closed fields. The discussion of determinants includes a correct definition of the exterior (alternating) product of dual vector spaces.

This talk is entirely review, except possibly for the spectral theorem (which you should at least have seen for symmetric/Hermitian or orthogonal/unitary operators).

# 1 Preliminaries

Let me spend one minute reviewing groups, rings, and modules. A **group** $G$ is a set together with a multiplication $G \times G \to G$ which is associative (i.e. $(ab)c = a(bc), \forall a, b, c \in G$), has a unit $e \in G$ satisfying $eg = ge = g, \forall g \in G$, and has inverses: for every $g \in G$ there exists $g^{-1} \in G$ such that $gg^{-1} = g^{-1}g = e$. An **abelian group** is a group $G$ which is also commutative: $ab = ba, \forall a, b \in G$. Usually we will regard the "multiplication" in an abelian group as an "addition" since usual addition is commutative.

Common examples of groups: permutation groups $S_n$ on $n$ letters and the alternating groups $A_n$ of even permutations; cyclic groups $\mathbb{Z}, \mathbb{Z}/n$; the rationals $\mathbb{Q}$, reals $\mathbb{R}$, and complex numbers $\mathbb{C}$ under addition; the "general linear" group $GL(V)$ of invertible linear transformations on a vector space $V$ (to be defined later), the groups $SL(V), O(V), SO(V), U(V), SU(V), Sp(V)$ of, in order: special linear group of invertible transformations having determinant 1, orthogonal group of transformations preserving a symmetric bilinear form over a (real) vector space, special orthogonal group of orthogonal transformations of determinant 1, unitary group of transformations preserving a Hermitian inner product over a complex vector space, special unitary group of unitary transformations of determinant 1, and the symplectic group of transformations preserving a symplectic form on a (real) vector space. For the groups $GL, SL$, one can get finite groups if working over a finite field; $O, SO$, and $Sp$ can also be generalized to finite fields to get finite groups. Other groups one could consider include groups of units of fields such as $\mathbb{R}^x, \mathbb{C}^x$, or even the noncommutative group of unit quaternions $\mathbb{H}^x$. One can also consider groups of automorphisms of various other objects, such as geometric spaces, other groups, etc.

A **ring** $R$ is an abelian group (which we consider the addition, with unit 0) together with an associative multiplication with unit 1 satisfying the distributive property: $a(b+c) = ab + ac, \forall a, b, c \in R$. In other words, defining a **monoid** to be a set with an associative

multiplication with unit, $R$ is both an abelian group under addition and a monoid under multiplication, which also satisfies the distributive property. An **abelian (or commutative) ring** $R$ is a ring which is also multiplicatively commutative: $ab = ba, \forall a, b \in R$.

We will always use the notation 0 for the additive unit and 1 for the multiplicative unit in a ring. Note that the distributive property implies in particular that $0x = x0 = 0$ for all $x \in R$. So, the zero element is never multiplicatively invertible. If every nonzero element is invertible, the ring is called a **division ring**, and a commutative division ring is called a **field**. More generally, for any ring $R$, we can define $R^x$ to be the group of **units**, or invertible elements, of $R$. $R^x = R \setminus \{0\}$ iff $R$ is a division ring.

Common examples of rings: $\mathbb{Z}, \mathbb{Z}/n$, rings $\mathbb{Z}[\sqrt{2}], \mathbb{Z}[i]$ or other such subrings of $\mathbb{C}$; direct products of these; rings of $n \times n$ matrices or their subrings, which can contain coefficients in any other ring. Common examples of fields: $\mathbb{R}, \mathbb{C}, \mathbb{F}_p = \mathbb{Z}/p$ for a prime $p$, the other finite fields $\mathbb{F}_{p^k}, k \geq 1$ and $p$ prime. For any ring $R$ we can consider the ring $R[x]$ of polynomials with those coefficients: so $\mathbb{R}[x], \mathbb{C}[x], \mathbb{F}_p[x]$ and their quotients are common rings; by iterating this process we can get rings of polynomials in any number of variables with coefficients in any ring.

Given any groups $G_1, G_2$ we define the notion of a **group homomorphism** $\phi : G_1 \to G_2$ to be a map sending the unit to the unit which is multiplicative: $\phi(e) = e, \phi(gh) = \phi(g)\phi(h) \forall g, h \in G_1$.

Similarly, if $R_1, R_2$ are rings, a **ring homomorphism** $\phi : R_1 \to R_2$ is a map such that $0 \mapsto 0, 1 \mapsto 1$ and $\phi$ is additive and multiplicative: $\phi(ab) = \phi(a)\phi(b), \phi(a + b) = \phi(a) + \phi(b)$. In other words, $\phi$ is simultaneously a group homomorphism under addition and a monoid homomorphism under multiplication.

Any homomorphism which has an inverse homomorphism is called a **isomorphism**. Any homomorphism from an object to itself is called an **endomorphism**; anything which is both an endomorphism and an isomorphism is called an **automorphism**. The collections of endomorphisms or automorphisms of an object $X$ are denoted $\mathrm{End}(X), \mathrm{Aut}(X)$, respectively. The collection of homomorphisms between $X$ and $Y$ is denoted by $\mathrm{Hom}(X, Y)$.

Note that if $G_1, G_2$ are abelian groups, then $\mathrm{Hom}(G_1, G_2)$ is also an abelian group: $(\phi_1 + \phi_2)(g) = \phi_1(g) + \phi_2(g)$. The zero element is the map $0(g) = 0, \forall g \in G_1$. Also, if $G$ is any group, $\mathrm{End}(G)$ is not just an abelian group, but a ring: given two endomorphisms $\phi_1, \phi_2$, we can add them by $(\phi_1 + \phi_2)(x) = \phi_1(x) + \phi_2(x)$ and multiply by composition: $(\phi_1 \circ \phi_2)(x) = \phi_1(\phi_2(x))$. The group $\mathrm{End}(G)^x$ of invertible endomorphisms is the same as $\mathrm{Aut}(G)$. In $\mathrm{End}(G)$, the 1 element is the identity, and the 0 element is the endomorphism sending everything to zero.

Note that, for any abelian group $G$, $\mathrm{Hom}(\mathbb{Z}, G) \cong G$ under the indentification $\phi \mapsto \phi(1)$. Also, for any ring $R$, $R \subset \mathrm{End}(R)$ under the inclusion $r \mapsto \phi_r, \phi_r(s) = rs, \forall s \in R$. So we say that $End(R)$ is an **$R$-algebra**, meaning a ring which contains $R$ (we'll talk more about this later).

Given a ring $R$, a **left $R$-module** $M$ is an abelian group together with a ring homomorphism $\rho : R \to \mathrm{End}(M)$. This is the same thing as a left multiplication $R \times M \to M, r \cdot m \mapsto \rho(r)(m)$ for $r \in R, m \in M$, which satisfies the properties $(r_1 r_2)(m) = r_1(r_2 \cdot m)$,

$(r_1 + r_2)(m) = r_1 m + r_2 m$, and $1 \cdot m = m, 0 \cdot m = 0, \forall m \in M$. A **right $R$-module** is the same thing but for a right multiplication instead of a left: this is equivalent to a ring homomorphism $\rho : R^{\text{op}} \to \text{End}(M)$, where $R^{\text{op}}$ is defined to be the ring with the same set of elements as $R$, but with reversed multiplication: to multiply $x \cdot y$ in $R^{\text{op}}$, we take $y \cdot x$ in $R$. That is, a right $R$-module is an "anti-homomorphism", a map $\rho : R \to \text{End}(M)$ satisfying $\rho(xy) = \rho(y)\rho(x)$, which is an additive homomorphism and sends 1 to 1.

Note that the distinction between left and right modules disappears if the ring is commutative: then we can canonically associate a left module with a right one and vice-versa by $rm = mr$ for $m \in M, r \in R$.

If the ring is not just commutative but a field, then the module can be called a **vector space** over the field. That is, a vector space is, by definition, a module over a field.

A homomorphism between two left $R$-modules $M_1, M_2$ is a map $\phi : M_1 \to M_2$ which is a homomorphism of abelian groups commuting with the multiplication by the ring: that is, $\phi(rm) = r\phi(m), \forall r \in R, m \in M$, and $\phi(m_1 + m_2) = \phi(m_1) + \phi(m_2), \forall m_1, m_2 \in M$. This is sometimes expressed by saying $\phi(r_1 m_1 + r_2 m_2) = r_1\phi(m_1) + r_2\phi(m_2)$. A homomorphism of right $R$-modules is the same thing but with the multiplication on the right. A homomorphism of vector spaces is the same as a homomorphism of modules, but where the ring is a field.

The space of homomorphisms $\text{Hom}(M_1, M_2)$ of two left $R$-modules is itself a left $R$-module: $(r_1\phi_1 + r_2\phi_2)(m) = r_1\phi_1(m) + r_2\phi_2(m)$. The space of endomorphisms $\text{End}(M)$ of an $R$-module is not only an $R$-module, but also a ring (because $M$ is an abelian group) which contains $R$: there is a ring homomorphism $R \to \text{End}(M)$ given by $r \mapsto \phi_r, \phi_r(m) = rm, \forall m \in M$. The ring homomorphism is injective and makes $R$ a subring of $\text{End}(M)$, so $\text{End}(M)$ is an **$R$-algebra**, generalizing the example where $M = R$ itself. The reason this is called an $R$-algebra is because any ring containing $R$ is canonically an $R$-module under multiplication by $R$. We see that this $R$-module structure on $\text{End}(M)$ is the same $R$-module structure we get by considering it to be $\text{Hom}(M, M)$: $r\phi(m) = (\phi_r \circ phi)(m) = (r\phi)(m)$. The multiplicative group $\text{End}(M)^x$ of units in this algebra is the group $\text{Aut}(M)$ of $R$-module automorphisms of $M$.

# 2 Vector spaces, basis, dimension, and rank

Now that we've gotten that out of our system (I'm actually omitting it from my talk, because I only have 50 minutes), we can focus on vector spaces. Here's the definition since I'm starting here for my talk: Recall that a field is a set with addition and multiplication, which are both associative and commutative, satisfying the distributivity property, and having distinct elements 0 and 1 which are the additive and multiplicative identities, respectively; finally, every element has a negative (=additive inverse), and every nonzero element has a reciprocal (=multiplicative inverse). Since the multiplication is commutative, it makes sense to write fractions, since we don't need to worry whether the denominator is inverted on the left or right.

Now, a vector space $V$ is an additive group which has a multiplication by a field $k$, which we call from now on the **field of scalars** or just **scalars**. The multiplication must

be an action in the sense that it gives a homomorphism from the field to the ring of group endomorphisms of the vector space: that is $(a+b)(v) = av+bv$ and $(ab)v = a(bv)$, as well as $0v = 0, 1v = v, \forall v \in V$. Note that, since every nonzero element of the field has a reciprocal, multiplication by anything nonzero must actually be an automorphism of the vector space. This is just a fancy way of saying that for any $c \in k \setminus \{0\}$, the maps $v \mapsto cv$, $v \mapsto \frac{1}{c}v$ are inverses.

A map $\phi : V \to W$ between vector spaces which satisfies $\phi(av + bw) = a\phi(v) + b\phi(w)$ is called a **linear map**, a **homomorphism**, or even just a **map** if there is no confusion. (This is the same as a $k$-module homomorphism, of course.) A subspace $V_0 \subset V$ of a vector space $V$ is a subset which is a vector space under the induced addition and scalar multiplication. Given a subspace $V_0 \subset V$, we can form the quotient $V/V_0$, which is defined as the set of distinct cosets $v + V_0 := \{v + w \mid w \in V\}$ as $v \in V$ varies: this is a vector space since $a(v + V_0) + b(w + V_0) = (av + bw) + V_0$, with $0 + V_0 = V_0$ the zero element.

A map of vector spaces is an **isomorphism** if it has an inverse (which is also linear). This is equivalent to the map being bijective, since any set-theoretic inverse must be linear: $\phi(\phi^{-1}(aw_1 + bw_2)) = aw_1 + bw_2 = \phi(a\phi^{-1}(w_1) + b\phi^{-1}(w_2)) \Rightarrow \phi^{-1}(aw_1 + bw_2) = a\phi^{-1}(w_1) + b\phi^{-1}(w_2)$. So surjectivity and injectivity imply a linear map is an isomorphism.

The term **vector space** comes from the examples of $\mathbb{R}^2, \mathbb{R}^3$, or more generally $\mathbb{R}^n$, where we can think of a vector as being an oriented line segment which can be translated anywhere (but keeping its direction and length fixed), adding two of them by sticking the tail of one at the head of the other. In other words, these vector spaces can be thought of as an $n$-tuple of real numbers which add coordinatewise.

This situation generalizes to $k^n$ for any field $k$: we can always consider $n$-tuples of elements of $k$. Actually, this happens a lot: any finite-dimensional vector space is $k^n$ where $n$ is the dimension. But let's review first the concept of dimension through bases.

Given a vector space $V$ over a field $k$, we can try to make it look like $k^n$ by starting to pick coordinates, or a basis. Given any vectors $v_1, \ldots, v_n \in V$, we can define the **span** as $\text{Span}(v_1, \ldots, v_n) = \{a_1v_1 + a_2v_2 + \ldots + a_nv_n, a_i \in k\}$. If all of these elements in the span are different for different choices of $a_i$, then we see that $\text{Span}(v_1, \ldots, v_n) \cong k^n, a_1v_1 + \ldots + a_nv_n \mapsto (a_1, \ldots, a_n)$. In this case, we say that the set $v_1, \ldots, v_n$ is **linearly independent**, and is a **basis** for the $\text{Span}(v_1, \ldots, v_n)$.

If not, then there are two choices $a_i, a_i'$ such that $a_1v_1 + \ldots a_nv_n = a_1'v_1 + \ldots + a_n'v_n$, which means $(a_1 - a_1')v_1 + \ldots + (a_n - a_n')v_n = 0$. In other words, the elements are not all the same just in the case when there is some nonzero combination of the $v_i$ that gives zero. In this case, we say that the $v_i$ are **linearly dependent**, and any relation of the form $a_1v_1 + \ldots + a_nv_n = 0$ is referred to as a **linear dependence**. In this case, we can pick any $a_i \neq 0$ and see that $v_i = \frac{1}{a_i}\left(\sum_{j \neq i} a_jv_j\right)$. That means that we could have thrown out $v_i$ to get the same span. By continually throwing out a vector every time the set is linearly dependent, we will eventually arise at a set $v_{i_1}, \ldots, v_{i_m}$ with is **linearly independent**, i.e. there is no linear dependence. In this case, we see that $\text{Span}(v_1, \ldots, v_n) = \text{Span}(v_{i_1}, \ldots, v_{i_m}) \cong k^m$. So we just proved that any vector space spanned by a list of $n$ vectors is isomorphic to $k^m$ for some $m \leq n$, with basis some collection of $m$ of the vectors in our original set.

We call this $m$ the **dimension** of the vector space. We haven't proved, though, that $m$ is well-defined: what if the same vector space can have bases of two different sizes? In other words, what if $k^m \cong k^n$ for $m \neq n$? As it turns out, this can't happen. A reason for that is the following: in $k^m$, any set of vectors of size $n > m$ is linearly dependent. This can be proved by viewing a set of $n$ vectors in $k^m$ as an $m \times n$-matrix of coefficients in $k$; by row reduction, if $n > m$, we can eliminate the matrix to one that is of the form

$$\begin{pmatrix} 0 & \dots & 0 & 1 & * & \dots & & & \\ 0 & \dots & 0 & 0 & 0 & \dots & 0 & 1 & \dots \\ \vdots & & \ddots & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.1}$$

so that the bottom $n - m$ rows are all zero (and possibly more): any of these $n - m$ rows gives an "independent" linear dependence. So **dimension** is well defined.

Additionally, given any vector space $V$ of dimension $n$, then any set of vectors which spans $V$ must be at least $n$ in size, for the same reason. Finally, any linearly independent set of $n$ vectors in $V$ must span all of $V$, since it spans an $n$-dimensional subspace, and thus every vector of $V$ is linearly dependent with these $n$ elements, i.e. in its span.

Actually, the proof that dimension is well-defined in fact has shown us more: Given any $n \times m$-matrix, there is a well-defined **rank** of the matrix, which is just the value $k$ such that there are exactly $k$ independent rows. Using column operations, we see that this is the same number as the number of independent columns: swapping rows and adding a multiple of one row to another will not change any linear independence of the columns and vice-versa. So column and row operations leave the number of linearly independent rows and columns the same. When we reduce the matrix to the form (2.1) we see that these numbers are the same: just the number of nonzero rows (equivalently the number of nonzero columns) which gives the rank. Clearly for any $n \times m$-matrix, the rank is $\leq \min(m, n)$.

Let us generalize to vector spaces that are not spanned by a finite number of vectors; for example, the vector space of infinite sequences $(a_0, a_1, a_2, \dots)$ of elements of $k$, or say the space of real functions $\mathbb{R} \to \mathbb{R}$ (we could restrict to continuous ones, or differentiable ones, or whatever). In the general case, we define the span of a possibly infinite subset $S \subset V$ as $\text{Span}(S) = \{a_1 v_1 + a_2 v_2 + \dots a_n v_n \mid v_1, v_2, \dots, v_n \in S\}$. We say $S$ is **linearly independent** if there is no linear dependence among any finite number of elements of $S$. We say that $S$ is a **basis** for $V$ if it spans $V$ and is linearly independent.

For any vector space that is not spanned by a finite number of vectors, we can still find a basis of the vector space using Zorn's lemma: we start with a vector, then find a linearly independent vector, then a vector linearly independent with the first two, etc., and eventually (after a possibly large infinite number of vectors) we have a linearly independent set which spans, i.e. a basis.

# 3 Direct sum, complementary subspaces

In $\mathbb{R}^3$, say, given some one-dimensional subspace (a line), $(ta, tb, tc), t \in \mathbb{R}$, we may want to find a complementary subspace, i.e. some plane that generates $\mathbb{R}^3$ together with the line. For example, $(x, y, z), ax + by + cz = 0$. Every vector in $\mathbb{R}^3$ can be uniquely expressed as a sum of a vector from each of these subspaces. For another example, consider $V =$ the $x$-axis, $W =$ the $yz$-plane.

In general, given vector subspaces $W_1, W_2 \subset V$, we define the sum $W_1 + W_2 = \{w_1 + w_2 \mid w_1 \in W_1, w_2 \in W_2\}$. If $W_1 + W_2 = V$ and $W_1 \cap W_2 = \{0\}$, then we say $V = W_1 \oplus W_2$, and write that $V$ **is the direct sum of** $W_1$ **and** $W_2$. We make this definition because $v_1 + w_1 = v_2 + w_2 \Leftrightarrow (v_1 - v_2) = (w_1 - w_2) \in V \cap W$, so that $V \cap W = \{0\}$ iff $v_1 + w_1 = v_2 + w_2 \Rightarrow v_1 = v_2, w_1 = w_2$ for any $v_1, v_2 \in V, w_1, w_2 \in W$.

Given any vector spaces $V_0 \subset V$, there always exists a subspace $W \subset V$ such that $V = V_0 \oplus W$. We call such a subspace a **complementary subspace to** $V_0$. To construct one, we can start by finding one vector not in $V_0$, then another vector not in the span of that with $V_0$, etc., just like finding a basis of $V$. In fact, one way to find $W$ is to find a basis of $V_0$ and extend to a basis of $V$; then the basis vectors not in $V_0$ span a complementary subspace.

For finite-dimensional vector spaces $W_1, W_2$, we readily see that $\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2$. More generally, given $W_1, W_2 \subset V$, we see that $\dim(W_1 + W_2) = \dim W_1 + \dim W_2 + \dim W_1 \cap W_2$.

We can extend the definition of direct sums to an arbitrary number of vector spaces: We say that $V = \bigoplus_{i \in I} W_i$ for subspaces $W_i \subset V$ if $W_i \cap \bigoplus_{j \neq i} W_j = \{0\}, \forall i \in I$. Note that being pairwise distinct is not enough! We want this definition to mean that all the $W_i$ are linearly independent. For example, if we have a basis $(v_i)$ of $V$, then $V = \bigoplus \mathrm{Span}(v_i)$. So a basis gives a direct-sum decomposition into one-dimensional subspaces. If we had required the intersections to be zero only pairwise, then for $\mathbb{R}^2$ we could have said that the plane $\mathbb{R}^2$ is the direct sum of any distinct set of $\geq 2$ lines in the plane!

Note finally that the way we defined direct sum, we have $\dim(V) = \sum_{i \in I} \dim V_i$.

# 4 Linear transformations and dual vector spaces

Given any vector spaces $V, W$, it makes sense to consider the space $\mathrm{Hom}(V, W)$ of homomorphisms between $V$ and $W$. This is just the space of homomorphisms of abelian groups which commutes with scalar multiplication: $\phi(av) = a\phi(v), \forall a \in k$. We also call this the space of **linear transformations** between $V$ and $W$. We see that $\mathrm{Hom}(V, W)$ is itself a $k$-vector space. If $V$ and $W$ are finite-dimensional of dimensions $m$ and $n$, respectively, then choosing any basis of $V$ and $W$, we see that $\mathrm{Hom}(V, W)$ is identified with $n \times m$-matrices. So, $\dim(\mathrm{Hom}(V, W)) = nm$. In particular, it is finite-dimensional if $V$ and $W$ are.

The rank that we defined as the number of linearly independent rows or columns now becomes the dimension of the image: if $\phi \in \mathrm{Hom}(V, W)$, then $\mathrm{rank}(\phi) = \dim(\mathrm{im}\ \phi)$, because when we choose any bases of $V$ and $W$ and look at the matrix, the span of the columns is the image so that the dimension is the number of linearly independent columns. This tells

us that rank is a really natural concept, since it doesn't need a choice of basis at all to be defined (provided you don't think of our use of dimension as requiring looking at a basis).

Given a homomorphism $\phi : V \to W$, we can consider the kernel of $\phi$, the vector space $\ker(\phi) = \{v \in V \mid \phi(v) = 0\} \subset V$. It is a vector space because $\phi(av + bw) = 0$ if $\phi(v) = \phi(w) = 0$. We see that the **image** $\text{im } \phi = \phi(V) \subset W$, defined by $\text{im } \phi := \phi(V) := \{w \in W \mid w = \phi(v), \text{ for some } v \in V\}$, is isomorphic to the quotient $V/\ker\phi$. That is, just as in the case of abelian groups, we see that $\phi : V/\ker\phi \to \phi(V)$ is well-defined and is surjective and injective, which means that it has an inverse (clearly as sets, and we see that it must also be linear). So, the dimension satisfies $V - \dim \ker(\phi) = \dim \phi(V) = \text{rank } \phi$.

If we have any map $\phi : V \to W$, then it is injective iff $\ker(V) = 0$, and it is surjective iff $\phi(V) = W$. In the case that $W$ is finite-dimensional, we see that surjectivity is the same as $\text{rank } \phi = \dim W$, since $\dim \phi(V) = \dim W \Rightarrow \phi(V) = W$ by our earlier remarks about dimension. If, moreover, $\dim V = \dim W$, then $\dim V = \dim \ker(\phi) + \dim \phi(V)$, shows that $\dim \ker(\phi) = 0$ ($\phi$ is injective) iff $\dim \phi(V) = \dim V = \dim W$ ($\phi$ is surjective). So in the finite-dimensional case, the map is an isomorphism iff the dimensions of the vector spaces are equal and the map is either injective or surjective (we need not check both).

If we let $W = k$, the one-dimensional vector space, then the space $\text{Hom}(V, k)$ is called the **dual to** $V$ and is denoted by $V^*$. For any dual vector $f \in V^*$ and any vector $v$, we can take $f(v) \in k$, that is we have a pairing $V \times V^* \to k$. By considering this pairing the other way, we can view $V$ as a space of linear functions on $V^*$: that is, fixing a vector $v \in V$ we get a function on $V^*, f \mapsto f(v)$. So this gives an embedding $V \hookrightarrow (V^*)^*$: it's obvious that it is a linear map, and it is injective because $f(v) = f(v')$ for all $f \in V^*$ implies $f(v - v') = 0, \forall f \in V^*$, but we could certainly construct an element of $V^*$ that is not zero on $v - v'$ if $v - v' \neq 0$ by forming a basis of $V$ that begins with $v - v'$ and considering the linear function that is 1 on $v - v'$ and zero on the other vectors in the basis. This requires Zorn's lemma if $V$ is not finite-dimensional. We have constructed a canonical embedding $V \hookrightarrow (V^*)^*$, i.e. an injective homomorphism (=monomorphism).

If $V$ is finite-dimensional, then we can see that $V^*$ has the same dimension as $V$: given any basis $v_1, \ldots, v_n$ of $V$, we can define a **dual basis** of $V^*$ which is just $v^1, \ldots, v^n$ satisfying $v^i(v_j) = \delta_{ij}$ (recall $\delta_{ij} = 1$ if $i = j$ and 0 otherwise). It is clear that the $v^i$ are linearly independent and span $V^*$, so $V^*$ has the same dimension as $V$. Since they are then both isomorphic to $k^n$, that means that $V \cong V^*$, sending a linear combination of the $v_i$ to the corresponding combination of the $v^i$. But this is not a natural isomorphism, because it depended on the choice of the $v_i$: if we changed the $v_i$ slightly, say $v_1 \mapsto v_1' 2v_1$, then $v^1 \mapsto (v')^1 = \frac{1}{2}v^1$ is the dual change of basis, but then our isomorphism sending $v_1'$ to $(v^1)'$ would send $v_1$ to $\frac{1}{4}v^1$ instead of $v^1$, so we get a different isomorphism of $V$ with $V^*$. That is, the isomorphism does not commute with change of basis, which is the main naturality requirement for isomorphisms (more later in the category theory talk).

Thus, $(V^*)^*$ also has the same dimension as $V$ in the finite-dimensional case. So the natural embedding $V \hookrightarrow (V^*)^*$ is an isomorphism by our earlier remarks.

On the other hand, if $V$ is infinite-dimensional, $V \hookrightarrow (V^*)^*$ is not surjective: given a basis of $V$, we can still construct the "dual basis" in $V^*$, and it is still linearly independent,

but it doesn't span all of $V^*$, only the set of functions which are nonzero on only a finite number of basis vectors of $V$. Then, the canonical embedding $V \hookrightarrow (V^*)^*$ sends $V$ to only those vectors which are nonzero on only a finite number of elements of that "dual basis", and thus is not surjective.

In fact, for infinite-dimensional spaces, we can still define dimension by the cardinality of a basis (which one can show is independent of the choice of basis much as we did earlier, using the well-ordering principle) and $V^*$ will have dimension greater than $V$: I believe it has dimension equal to $2^{\dim V}$. So $(V^*)^*$ is also quite a bit larger than $V$.

# 5 Dualizing linear maps

Now that we understand the concept of a dual vector space, it is time to understand how to take duals of linear maps. Given a homomorphism $\phi : V \to W$, we can define its **dual**, the map $\phi^* : W^* \to V^*$, by the definition $\phi^*(f)(v) = f(\phi(v)), \forall v \in V, f \in W^*$. If $V, W$ are finite-dimensional, then the double-dual $(\phi^*)^* : (V^*)^* \to (W^*)^*$ is identified with $\phi$ under the canonical identifications $V \cong (V^*)^*, W \cong (W^*)^*$ (making the canonical identifications, $(\phi^*)^*(v)(f) = [\phi^*(f)](v) = f(\phi(v)), \forall v \in V = (V^*)^*, f \in W^*$, so $(\phi^*)^*(v) = \phi(v)$).

We see that $\mathrm{Hom}(V, W) \xhookrightarrow{*} \mathrm{Hom}(W^*, V^*)$ under the dual, and this is an isomorphism if $V$ and $W$ are finite-dimensional since the two spaces both have the same dimension.

Dualizing commutes with composition: if we have linear maps $T : U \to V, S : V \to W$, then $S \circ T : U \to W$ has dual $(S \circ T)^* = T^* \circ S^*$, which follows immediately from the definition.

Using duals, we can define, for any subspace $V_0 \subset V$, the natural perpendicular space: $V_0^\perp := \ker(i^*) \subset V^*$, where $i^* : V^* \to V_0^*$ is the dual of the natural inclusion $i : V_0 \hookrightarrow V$. We call $i^*$ the **restriction** of $V^*$ to $V_0^*$ because it just restricts a function to $V_0$. That is, $V_0^\perp = \{f \in V^* \mid f(v) = 0, \forall v \in V_0\}$. Another way to define $V_0^\perp$ is by $V_0^\perp := (V/V_0)^*$, which is identified with the given subset of $V^*$ by the map $q^* : (V/V_0)^* \hookrightarrow V^*$ dual to the quotient $q : V \to V/V_0$. Indeed, $q^*$ just sends a linear function on $V/V_0$ to the corresponding function of $V$ which is zero on $V_0$. In the finite-dimensional case, this says that $\dim V_0^\perp = \dim (V/V_0)^* = \dim V/V_0 = \dim V - \dim V_0$.

We saw that the inclusion $i$ is injective and the restriction $i^*$ is surjective (actually, for surjectivity, we need to extend a linear function of $V_0$ to one of $V$, which we can do by extending a basis of $V_0$ to one for $V$ for example); on the other hand, the quotient $q$ is surjective and the map $q^*$ is injective. In other words, we have just seen that the dual of any injective map is surjective, and the dual of any surjective map is injective.

The perpendicular has a nice property: if we have any linear map $T : V \to W$ with dual $T^* : W^* \to V^*$, then $\mathrm{im}\ T^* = (\ker T)^\perp$. Indeed, we can factor the map as $T = i \circ q$, where $q : V \twoheadrightarrow V/V_0$ is the quotient, and $i : V/V_0 \hookrightarrow W$ is injective. Then $\mathrm{im}\ T^* = \mathrm{im}\ (i \circ q)^* = \mathrm{im}\ (q^* \circ i^*) = \mathrm{im}\ q^* = (\ker T)^\perp$. Here we used the previous observation that $i^*$ is surjective.

In the finite-dimensional case, we see from surjectivity of the restriction $i^*$ that $\mathrm{rank}\ i^* = \dim V_0^* = \dim V_0 = \mathrm{rank}\ i$. This property of the rank being invariant under dualization generalizes to any linear map $T : V \to W$:: we see that $T = i \circ q$ for $q : V \twoheadrightarrow V/(\ker T)$, and

$i : V/(\ker T) \hookrightarrow W$. Then rank $T$ = rank $i$ = rank $q$ = rank $i^*$ = rank $q^*$ = rank $T^*$.

Using bases of $V$ and $W$ and their dual bases for $V^*$ and $W^*$, we see that the matrices for $T^*$ and $T$ are related by transposition: we reflect across the diagonal, turning an $n \times m$-matrix into an $m \times n$-matrix. So we are just saying that the rank of a matrix is equal to the rank of its transpose, which is obvious from the definition of rank of a matrix. [Actually, we defined rank of an operator as the dimension of the column space of a corresponding matrix, and the fact that this equals the rank of the dual operator is another way to see that the dimension of the column space is the same as the dimension of the row space.]

# 6 Inner products and normal transformations

We have seen how if $V$ is a finite-dimensional vector space, then there are noncanonical isomorphisms between $V$ and its dual space $V^*$, for example by choosing a basis and then taking the dual basis. We saw that this identification depends on the basis. But it doesn't depend completely on the basis: if we swapped the order of two of the basis vectors (or any other permutation of them for that matter), one sees that the map is the same. So the time has come to characterize what an identification of $V$ with $V^*$ really is.

Let's start with a homomorphism $T : V \to V^*$. This assigns each $v \in V$ with a linear function $f_v = T(v) \in V^*$. This is evidently the same thing as a map $B : V \times V \to k, B(v, w) := f_v(w) = T(v)(w)$ which is bilinear: $B(a_1v_1+a_2v_2, w) = a_1B(v_1, w)+a_2B(v_2, w)$ and $B(v, a_1w_1 + a_2w_2) = a_1B(v, w_1) + a_2B(v, w_2)$. Let's drop the $B$ now and just think of it as a pairing, $(v, w) := B(v, w) \in k$ for any $v, w \in V$. Such a bilinear pairing $(,)$ is called a **bilinear form**.

For the homomorphism to be an isomorphism, it is necessary and sufficient for $V$ to be finite-dimensional and for the homomorphism to be injective: that is, $f_v \neq 0$ for any nonzero choice of $v$. In terms of the bilinear form, this says that for every $v \neq 0$, there exists a $w$ such that $B(v, w) \neq 0$. An equivalent condition is that the homomorphism is surjective, or that the dual $T^* : V \to V^*$ is injective, which means that for every $w \neq 0$, there exists $v$ such that $B(v, w) \neq 0$.

We define a bilinear form satisfying these properties to be **nondegenerate**: (i) $\forall v \in V \setminus \{0\}$, $\exists w \in V$ such that $B(v, w) \neq 0$; (ii) $\forall w \in V \setminus \{0\}$, $\exists v \in V$ such that $B(v, w) \neq 0$. We just proved that the two conditions are equivalent if $V$ is finite dimensional, and are both equivalent to the map $V \to V^*$ being an isomorphism.

It would be nice if the bilinear form were symmetric: $B(v, w) = B(w, v), \forall v, w \in V$. We see that this is equivalent, in the finite-dimensional case, to saying that $T$ and $T^*$ induce the same bilinear form.

If we restrict to the case $k = \mathbb{R}$, a nice property for the bilinear form to have is $(v, v) > 0, \forall v \in \mathbb{R} \setminus \{0\}$, which we call **positive-definiteness**. Positive-definiteness in particular implies nondegeneracy: $B(v, v) > 0$ gives both properties (i) and (ii) of nondegeneracy when we take $v = w$.

All of these properties: positive-definiteness, symmetric, and bilinear, are all satisfied by the dot product $(a_1, \ldots, a_n) \cdot (b_1, \ldots, b_n) = a_1b_1 + \ldots + a_nb_n$; in fact, every positive-definite

symmetric bilinear form over a finite-dimensional real vector space is equivalent to $\mathbb{R}^n$ with the dot product.

**Definition 6.1.** Over a real vector space $V$, an **inner product** is a positive-definite, symmetric, bilinear form. That is, a map $B : V \times V \to k$ that is linear in each component ($B(a_1v_1 + a_2v_2, w) = a_1B(v_1, w) + a_2B(v_2, w)$ and similarly for the other component), is symmetric ($B(v, w) = B(w, v)$ for all $v, w \in V$), and positive definite ($B(v, v) > 0$ for any $v \neq 0$). An **inner product space** is a vector space equipped with an inner product.

Over complex vector spaces, however, we can't get a positive-definite bilinear form, since $(v, v) > 0$ implies $(iv, iv) < 0$ for $i = \sqrt{-1}$. We can have nondegenerate bilinear forms, and in fact all such forms are equivalent for a complex vector space (we can get from one to another by changing basis).

To fix this problem, one can generalize the notion of bilinear form so that, in the case of $\mathbb{C}^n$, our form is $(v, w) \mapsto v \cdot \bar{w}$, where $\cdot$ is the dot product. This however is no longer bilinear but **sesquilinear**: $(v, aw) = a(v, w)$ but $(av, w) = \bar{a}(v, w)$, while still being real-linear. That is, a sesquilinear form is conjugate-linear in the first component and complex-linear in the second component.

With this definition, we can define a **Hermitian inner product** to be a sesquilinear form satisfying $(v, v) > 0, v \neq 0$ (positive-definiteness), and $(v, w) = \overline{(w, v)}, \forall v, w \in V$ (conjugate-symmetry).

We summarize with the

**Definition 6.2.** For a complex vector space $V$, a **(Hermitian) inner product** is a sesquilinear form ($B(a_1v_1 + a_2v_2, w) = \bar{a}_1B(v_1, w) + \bar{a}_2B(v_2, w)$ and $B(v, a_1w_1 + a_2w_2) = a_1B(v, w_1) + a_2B(v, w_2)$), which is conjugate-symmetric ($B(v, w) = \overline{B(w, v)}$), and positive-definite ($B(v, v) > 0$ for any $v \neq 0$.) Again, an **inner product space** is a vector space equipped with an inner product.

In particular, any inner product over a finite-dimensional real vector space gives an isomorphism $V \cong V^*$, and any Hermitian inner product over a finite-dimensional complex vector space gives a conjugate-linear bijection $V \to V^*$. This allows us in both cases to identify $V$ with $V^*$; note that if we then consider the inner product to be on $V^*$, the resulting identification $V^* \to (V^*)^* = V$ is the inverse map.

Recall again that this identification is given by $v \mapsto f_v$, $f_v(w) = (v, w)$. For any endomorphism $T : V \to V$, we can consider the map $f_{T^\dagger v}V \to k$ given by $w \mapsto (v, Tw)$. In particular, $f_{T^\dagger v} \in V^*$, so $f_{T^\dagger v} = f_w$ for some unique $w \in V$, which we call (you guessed it) $T^\dagger v$. We can define a map $T^\dagger : V \to V$ which sends each $v$ to $T^\dagger v$ in this sense. It is clearly a linear map, which we call the **adjoint of** $T$. That is,

**Definition 6.3.** Given an endomorphism $T : V \to V$ of a vector space $V$ with inner product $(,)$, the **adjoint**, $T^\dagger : V \to V$ is defined to be the unique map such that $(v, Tw) = (T^\dagger v, w)$ for all $v, w \in V$.

By definition, we see that $(T^\dagger)^\dagger = T$.

What is the relation between $T^\dagger$ and $T^*$? We can see from the definition that viewing the inner product as an identification $V \to V^*$, then $T^\dagger$ is the map $V \to V$ we get by applying this identification to $T^* : V^* \to V^*$. That is why the adjoint is sometimes written as $T^*$, but here we try to separate the concept of dual from adjoint.

We see just as with duals that $(ST)^\dagger = T^\dagger S^\dagger$.

What does an identification $V \cong V^*$ do to the perpendicular space $W^\perp \subset V^*$ to a subspace $W \subset V$? We see that $W^\perp$ becomes a subspace of $V$ of dimension $\dim V - \dim W$. This is true for any linear (or conjugate-linear) identification arising from a nondegenerate bilinear (or sesquilinear) form. In the case of an inner product, we see that $V = W \oplus W^\perp$, i.e. $W^\perp$ is a complement to $W$, because $W \cap W^\perp = \{0\}$ for a positive-definite space (in general, all vectors in this intersection satisfy $(v, v) = 0$: such a vector space is called **isotropic**.) We thus call $W^\perp \subset V$ the **orthogonal complement to** $W$.

Now, let us discuss all this in terms of bases. For any basis $S \subset V$, we call the **standard inner product** the unique inner product such that $(v, w) = \delta_{v,w}$ for any $v, w \in S$: that is, the inner product of two basis vectors is 1 if they are the same and 0 otherwise. From the other point of view, given an inner-product space, a basis of this type is called an **orthonormal basis**; a set of elements with this inner product is called an **orthonormal set** if it doesn't necessarily span. Orthonormal bases exist for any finite-dimensional inner product space, which can be found inductively just as we found a basis: given any orthonormal set $S \subset V$, if it doesn't span $V$ we just pick any vector in $S^\perp \subset V$ to increase the basis. Such a vector exists because we can find a nonzero element of $V^*$ which vanishes on $\mathrm{Span}(S) \neq V$, which must be identified with some vector of $V$ by the inner product.

For infinite-dimensional inner product spaces, orthonormal bases do not always exist! (However, there is always an orthonormal set the closure of whose span is all of $V$, where we use the topology given from the metric $d(v, w) = (v - w, v - w)$; more on this in the Analysis II talk.)

In the finite-dimensional case, we just proved that any inner-product space is isomorphic to $k^n$ for some $n$, with the standard inner product (recall $k = \mathbb{R}$ or $\mathbb{C}$ as we defined inner product). Let us restrict to this case for now, generalizations to Hilbert spaces coming in another talk.

In terms of an orthonormal basis of $V$, we see that $T^\dagger$ is just the conjugate-transpose of $T$ (which includes just transpose in the real case). The "conjugate" part can be viewed as coming from the fact that the identification $V \to V^*$ coming from the inner-product sends $c$ times a basis vector to $\bar{c}$ times the dual vector, for $c \in \mathbb{C}$. **Note that this is only true for an orthonormal basis!**

With respect to any inner product (or any bilinear or sesquilinear form) it makes sense to consider the space of operators $T$ such that $(Tv, Tw) = (v, w)$ for all $v, w \in V$: that is, operators preserving the form. For a real inner product (or bilinear form), such operators are called **orthogonal**, and for a complex inner product, such operators are called **unitary**. Using adjoint, this just says that $T^\dagger T = \mathrm{Id}$, the identity. In terms of matrices, we are saying that the matrix and its conjugate-transpose are inverses. If we look at what this means when

multiplying the matrix, it says that the standard inner product of any two distinct rows (or columns) is zero, and the standard inner product of a row (or column) with itself is one. In other words, the rows (or columns) form an **orthonormal basis** of the vector space $V$.

Another special type of transformation is one that equals its own adjoint: $T = T^\dagger$; in terms of an orthonormal basis, this says it equals its own conjugate transpose. Such matrices are called "Hermitian" in the complex case, or "symmetric" in the real case (the symmetry is across the diagonal!).

All of these transformations, Hermitian, symmetric, unitary, orthogonal, you may recall from linear algebra obey a certain **spectral theorem**: that says that they are completely diagonalizable over $\mathbb{C}$; the symmetric matrices are additionally diagonalizable over $\mathbb{R}$. More precisely, these are precisely the matrices for which there is an orthonormal basis of **eigenvectors**, that is, an orthonormal basis in which the transformation is just a diagonal matrix. This theorem is the subject of the next section. We just end with the statements:

**Definition 6.4.** A **normal operator** $T : V \to V$ on a vector space with an inner product is an operator such that $TT^\dagger = T^\dagger T$. In particular, this includes Hermitian, symmetric, unitary, or orthogonal transformations.

**Theorem 6.5.** *(Complex spectral theorem). Let $V$ be any finite-dimensional complex inner product space. $T : V \to V$ is a normal operator iff there exists an orthonormal eigenbasis of $V$, i.e. an orthonormal basis in which $T$ is diagonal.*

**Theorem 6.6.** *(Real spectral theorem). Let $V$ be any finite-dimensional real inner product space. $T : V \to V$ is a symmetric operator iff there exists an orthonormal eigenbasis of $V$.*

# 7 Proof of the spectral theorem

The condition $TT^\dagger = T^\dagger T$ is exactly the same as the condition $(Tv, Tv) = (T^\dagger v, T^\dagger v), \forall v \in V$. So, this means that $Tv = 0 \Leftrightarrow T^\dagger v = 0$, using positive-definiteness. This shows that $T^2 v = 0$ implies $T^\dagger T v = 0$, which then implies $(Tv, Tv) = 0$ so $Tv = 0$. Inductively we then see that $T^k v = 0 \Leftrightarrow Tv = 0 \Leftrightarrow T^\dagger v = 0$. We have proved

**Lemma 7.1.** *If $T$ is a normal operator, then $T^k v = 0 \Leftrightarrow Tv = 0 \Leftrightarrow T^\dagger v = 0$. That is, looking at the vector space $V_0 = \{v \in V \mid T^j v = 0, \text{ for some } j \geq 1\}$, we have $V_0 = ker\ T = ker\ T^\dagger$.*

**Corollary 7.2.** *For any $\lambda \in k$, we have $V_\lambda := ker\ (T - \lambda Id)^j = ker\ (T - \lambda Id) = ker\ (T^\dagger - \bar{\lambda} Id)$ for any (eigenvalue) $\lambda$.*

*Proof.* We see that $T - \lambda \mathrm{Id}$ is still normal since $(T - \lambda \mathrm{Id})^\dagger = T^\dagger - \bar{\lambda}\mathrm{Id}$, and $T, T^\dagger$, and Id all commute. □

The space $V_\lambda$ appearing above is called the **generalized eigenspace** of $\lambda$, with the **eigenspace of** $\lambda$ being the case $j = 1$, i.e. vectors satisfying $Tv = \lambda v$. It is a nice fact that if $k$ is algebraically closed, we can decompose the vector space into generalized eigenspaces (for an arbitrary operator, without using an inner product space):

**Theorem 7.3.** *If $V$ is a finite-dimensional space over an algebraically closed field $k$ (i.e. one where every polynomial has a root, such as $k = \mathbb{C}$), then $V$ can be decomposed into a direct sum of generalized eigenspaces for a finite number of eigenvalues. That is,*

$$V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_l}, \tag{7.1}$$

*where $V_{\lambda_i}$ is the generalized eigenspace with eigenvalue $\lambda_i$.*

**Corollary 7.4.** *The complex spectral theorem.*

*Proof.* Since $\mathbb{C}$ is algebraically closed (the fundamental theorem of algebra), $V$ decomposes into a direct sum of generalized eigenspaces. Corollary 7.2 just says that these are all eigenspaces of $T$ and $T^\dagger$, with the eigenvalue under $T^\dagger$ of $V_\lambda$ being $\bar{\lambda}$. Also, the eigenspaces for different eigenvalues are orthogonal: if $v \in V_\lambda, w \in V_{\lambda'}$, then $(v, Tw) = (v, \lambda' w) = \lambda'(v, w) = (T^\dagger v, w) = (\bar{\lambda} v, w) = \lambda(v, w)$, so $\lambda \neq \lambda'$ implies $(v, w) = 0$. So, if we pick orthonormal bases of each $V_\lambda$, then we get an orthonormal basis of all of $V$; and in this basis, $T$ is diagonal with the eigenvalues on the diagonal (with multiplicity of $\lambda$ equalling dim $V_\lambda$).

For the opposite direction, if $T$ is a diagonal matrix in an orthonormal basis of $V$, then $T^\dagger$ is just the complex conjugate in this basis, so $TT^\dagger = T^\dagger T$, with diagonal equalling the absolute-value squared of each eigenvalue with multiplicity. Note that the use of orthonormal here was to see that $T^\dagger$ is the conjugate-transpose (which is not in general true for a non-orthonormal basis). □

**Corollary 7.5.** *The real spectral theorem.*

*Proof.* If $T$ is symmetric (or even Hermitian in the complex case), then $v \in V_\lambda \setminus \{0\}$ shows that $\lambda(v, v) = (v, Tv) = (T^\dagger v, v) = (Tv, v) = \bar{\lambda}(v, v)$, so that $\lambda = \bar{\lambda}$, or $\lambda \in \mathbb{R}$. Viewing $T$ as a linear transformation on $\mathbb{C}^n$ by choosing a basis of the real vector space $V$ and replacing $T$ with its matrix, we see that a (possibly complex) orthonormal change-of-basis will diagonalize $T$. But this change-of-basis can be taken to be real, since if $\lambda$ is an eigenvalue of $T$ viewed as a complex matrix, it is real from the above, and then dim $V_\lambda =$ dim $V -$rank $(T - \lambda \mathrm{Id})$ (since the generalized eigenspace is actually the eigenspace, or $j = 1$ in the definition of generalized eigenspace since $T$ is normal). But rank can be calculated over the reals, since it involves linear dependence of a matrix with real coefficients. So viewing $T$ as a real transformation again, $V_\lambda \subset V$ has the same dimension we found considering it to be a complex matrix. So, as a real vector space, $V$ has a generalized-eigenspace decomposition for $T$ with the same eigenvalues and multiplicities that we found considering $T$ to be a complex matrix. Corollary 7.2 again shows us this is an eigenspace decomposition, and again (over the real numbers) we can get an orthonormal basis of eigenvectors of $T$ by choosing any orthonormal basis of each eigenspace.

Alternatively, we could have shown that the complex eigenspace decomposition actually was given by a real change-of-basis since taking the complex conjugate of the coefficients of a real eigenvector gives another eigenvector with the same eigenvalue, so each eigenspace is invariant under complex conjugation and is hence given by a real change-of-basis.

Anyway, the converse follows just as before: any diagonal matrix (over the reals) is invariant under transposition and is hence a symmetric matrix. □

*Proof.* (Proof of Theorem 7.3) Note first that this follows from Jordan decomposition, which will be done using more general machinery in a later talk. We prove the theorem in a more elementary way. Since the space of endomorphisms (or matrices if we pick a basis) is $(\dim V)^2$-dimensional, the matrices $\mathrm{Id}, T, T^2, \ldots, T^{(\dim V)^2-1}$ must be linearly dependent, so there is a polynomial of degree $\leq (\dim V)^2 - 1$ that $T$ satisfies as an endomorphism. (Actually, there is a polynomial of degree $\leq n$, but we'll get to that later.) If we are over an algebraically closed field, such as $\mathbb{C}$, then the polynomial factors into linear factors.

Now, we have something like $(T-a_1)(T-a_2)\cdots(T-a_m) = 0$. So, not all of the $T-a_i$ can be invertible. So there is some $a_i$ and some vector $v$ such that $(T - a_i\mathrm{Id})v = 0$. This $v$ is an **eigenvector of eigenvalue** $a_i$. We have proved that **every endomorphism of a finite-dimensional vector space over an algebraically closed field has an eigenvalue and eigenvector**.

This result does not hold when any of the assumptions are false! For example, a rotation of $\mathbb{R}^2$ has no eigenvector, and the shift-right operator on the space of sequences of elements of $k$, indexed by integers, with finitely many nonzero entries has no eigenvector (even over an algebraically closed field $k$).

So, if $T$ is any endomorphism, and $\lambda$ any eigenvalue (which means that there exists $v \neq 0$ such that $Tv = \lambda v$), then we can define $V_\lambda \neq 0$ as the generalized eigenspace as above. We can find $j \geq 1$ such that $(T - \lambda)^j V_\lambda = \{0\}$; for instance, taking $j = \dim V_\lambda$ works, since if at any point $\mathrm{rank}(T - \lambda)^j = \mathrm{rank}(T - \lambda)^{j+1}$, the rank will never change as $j$ increases, so it had better be 0 since we defined $V_\lambda$ so that everything is eventually killed!

Now, $\ker[(T - \lambda)^j] = V_\lambda$ since $\subset$ is true by definition and $\supset$ is what we just deduced. So we see that $V = V_\lambda \oplus \mathrm{im}\ (T - \lambda\mathrm{Id})^j$, as the two vector spaces do not intersect. Since $T(\mathrm{im}\ (T-\lambda\mathrm{Id})^j) = \mathrm{im}\ (T-\lambda\mathrm{Id})^j$, we can restrict to $\mathrm{im}\ (T-\lambda\mathrm{Id})^j$, which has strictly smaller dimension, and apply induction to find that

$$V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_l}, \tag{7.2}$$

for some finite list of eigenvalues $\lambda_1, \ldots, \lambda_l$. □

*Remark* 7.6. We see from the construction that $T$ actually satisfies the polynomial

$$(x - \lambda_1)^{\dim V_{\lambda_1}} \cdots (x - \lambda_l)^{\dim V_{\lambda_l}} \tag{7.3}$$

of degree $\dim V$, which is called the **characteristic polynomial**. This can also be defined as $\det(x\mathrm{Id} - T)$; so if we have a non-algebraically closed field, we see that the polynomial given from the above decomposition taken over the algebraic closure actually lives in the field we started with (which makes sense from the point of view of Galois theory since all the Galois automorphisms of the algebraic closure fixing the ground field will fix the set and multiplicity of eigenvalues).

To prove more generally that $T$ satisfies this polynomial $f(x) := \det(x\mathrm{Id} - T)$, it is easiest to multiply the matrix $(xI - T)$ by its matrix of cofactors $C$ (having in the $i, j$-th position the determinant of the submatrix obtained by deleting the $i$-th row and the $j$-th column) and we see that $C(xI - T) = \det(xI - T)\mathrm{Id}$, so plugging in $T$ we get $0 = f(T)$.

This is known as the **Cayley-Hamilton theorem** and will be generalized in second-quarter algebra and used to prove fundamental results such as "Nakayama's Lemma".

Written with matrices, the spectral theorem looks like:

**Theorem 7.7.** *If $T$ is any $n \times n$-matrix such that $TT^\dagger = T^\dagger T$, then there exists a unitary matrix $U$ such that $T = UDU^{-1}$ where $D$ is a diagonal matrix. If $T$ is in fact symmetric with real coefficients, we can take $U$ to be an orthogonal matrix (equivalently, a unitary matrix with real coefficients).*

*Remark* 7.8. Working in the real case, it would be nice to also have a spectral decomposition of normal operators that aren't symmetric (e.g. to include orthogonal transformations). We may not even get a single real eigenvector in this case; for example, rotations in $\mathbb{R}^2$ that are not 180° or 360°. However, since the matrix is real, we see that for every eigenvalue $\lambda \in \mathbb{C}$, $\bar{\lambda}$ is also an eigenvalue, and $\dim(V_\lambda) = \dim(V_{\bar{\lambda}})$, since we can conjugate all of our work without changing the original matrix (or see Remark 7.6 regarding the characteristic polynomial). In fact, by pairing up elements with conjugate eigenvalues in an orthonormal eigenbasis in which the normal matrix is diagonal, we can write any normal matrix $N$ as $N = O_1 N_2 (O_1)^{-1}$, where $O_1$ is (real) orthogonal, and $N_2$ is a direct sum of one-by-one real matrices and two-by-two orthogonal matrices (actually, with positive determinant: rotations and scalings, or multiplication by a complex number viewing $\mathbb{R}^2$ as the complex plane: this complex number is an eigenvalue). That is, $N_2$ is block-diagonal with these pieces.

More generally, the fact that $\mathbb{C}$ is given from $\mathbb{R}$ just by adjoining $\sqrt{-1}$, or equivalently that any polynomial splits into linear and quadratic factors (applied to the characteristic polynomial), shows that any real matrix is equivalent to a block upper-triangular one with one-by-one or two-by-two blocks along the diagonal, the latter being orthogonal matrices with positive determinant. Using Jordan decomposition we could restrict the above-diagonal to just have some ones (between rows with the same one-by-one block) and some two-by-two identity blocks (between pairs of rows with the same two-by-two orthogonal block). Being normal means not only that there is nothing on the superdiagonal, but moreover that the change-of-basis was actually orthogonal.

There is uniqueness of the diagonal form (in the normal case) and of the Jordan form (in the general complex or even real case) up to permutation of blocks (which include groups of diagonal blocks that are linked by superdiagonal identity matrices). In the real case to get uniqueness we need the two-by-two blocks to be scalings by a positive number composed with rotations counterclockwise by 0° to 180°, not inclusive. We can generalize this type of Jordan form to give a Jordan form of an arbitrary matrix over any nonalgebraically closed field, where we allow the diagonal blocks to be of arbitrary size, but expressing (in a unique way) an irreducible polynomial which is a factor of the characteristic polynomial. The upper-triangular part would be some identity matrices in-between identical diagonal blocks. The correct way to understand these Jordan forms for nonalgebraically closed fields $k$ is to think of each $n \times n$-matrix in a diagonal or superdiagonal block as a single number over an algebraic extension field of $k$ of degree $n$, which adjoins a single root of the corresponding degree-$n$ irreducible factor of the characteristic polynomial.

# 8 Determinants

We finally move on to determinants (having already defined the characteristic polynomial!) Over the real numbers, the determinant of a two-by-two matrix gives the signed area of the parallelogram that the two vectors make: the sign is positive or negative according to the right-hand rule. In general, the determinant of an $n \times n$-matrix over the reals is $\pm$ the volume of the parallelopiped that the $n$ vectors make, the sign given by a certain orientation one can define from the order of the vectors.

For two-by-two matrices, we have the familiar formula

$$\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc, \tag{8.1}$$

which one could readily prove has the area property described above.

In general, determinant is a way of telling whether a matrix is invertible or not: if the determinant is nonzero or not. In the case of $\mathbb{R}^n$, if the determinant is nonzero, then the vectors must span, since they make a parallelopiped with nonzero volume; on the other hand, the volume is zero just in the case when the span lies in some hyperplane.

Let's be more precise: let $V^m$ be the direct product of $m$ copies of $V$::

**Definition 8.1.** Given any vector space $V$, we define the vector space $\Lambda^m V^* := \{m$-linear maps $A : V^m \to k$ such that $A(v_1, v_2, \ldots, v_m) = 0$ if any $v_i = v_j, i \neq j$.

In particular, if we consider $w_i = v_i + v_j = w_j$, then we see that $A(v_1, v_2, \ldots, v_{i-1}, w_i, v_{i+1}, \ldots, v_{j-1}, w_j$ $0$ implies that $A(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{j+1}, \ldots, v_m) = -A(v_1, \ldots, v_n)$, so $\Lambda^m V^*$ includes only completely skew-symmetric, $m$-linear maps from collections of $m$ vectors to the field $k$. Actually, this is an equivalent definition, except if the characteristic of $k$ is 2, when we need the alternating definition (i.e. that $A(v_1, \ldots, v_n) = 0$ if they are not all distinct.)

In fact, if we change the order of the $v_i$'s in any way, we get $\pm$ the result we had before by applying $A$, depending on whether we made an odd or even number of swaps (one can prove that the parity of the number of swaps is invariant, and is equal to the parity of the number of pairs $i, j$ such that $v_i, v_j$ end up in the wrong order). If we had applied permutation $\sigma \in S_n$, we define $\text{sign}(\sigma) = \pm 1$ to be this sign (just as in group theory).

Moreover, if $v_m = a_1 v_1 + \ldots a_{m-1} v_{m-1}$ is in the span of the first $m-1$ vectors, then $A(v_1, \ldots, v_m) = a_1(v_1, \ldots, v_{m-1}, v_1) + a_2(v_1, v_2, \ldots, v_{m-1}, v_2) + \ldots + a_{m-1} A(v_1, \ldots, v_{m-1}, v_{m-1}) =$ $0$. So already we see that we can detect if a set of $m$ vectors is linearly dependent or not by seeing whether all $A \in \Lambda^m V^*$ kill it or not (we haven't actually shown that a linearly independent set has an $A$ that is nonzero on it, but we will see this.)

Now, for any operator $T : V \to V$, we can define an operator $(T^*)^{\wedge m} : \Lambda^m V^* \to \Lambda^m V^*$ by $(T^*)^{\wedge m}(A)(v_1, \ldots, v_m) := A(Tv_1, \ldots, Tv_m)$. One can see that $(T^*)^{\wedge m}$ must be a linear tranformation because $T$ is, using $m$-linearity and skew-symmetry of $A$.

We see that the set $\Lambda^n V^*$ must be one-dimensional, because if we have a basis $(v_1, \ldots, v_n)$, then $A(w_1, \ldots, w_n)$ is determined by $A(v_1, \ldots, v_n)$ for any $w_i$, using the $n$-linearity, alternating condition, and skew-symmetry. So we see that $(T^*)^{\wedge n}$ must be multiplication by some

scalar, which we call the **determinant** of $T$. In particular, it is clear that this scalar is 1 if $T$ is the identity.

There is one problem with the above: we didn't really prove that $\Lambda^n V^*$ is one-dimensional because we didn't show that $\Lambda^m V^* \neq \{0\}$ for any $m$! To do this, I'll just give the determinant formula. Take a linearly independent set $v_1, \ldots, v_m \in V$. Complete this to a basis $v_1, \ldots, v_n$ of $V$. Then we can define an $A$ which has the property $A(v_1, \ldots, v_m) = 1$, but $A(v_j, w_1, \ldots, w_{m-1}) = 0$ for any $j > m$ and any $w_i$'s. We can do this by writing vectors in the basis $(v_i)$, and defining, for $w_i = a_{i1}v_1 + \ldots a_{in}v_n$,

$$A(w_1, \ldots, w_m) = \sum_{\sigma \in S_m} \text{sign}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{m\sigma(m)}, \tag{8.2}$$

which can be verified to have the desired property. In particular, $A(v_1, \ldots, v_m) = 1 \neq 0$. So we can detect linear independence with the set $\Lambda^m V^*$.

Setting $m = n$, the above formula (8.2) gives the **determinant** of the given $n \times n$-matrix; i.e. this is how we define the determinant for matrices.

One can actually see that, for a basis $v_1, \ldots, v_n$ of $V$, the $A$'s defined for a choice of $1 \leq i_1 < \ldots < i_m \leq n$ by $A(v_{i_1}, \ldots, v_{i_m}) = 1$, $A(v_j, w_1, \ldots, w_{m-1}) = 0$ for $j \notin \{i_1, \ldots, i_m\}, \forall w_i \in V$, form a basis of $\Lambda^m V^*$ and hence this latter space has dimension $\binom{n}{m}$. More on this will be in the multilinear algebra course!!

*Remark* 8.2. (cf. Remark 7.6) To see that the characteristic polynomial we defined earlier is the same as $\det(x\text{Id} - T)$, note first that $x\text{Id} - T$ has the same generalized eigenspace decomposition as $T$, with each eigenvalue $\lambda$ of $T$ becoming the eigenvalue $x - \lambda$ with the same multiplicity (and generalized eigenspace). Then, note that determinant is the same as the product of the determinants when restricted to each generalized eigenspace, because we can pick a basis of $V$ by taking bases of each eigenspace (this generalizes to any decomposition of $V$ into spaces left invariant by $T$). We can then prove inductively that the determinant of $T$ on $V_\lambda$ is $\lambda^{\dim V_\lambda}$: we induct on the minimum $j \geq 1$ such that $V_\lambda = \ker(T - \lambda)^j$. The result is obvious if $j = 1$, since $T|_{V_\lambda} = \lambda\text{Id}$ in this case. Otherwise, we see that $V_\lambda \cong \ker(T - \lambda) \oplus V/[\ker(T - \lambda)]$. We see that the determinant is equal to the product of the determinant of $T$ on each of these; more generally, if $W \subset V$ is any $T$-invariant subspace, $TW \subset W$, then picking any basis of $W$ and extending to a basis of $V$ shows that $\det T = \det T|_W \det T_{V/W}$, since what $T$ does to the part of the basis from $W$ is the first term, and what $T$ does to the rest of the basis modulo things from $W$ is the second term (i.e. in $A(v_1, \ldots, v_n)$, we can consider the $v_i, i > m$ to be taken modulo $\text{Span}(v_1, \ldots, v_m)$). But $(T - \lambda^{j+1})V_\lambda = \{0\}$ implies $(T - \lambda^j)|_{V/\ker(T-\lambda)} = 0$, so the induction hypothesis shows that $\det T = \lambda^{\dim \ker(T-\lambda)} \lambda^{\dim V/\ker(T-\lambda)} = \lambda^{\dim V}$, as desired.

# 9 The classical linear groups

Now that we have defined all of the types of matrices, we can define the classical matrix groups. For any vector space $V$ over a field $k$, $\text{GL}(V)$ is the group of invertible endomorphisms

of $V$. For $V = k^n$ we write this as $\mathrm{GL}(n,k)$. We define $SL(V)$ to be the group of matrices of determinant one (one can verify that $\det T \det S = \det(TS)$ by our definition, so this makes a group), which is called $SL(n,k)$ for $k^n$. For any vector space $V$ with a bilinear form $B$, we let $O(B,V)$ be the set of transformations preserving the form in the sense that $B(Tv,Tw) = B(v,w)$ for any $v,w \in V$ (orthogonal group). Also, $SO(B,V) = O(B,V) \cap SL(V)$ is the special orthogonal group. For $k^n$ with the dot product, we write these as $O(n,k)$ and $SO(n,k)$. For the case $k = \mathbb{R}^n$, we may consider something like the dot product but with a basis $v_1, \ldots, v_n$ where $B(v_i,v_i) = 1$ if $i \leq m$ and $B(v_i,v_i) = -1$ if $i > m$, then we write this group as $O(n,m)$ (we usually only consider such things over $\mathbb{R}$.) We also have $SO(n,m)$. Similarly, the group of unitary matrices over a complex vector space $V$ is $U(V)$, and $SU(V) = U(V) \cap SL(V)$. For $V = \mathbb{C}^n$ we denote this by $U(n), SU(n)$ (unitary matrices are usually only defined over complex numbers).

Finally, the symplectic group $Sp(V)$ is the group of matrices preserving a symplectic form in the same sense as given above for a bilinear form. Only we didn't define symplectic forms! These forms are just nondegenerate bilinear forms which are **alternating** instead of symmetric: this means $B(v,v) = 0$ for all $v$. In particular, this shows $B(v,w) = -B(w,v)$; in fact, this skew-symmetry is equivalent to the alternating condition except for fields of characteristic two, that is where $1+1=0$: in those cases we need to say $B(v,v) = 0 \forall v \in V$ because the skew-symmetry condition only makes the form symmetric (and the general phenomena one sees with skew-symmetry generalizes better to the alternating condition, not the symmetry condition, in this case). Similarly, for $k^{2n}$ with the standard symplectic form $B((a_1,\ldots,a_{2n}),(b_1,\ldots,b_{2n})) = \sum_{i=1}^{n}(a_i b_{n+i} - a_{n+i} b_i)$, we call this $Sp(2n,k)$. As you probably guessed, one can show that any symplectic form on a finite-dimensional vector space $V$ is equivalent (under change-of-basis) to this one (for $n = \frac{1}{2}\dim V$)! (There are no symplectic forms on odd-dimensional vector spaces.)