Linear Algebra in brief

Adam A. Allan, Ph.D.

February 12, 2026

Throughout this chapter k will denote a field.

In this chapter we quickly develop the main results for linear transformations defined on a finite dimensional vector space. Our treatment will differ from that of a first course in linear algebra in that we will try to avoid choosing a basis for the vector space whenever possible. Our focus will be on the decomposition theorems for linear transformations and on non-degenerate bilinear forms.

We also assume knowledge of the basic theory of equations over k and of matrices, including the row reduction algorithm and matrix multiplication. The definition of the determinant is provided in the appendix on multilinear algebra and its main properties are developed there.

0.1 Definitions

A **vector space** V over the field k is an additive group that comes equipped with a function $k \times V \to V$ that respects the operations of addition and multiplication in V and k. We call elements of V **vectors** and write them as u, v, w, etc, and we call elements of k **scalars** and write them as a, b, c, etc. We refer to the function $k \times V \to V$ as **scalar multiplication** and write av as the image of (a, v) under this function. In more concrete terms, to say that scalar multiplication respects the operations of addition and multiplication means that for all v, w in V and a, b in k we have the equalities

$$\begin{aligned} a(v+w) &= av + aw \\ (a+b)v &= av + bv \\ (ab)v &= a(bv) \end{aligned}$$

We also assume that $1v = v$ for all v in V. A **linear transformation** is a function $T : V \to W$ from one vector space to another that satisfies

$$\begin{aligned} T(v_1 + v_2) &= T(v_1) + T(v_2) \\ T(av) &= aT(v) \end{aligned}$$

for all v_1, v_2, v in V and a in k. If T is a bijection then we call T an **isomorphism**, and in this case T^{-1} is itself a linear transformation (as is easy to check). More generally we write $\mathrm{Hom}_k(V, W)$ (or $\mathrm{Hom}(V, W)$ for short) as the set of all linear transformations $T : V \to W$. Importantly, $\mathrm{Hom}(V, W)$ is itself a vector space over k with addition and scalar multiplication given by

$$\begin{aligned}(T_1 + T_2)(v) &= T_1(v) + T_2(v)\\ (aT)(v) &= aT(v)\end{aligned}$$

In case $W = V$ we call T a **linear operator** for emphasis. Moreover, if $S : U \to V$ and $T : V \to W$ are linear transformations then so is $T \circ S : U \to W$. Finally, we write $\mathrm{Hom}(V, V)$ as $E(V)$ and note that $E(V)$ is a k-algebra with identity I defined by $Iv = v$ for all v in V.

So linear transformations are one major object of study, and bilinear forms are the other. A **bilinear form** on V is a map $B : V \times V \to k$ such that for all u, v, w in V and a in k we have the equalities

$$\begin{aligned}B(u + v, w) &= B(u, w) + B(v, w)\\ B(u, v + w) &= B(u, v) + B(u, w)\\ B(au, v) &= aB(u, v) = B(u, av)\end{aligned}$$

Now suppose B is a fixed bilinear form and define G as the set of all invertible linear transformations $T : V \to V$ satisfying $\mathcal{B}(Tv_1, Tv_2) = B(v_1, v_2)$ for all v_1, v_2 in V. We see that G is a group. Through appropriate choice of B we will be able to create the classical families of matrix groups.

0.2 Bases

An example of a vector space is the set k^n consisting of all **column vectors**

$$v = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

with each a_i in k. Addition and scalar multiplication are given 'pointwise'

$$\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix} \qquad a \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} aa_1 \\ \vdots \\ aa_n \end{pmatrix}$$

In a certain sense, the only vector spaces that will be of interest to us are the k^n for some natural number n. To show this we need the definition of a basis. A **basis** $\mathcal{B}$ of V is a set of vectors with the property that for every v in V there are uniquely determined scalars $\{a_w(v) : w \in \mathcal{B}\}$ with only finitely many of the $a_w(v)$ non-zero and such that

$$v = \sum_{w \in \mathcal{B}} a_w(v) w$$

We understand this sum as only taking place across the non-zero entries and equalling 0 if all entries are zero. It is clear that k^n has basis $\{e_1, \ldots, e_n\}$ where e_i has zeroes in all entries except for a 1 in the ith row. We call this the **canonical basis** for k^n.

Theorem 1. Every vector space V has a basis, and any two bases have the same cardinality.

Proof. We say a set of vectors $\mathcal{B}$ is **linearly independent** if it has the property that whenever $\{a_w : w \in \mathcal{B}\}$ are scalars with only finitely many of them non-zero and $\sum_{w \in \mathcal{B}} a_w w = 0$ then all $a_w = 0$. We say a vector v is **spanned by** a set of vectors $\mathcal{B}$ provided there are scalars $\{a_w(v) : w \in \mathcal{B}\}$ with only finitely many of them non-zero satisfying $v = \sum_{w \in \mathcal{B}} a_w(v) w$. So a set of vectors is a basis if and only if it spans all of V and is linearly independent.

Now let $\mathcal{S}$ be the set of linearly independent subsets of V. Notice that subset inclusion $\subseteq$ defines a partial order on $\mathcal{S}$. We want to invoke Zorn's lemma to prove that $\mathcal{S}$ has a maximal element $\mathcal{B}$. So suppose $\mathcal{C} = \{\mathcal{B}_i : i \in I\}$ is a chain in $\mathcal{S}$ and define $\mathcal{B}_\mathcal{C} = \bigcup_{i \in I} \mathcal{B}_i$. Suppose we have an expression of the form $\sum a_w w = 0$ where w are in $\mathcal{B}_\mathcal{C}$ and all but finitely many of the a_w are non-zero. For each non-zero a_w choose $\mathcal{B}_i$ containing w, let $\mathcal{B}'$ be the largest of the subsets among these $\mathcal{B}_i$ (which is possible since there were only finitely

many $\mathcal{B}_i$ chosen), and notice that this contradicts the fact that $\mathcal{B}'$ is linearly independent. This contradiction shows that $\mathcal{B}_\mathcal{C}$ is linearly independent, and so all chains in $\mathcal{S}$ have upper bounds. Zorn's lemma now guarantees the existence of a maximal element $\mathcal{B}$ in $\mathcal{S}$. Notice that if $\mathcal{B}$ does not span all of V then there is $v \neq 0$ in V that is not spanned by $\mathcal{B}$. But then $\mathcal{B} \cup \{v\}$ is linearly independent, contradicting the maximality of $\mathcal{B}$. This contradiction shows that $\mathcal{B}$ spans all of V and is hence a basis.

Now suppose $\mathcal{B}_1$ and $\mathcal{B}_2$ are two bases of V. We argue that $\mathcal{B}_1$ and $\mathcal{B}_2$ have the same cardinality by considering two separate cases: $|\mathcal{B}_1|$ or $|\mathcal{B}_2|$ is finite versus $|\mathcal{B}_1|$ and $|\mathcal{B}_2|$ are infinite. So assume first that one of $|\mathcal{B}_1|$ or $|\mathcal{B}_2|$ is finite, and suppose for the sake of contradiction that $|\mathcal{B}_2| \neq |\mathcal{B}_1|$. Let's say that $\mathcal{B}_1$ has elements $v_1, \ldots, v_n$ and that $\mathcal{B}_2$ contains more than n elements, say $w_1, \ldots, w_{n+1}$. Choose scalars a_{ij} satisfying

$$w_i = \sum_{j=1}^{n} a_{ji} v_j$$

Let $x_1, \ldots, x_{n+1}$ denote variables with values in k and note that

$$\sum_{i=1}^{n+1} x_i w_i = \sum_{i=1}^{n+1} \sum_{j=1}^{n} a_{ji} x_i v_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{n+1} a_{ji} x_i \right) v_j$$

So we would have a linear dependence relation $\sum_{i=1}^{n+1} x_i w_i = 0$ provided the equations $\sum_{i=1}^{n+1} a_{ji} x_i = 0$ have a solution where not all $x_i = 0$. Since there are more variables than equations, and the equations are homogeneous, we know from the theory of equations that there is a non-trivial solution. This contradicts the linear independence of $\mathcal{B}_2$. Thus $|\mathcal{B}_2| = |\mathcal{B}_1|$ if $|\mathcal{B}_1|$ is finite.

In the second case we assume $|\mathcal{B}_1|$ and $|\mathcal{B}_2|$ are both infinite. For v in $\mathcal{B}_1$ write $v = \sum_{w \in \mathcal{B}_2} a_w(v) w$ where only finitely many $a_w(v) \neq 0$ and define

$$f(v) = \{w : a_w(v) \neq 0\}$$

Since V is spanned by $\mathcal{B}_1$ and by no proper subset of $\mathcal{B}_2$ we must have

$$\mathcal{B}_2 = \bigcup_{v \in \mathcal{B}_1} f(v)$$

We then get $|\mathcal{B}_2| \leqslant |\mathcal{B}_1|$. Similarly $|\mathcal{B}_1| \leqslant |\mathcal{B}_2|$ so that $|\mathcal{B}_1| = |\mathcal{B}_2|$. ∎

The cardinality of any basis for V is called the **dimension** of V, and is denoted by $\dim_k V$ or $\dim V$ is k is clear from context. From now on we assume every vector space under consideration is finite dimensional. Now let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis with n elements ordered in some particular way and define the **coordinate** function $[\cdot]_{\mathcal{B}} : V \to k^n$ for V relative to $\mathcal{B}$ by

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$$

where $a_1, \ldots, a_n$ are chosen in k to satisfy $v = a_1 v_1 + \cdots + a_n v_n$. Notice

$$[v + w]_{\mathcal{B}} = [v]_{\mathcal{B}} + [w]_{\mathcal{B}}$$
$$[av]_{\mathcal{B}} = a[v]_{\mathcal{B}}$$

for v, w in V and a in k. This is to say that $[\cdot]_{\mathcal{B}}$ is an isomorphism. So every finite dimensional vector space is isomorphic to k^n for $n = \dim_k(V)$. Now given a linear transformation T it is often useful to identify both the domain and codomain with an appropriate space of column vectors and to then identify the action of T with matrix multiplication.

Proposition 2. Suppose $\mathcal{B}_1 = \{v_1, \ldots, v_m\}$ is a basis of V and $\mathcal{B}_2 = \{w_1, \ldots, w_n\}$ is a basis of W. For $T : V \to W$ a linear transformation, choose scalars t_{ij} satisfying

$$Tv_i = \sum_{j=1}^{n} t_{ji} w_j$$

Define the $n \times m$ matrix $[T]_{\mathcal{B}_2}^{\mathcal{B}_1}$ by $[T]_{\mathcal{B}_2}^{\mathcal{B}_1}(i, j) = t_{ij}$. For all v in V we have

$$[Tv]_{\mathcal{B}_2} = [T]_{\mathcal{B}_2}^{\mathcal{B}_1} [v]_{\mathcal{B}_1}$$

Proof. Write $v = \sum_{i=1}^{m} a_i v_i$ and note that

$$Tv = \sum_{i=1}^{m} a_i T v_i = \sum_{i=1}^{m} \sum_{j=1}^{n} t_{ji} a_i w_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} t_{ji} a_i \right) w_j$$

On the other hand, the definition of matrix multiplication gives

$$[T]^{\mathcal{B}_1}_{\mathcal{B}_2}[v]_{\mathcal{B}_1}(j) = \sum_{i=1}^{m}[T]^{\mathcal{B}_1}_{\mathcal{B}_2}(j,i)[v]_{\mathcal{B}_1}(i) = \sum_{i=1}^{m} t_{ji}a_i$$

This establishes the result. ∎

Using row reduction on the matrix $[T]^{\mathcal{B}}_{\mathcal{B}}$ we can find a basis for the **kernel** of T, denoted $\mathrm{Ker}(T)$, which consists of all v satisfying $Tv = 0$. The dimension of the kernel is called the **nullity** of T. Using row reduction on the transpose of $[T]^{\mathcal{B}}_{\mathcal{B}}$ we can find a basis for the image of T, denoted $\mathrm{Im}(T)$, whose dimensions is called the **rank** of T. If we allow $v_1, \ldots, v_r$ to be a basis for $\mathrm{Ker}(T)$ and choose $v_{r+1}, \ldots, v_m$ so that $v_1, \ldots, v_m$ is a basis for V, then $Tv_{r+1}, \ldots, Tv_m$ is a basis for $\mathrm{Im}(T)$. So we've proven the rank-nullity theorem, which states

$$\dim \mathrm{Ker}(T) + \dim \mathrm{Im}(T) = \dim V$$

It's straightforward to show that if $S : U \to V$ and $T : V \to W$ where U, V, and W have bases $\mathcal{B}_1$, $\mathcal{B}_2$, and $\mathcal{B}_3$ respectively then

$$[T \circ S]^{\mathcal{B}_1}_{\mathcal{B}_3} = [T]^{\mathcal{B}_2}_{\mathcal{B}_3}[S]^{\mathcal{B}_1}_{\mathcal{B}_2}$$

Suppose V has bases $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $\mathcal{B}' = \{v'_1, \ldots, v'_n\}$ and that P is the matrix defined by

$$P = [[v'_1]_{\mathcal{B}}, \ldots, [v'_n]_{\mathcal{B}}]$$

Then $P = [I]^{\mathcal{B}'}_{\mathcal{B}}$. This means that

$$[u]_{\mathcal{B}} = [I]^{\mathcal{B}'}_{\mathcal{B}}[u]_{\mathcal{B}'}$$

Of course $[I]^{\mathcal{B}}_{\mathcal{B}'}[I]^{\mathcal{B}'}_{\mathcal{B}} = [I]^{\mathcal{B}'}_{\mathcal{B}'} = I$ is the identity matrix so that $[I]^{\mathcal{B}}_{\mathcal{B}'} = P^{-1}$. We can now derive the **change of basis formula**

$$[T]^{\mathcal{B}'}_{\mathcal{B}'} = [I]^{\mathcal{B}}_{\mathcal{B}'}[T]^{\mathcal{B}}_{\mathcal{B}}[I]^{\mathcal{B}'}_{\mathcal{B}} = P^{-1}[T]^{\mathcal{B}}_{\mathcal{B}}P$$

Lastly, we define the **dual space** $V^* = \mathrm{Hom}(V, k)$. Elements of V^* are called **linear functionals**. Given a basis $\mathcal{B} = \{v_1, \ldots, v_m\}$ of V there is a **dual basis** $\mathcal{B}^*$ of V^* defined by $\mathcal{B}^* = \{v_1^*, \ldots, v_m^*\}$ where $v_i^*(v_j) = \delta_{ij}$. For f in V^* we have the identity

$$f = \sum_{i=1}^{m} f(v_i)v_i^*$$

which is true since it's true for the elements of $\mathcal{B}$. Moreover, given a linear transformation $T : V \to W$ there is induced a linear transformation $T^* : W^* \to V^*$ defined by $T^*f = f \circ T$. We can compute a matrix that represents T^* if we allow $\mathcal{B}' = \{w_1, \ldots, w_n\}$ and $\mathcal{B}'^* = \{w_1^*, \ldots, w_n^*\}$. Then

$$(T^*w_i^*)(v_j) = (w_i^* \circ T)(v_j) = w_i^*(Tv_j) = \sum_{l=1}^{n} w_i^*(t_{lj}w_l) = t_{ij}$$

This means that

$$T^*w_i^* = \sum_{j=1}^{m} t_{ij}v_j^*$$

and hence $[T^*]_{\mathcal{B}^*}^{\mathcal{B}'^*}$ is the matrix transpose of the matrix $[T]_{\mathcal{B}'}^{\mathcal{B}}$.

0.3 Decomposition Theorems

Suppose T is a linear operator on V. We say the subspace W of V is **T-invariant** provided that Tw is in W for every w in W. In this section we study the ways in which V can be decomposed in terms of T-invariant subspaces.

To begin, we say V is the direct sum of the subspaces W_1 and W_2 provided that $W_1 \cap W_2 = 0$ and every v in V can be expressed as $v = w_1 + w_2$ with w_1 in W_1 and w_2 in W_2. In this case we write $V = W_1 \oplus W_2$. In general we can write $V = \bigoplus_{i \in I} W_i$ for an indexing set I, which must be finite since we assume V is finite dimensional. So we want to express V as a direct sum of T-invariant subspaces.

The simplest examples of T-invariant subspaces are those of dimension 1. So suppose the T-invariant subspace W is spanned by the non-zero vector v, and note that $Tv = av$ for some scalar a. We call a scalar a that arises in this way an **eigenvalue** of T, and we call any non-zero vector w satisfying $Tw = a$ an **eigenvector** corresponding to the eigenvalue a. Note that $Tv = av$ means

$(T - aI)v = 0$ where $I : V \to V$ is the identity operator on V. In particular, since $T - aI$ has the non-zero vector v in its kernel its determinant must equal 0. That means $p(a) = 0$ where

$$p(x) = \det(T - xI)$$

is the **characteristic polynomial** of T. Finally, since the matrices representing I, T, T^2, $\ldots$, T^{n^2} cannot all be linearly independent, we must have a non-trivial linear dependence relation among them. This means there are non-zero polynomials $f(x)$ satisfying $f(T) = 0$. We choose a monic one of minimal degree, say $m(x)$, which we call the **minimal polynomial** of T. A polynomial $f(x)$ satisfies $f(T) = 0$ if and only if $f(x)$ divides $m(x)$. We are now ready to establish our first decomposition result.

Proposition 3 (Diagonalizability)**.** Given a linear operator T acting on V, V can be expressed as a sum of 1-dimensional T-invariant subspaces if and only if the minimal polynomial for T is a product of distinct linear polynomials.

Proof. Suppose V can be expressed as a sum of 1-dimensional T-invariant subspaces. Let $a_1, \ldots, a_r$ be the distinct eigenvalues corresponding to these subspaces and notice that each subspace is contained in the kernel of $T - a_i I$ for some value of i. But then each of the subspaces is in the kernel of $\prod_{i=1}^{n}(T - a_i I)$ and so V equals the kernel of $\prod_{i=1}^{n}(T - a_i I)$. This means $\prod_{i=1}^{n}(T - a_i I) = 0$, so that $m(x)$ divides $\prod_{i=1}^{n}(x - a_i)$. This establishes that the minimal polynomial is a product of distinct linear polynomials.

Conversely, suppose the minimal polynomial for T is a product of distinct linear polynomials, say $x - a_1, \cdots, x - a_r$. Notice that

$$\prod_{i \neq 1} \frac{x - a_i}{a_1 - a_i} + \cdots + \prod_{i \neq r} \frac{x - a_i}{a_r - a_i} = 1$$

since the left hand side is a polynomial with degree at most $r - 1$ that agrees with 1 at $x = a_1, \ldots, a_r$. From this formula we conclude that for v in V we have $v = v_1 + \cdots + v_r$ where

$$v_i = \prod_{j \neq i} \frac{T - a_j I}{a_i - a_j} v$$

Note that $(T - a_i I)v_i = 0$ since $\prod(T - a_j I) = 0$, so that $V = \sum_{i=1}^{r} W_i$ where W_i is the kernel of $T - a_i I$. Suppose we have a non-trivial linear

dependence relation of the form $\sum w_i = 0$ where each w_i is in W_i. We may choose such a relation with a minimal number of non-zero w_i, say $w_{i_1}, \ldots, w_{i_s}$. Then $w_{i_1} + \cdots + w_{i_s} = 0$ and applying T to this equation produces

$$a_{i_1} w_{i_1} + \cdots + a_{i_s} w_{i_s} = 0$$

But since $a_{i_1}, \ldots, a_{i_s}$ are distinct we can use the two equations to produce a new non-trivial linear dependence relation with fewer than s many w_i, thus contradicting our choice of s. This contradiction shows that $V = \bigoplus_{i=1}^{r} W_i$. Since any basis for W_i decomposes W_i into a sum of 1-dimensional T-invariant subspaces, the proof is complete. ∎

In the case of the previous proposition we say that T is **diagonalizable** because the matrix $[T]_{\mathcal{B}}^{\mathcal{B}}$ that represents T with respect to a basis consisting of eigenvectors must be diagonal with eigenvalues as the diagonal entries. Of course each eigenvalue will occur in the diagonal as many times as it occurs as a root of the characteristic polynomial. However, even if the characteristic polynomial factors completely, the operator T need not be diagonalizable. For instance, we could have T act on the 2-dimensional vector space V that has basis $\{v_1, v_2\}$ by $Tv_1 = 0$ and $Tv_2 = v_1$.

For the next decomposition result we need a new definition. We call a properly ascending sequence $0 = V_0 \subset V_1 \subset \cdots \subset V_r = V$ of subspaces in V a **flag** of V. We say the flag is complete if $r = n$, in which case there must be a subspace in the flag of dimension i for any i between 0 and n. For W a T-invariant subspace of V and v an element of V, define the T-conductor $S_T(v; W)(x)$ as the monic polynomial of minimal degree among all polynomials $f(x)$ for which $f(T)v$ is in W.

Proposition 4 (Triangulability)**.** Given a linear operator T acting on V, V has a complete flag of T-invariant subspaces if and only if the minimal polynomial $m(x)$ for T splits over k.

Proof. Suppose there is a complete flag $0 = V_0 \subset V_1 \subset \cdots \subset V_n = V$ of T-invariant subspaces. Let v be in $V \backslash V_{n-1}$ and write $m_1(x) = s_T(v; V_{n-1})$ and $m_2(x)$ for the minimal polynomial of T restricted to V_{n-1}. By induction (with base case $V = 0$), the polynomial $m_2(x)$ splits over k. Since $m(x)$ is divisible by $m_1(x)m_2(x)$, we conclude that $m(x)$ splits over k as well.

Now suppose $m(x)$ splits over k. We construct a complete flag of T-invariant subspace inductively, beginning with $V_0 = 0$. Having defined $V_0 \subset \cdots \subset V_i$, if $i < n$ then let v be an element of V not contained in V_i. Note that $S_T(v; V_i)$ divides $m(x)$ and hence splits over k. Write

$$S_T(v; V_i)(x) = f(x)(x - a)$$

for some polynomial $f(x)$ and some a in k. Set $v' = f(T)v$ and notice by minimality that v' is not in W. On the other hand, $(T - aI)v'$ is in W and

$$Tv' = av' + (T - aI)v'$$

So if we set $V_{i+1} = V_i \oplus kv'$ then V_{i+1} is a T-invariant subspace of V with dimension $i+1$, and V_{i+1} contains V_i by construction. Proceeding inductively we obtain a complete flag of T-invariant subspaces. ∎

If $v_1, \ldots, v_n$ are chosen so that $V_i = V_{i-1} \oplus kv_i$ for $1 \leqslant i \leqslant n$, then $[T]_{\mathcal{B}}^{\mathcal{B}}$ is an upper triangular matrix where $\mathcal{B} = \{v_1, \ldots, v_n\}$. It is for this reason that T is called triangulable. It can be shown that a family of commuting diagonalizable matrices is simultaneously diagonalizable, and a family of commuting triangulable matrices is simultaneously triangulable.

Theorem 5 (Cayley-Hamilton)**.** The minimal polynomial of a linear operator divides its characteristic polynomial. The characteristic and minimal polynomials have the same roots, except for multiplicities.

Proof. Suppose T is a linear operator acting on V and let $m(x)$ denote the minimal polynomial of T and $p(x)$ its characteristic polynomial. We first show that m and p have the same roots. On the one hand, if $p(a) = 0$ then there is a non-zero vector v with $Tv = av$. Since $m(T)v = m(a)v$ and $m(T)v = 0$, we get $m(a) = 0$ so that a is a root of $m(x)$. On the other hand, if a is a root of $m(x)$ and we write $m(x) = f(x)(x - a)$ for some polynomial $f(x)$, then there is v in V with $f(T)v \neq 0$, and since $(T - aI)f(T)v = 0$ we get that $Tv' = av'$ where $v' = f(T)v \neq 0$. This means that $p(a) = 0$.

We now show that $m(x)$ divides $p(x)$ by demonstrating that $p(T) = 0$. In case $m(x)$ splits over k, we know that V has a complete flag of T-invariant subspaces $0 = V_0 \subset V_1 \subset \cdots \subset V_n = V$. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be the basis for V obtained by writing $V_i = V_{i-1} \oplus kv_i$ for $1 \leqslant i \leqslant n$. There are scalars $a_1, \ldots, a_n$ with the property that $(T - a_i I)v_i$ is in V_{i-1}. So

$[T]^{\mathcal{B}}_{\mathcal{B}}$ is a triangular matrix with diagonal entries $a_1, \ldots, a_n$, and of course $p(x) = (x - a_1) \cdots (x - a_n)$. We now compute the action of $p(T)$ on the flag $V_0 \subset \cdots \subset V_n$. Since $(T - a_i I)V_i \subseteq V_{i-1}$ for all i, by induction $p(T)V = 0$. So $p(T) = 0$ as required.

Lastly, if $m(x)$ does not split over k then define F to be the splitting field for $m(x)$ over k. Define the F-vector space $V^F = V \otimes_k F$ and a linear operator T^F on V^F by $T^F(v \otimes a) = Tv \otimes a$. Of course $[T^F]^{\mathcal{B}^F}_{\mathcal{B}^F} = [T]^{\mathcal{B}}_{\mathcal{B}}$ for $\mathcal{B} = \{v_i\}$ any basis of V where $B^F = \{v_i \otimes 1\}$. So T^F also has characteristic polynomial $p(x)$. For $f(x)$ any polynomial in $k[x]$ we get

$$f(T^F) = f(T) \otimes I$$

Since $p(T^F) = 0$ we get $p(T) = 0$ as well, thus completing the proof. ∎

From the Cayley-Hamilton theorem we can sharpen our abstractly obtained bound $\deg m(x) \leqslant (\dim V)^2$ to the more realistic bound

$$\deg m(x) \leqslant \dim V$$

Theorem 6 (Primary Decomposition)**.** Given a linear operator T acting on V with minimal polynomial $m(x)$ and prime factorization

$$m(x) = p_1(x)^{e_1} \cdot \ldots \cdot p_r(x)^{e_r}$$

we have the decomposition

$$V = \operatorname{Ker}(p_1(T)^{e_1}) \oplus \cdots \oplus \operatorname{Ker}(p_r(T)^{e_r})$$

Moreover, $p_i(x)^{e_i}$ is the minimal polynomial of T acting on $\operatorname{Ker}(p_i(T)^{e_i})$.

Proof. The kernel of $f(T)$ is T-invariant for any polynomial $f(x)$ since

$$f(T)Tw = Tf(T)w = T0 = 0$$

Define polynomials $q_i(x) = \frac{m(x)}{p_i(x)^{e_i}}$ and note that 1 is the only (monic) polynomial that divides every $q_1(x), \ldots, q_r(x)$. By Brahmagupta's result there are polynomials $f_1(x), \ldots, f_r(x)$ satisfying

$$f_1(x)q_1(x) + \cdots + f_r(x)q_r(x) = 1$$

So for v in V we have $v = \sum v_i$ where $v_i = f_i(T)q_i(T)v$. Note that

$$p_i(T)^{e_i} v_i = p_i(T)^{e_i} f_i(T) q_i(T) v = f_i(T) m(T) v = 0$$

so that v_i is in $\mathrm{Ker}(p_i(T)^{e_i})$ for each i, and hence $V = \sum \mathrm{Ker}(p_i(T)^{e_i})$. Now suppose $\sum v_i = 0$ where each v_i is in $\mathrm{Ker}(p_i(T)^{e_i})$. Since $q_i(T)$ annihilates $\mathrm{Ker}(p_j(T)^{e_j})$ for $j \neq i$ we get

$$0 = q_i(T) \left(\sum_j v_j \right) = \sum_j q_i(T) v_j = q_i(T) v_i$$

Then we also have

$$v_i = f_1(T) q_1(T) v_i + \cdots + f_r(T) q_r(T) v_i = 0$$

This means that we have a direct sum $V = \bigoplus \mathrm{Ker}(p_i(T)^{e_i})$. Lastly, the minimal polynomial of T acting on $\mathrm{Ker}(p_i(T)^{e_i})$ divides $p_i(x)^{e_i}$ and the minimal polynomial of T divides the product of the minimal polynomials of T acting on the various subspaces $\mathrm{Ker}(p_i(T)^{e_i})$. From this we see that T acting on $\mathrm{Ker}(p_i(T)^{e_i})$ must have minimal polynomial $p_i(x)^{e_i}$. ∎

Note that the proofs to the diagonalizability theorem and the primary decomposition theorem are very similar, with the exception that we could give explicit formulas for $f_1(x), \ldots, f_r(x)$ as $f_i(x) = \prod_{j \neq i} (a_i - a_j)^{-1}$ in case $m(x)$ splits over k into distinct linear factors.

For the next result we require the definition that an operator N is **nilpotent** provided that $N^d = 0$ for some exponent $d \geqslant 1$.

Corollary 7. Given a linear operator T acting on V with minimal polynomial that splits over k, there is a diagonalizable operator D and a nilpotent operator N for which

$$T = D + N \quad \text{and} \quad DN = ND$$

Moreover, D and N are polynomials in terms of T and are uniquely determined by the two conditions above.

Proof. Suppose the minimal polynomial $m(x)$ of T splits as

$$m(x) = (x - a_1)^{e_1} \cdots (x - a_r)^{e_r}$$

In the notation of the proof to the Primary Decomposition theorem set

$$E_i = f_i(T)q_i(T)$$

for $1 \leqslant i \leqslant r$. Note that $E_iE_j = 0$ for $i \neq j$ and

$$E_1 + \cdots + E_r = I$$

so that $E_i^2 = E_i$ for all i. Now define linear operators D and N on V by

$$\begin{aligned} D &= a_1E_1 + \cdots + a_rE_r \\ N &= T - D \end{aligned}$$

It is clear that D and N are polynomials in T so that $DN = ND$, and of course $T = D + N$. Notice that D is diagonalizable since

$$V = \bigoplus_{i=1}^{r} \mathrm{Ker}((T - a_iI)^{e_i}) = \bigoplus_{i=1}^{r} \mathrm{Im}(E_i)$$

and $Dv = a_iv$ for v in $\mathrm{Im}(E_i)$. On the other hand

$$\begin{aligned} T &= TE_1 + \cdots + TE_r \\ N &= (T - a_1I)E_1 + \cdots + (T - a_rI)E_r \\ N^d &= (T - a_1I)^dE_1 + \cdots + (T - a_rI)^dE_r \end{aligned}$$

so that $N^e = 0$ where e is the maximum value of $\{e_1, \ldots, e_r\}$. This means that N is nilpotent. Now suppose $T = D' + N'$ with D' diagonalizable, N' nilpotent, and $D'N' = N'D'$. Since D and N commute with each other, we see that they commute with T. Since D and N are polynomials in T, all of the operators D, D', N, and N' commute with one another. Since

$$D - D' = N' - N$$

we see that $D - D'$ is a nilpotent operator. But since it's diagonalizable it must equal 0. Then $D' = D$, which also implies that $N' = N$, as needed. ∎

The decomposition of $T = D + N$ is also an easy consequence of the Jordan canonical form presented below. For the next result we introduce the notation that $Z(v;T)$ is the subspace of V generated by the vectors of the form $f(T)v$ for $f(x)$ a polynomial over k.

Theorem 8 (Cylic Decomposition)**.** Given a linear operator T acting on V, there are $v_1, \ldots, v_r$ with minimal polynomials $m_1(x), \ldots, m_r(x)$ for which

(i) $V = Z(v_1;T) \oplus \cdots \oplus Z(v_r;T)$

(ii) $m_{i+1}(x)$ divides $m_i(x)$ if $1 \leqslant i \leqslant r-1$

Proof. Choose v_1 in V with

$$\dim Z(v_1;T) = \max\{\dim Z(v;T) : v \in V\}$$

Suppose we have chosen $v_1, \ldots, v_i$ subject to the conditions

(i) $\sum_{j=1}^{i} Z(v_j;T) = \bigoplus_{j=1}^{i} Z(v_j;T)$

(ii) If $1 \leqslant j < i$ and $W_j = \sum_{l=1}^{j} Z(v_l;T)$ then

$$\deg S_T(v_{j+1};W_j) = \max\{\deg S_T(v;W_j) : v \in V\}$$

(iii) For $m_j(x)$ the minimal polynomial of v_j, $m_{j+1}(x) \mid m_j(x)$ for $1 \leqslant j < i$

If $\sum_{j=1}^{i} Z(v_i;T) = V$ then we are done. Otherwise, write $W_i = \bigoplus_{j=1}^{i} Z(v_j;T)$ and choose v'_{i+1} in V satisfying

$$\deg S_T(v'_{i+1};W_i) = \max\{\deg S_T(v;W_i) : v \in V\}$$

Select polynomials $g_1(x), \ldots, g_i(x)$ for which

$$s_T(v'_{i+1};W_i)(T)v'_{i+1} = g_1(T)v_1 + \cdots + g_i(T)v_i$$

Choose $h_j(x)$ and $r_j(x)$ with $\deg r_j(x) < \deg s_T(v'_{i+1};W_i)(x)$ and

$$g_j(x) = h_j(x)s_T(v'_{i+1};W_i)(x) + r_j(x)$$

Define the vector v_{i+1} by

$$v_{i+1} = v'_{i+1} - \sum_{j=1}^{i} h_j(T)v_j$$

Note that $s_T(v'_{i+1}; W_i) = s_T(v_{i+1}; W_i)$. Moreover

$$s_T(v_{i+1}; W_i)(T)v_{i+1} = \sum_{j=1}^{i} r_j(T)v_j \tag{1}$$

We claim that $r_j(x) = 0$ for $1 \leqslant j \leqslant i$. Suppose for the sake of contradiction that some $r_j(x)$ is non-zero. Let i^* denote the largest value of j for which $r_j(x)$ is non-zero. Note that $s_T(v_{i+1}; W_i)(x)$ divides $s_T(v_{i+1}; W_{i^*-1})(x)$ since $s_T(v_{i+1}; W_{i^*-1})(T)$ sends v_{i+1} into W_i, and so we can write

$$s_T(v_{i+1}; W_{i^*-1})(x) = g(x)s_T(v_{i+1}; W_i)(x)$$

for some polynomial $g(x)$. Then multiplying (1) by $g(T)$ produces

$$s_T(v_{i+1}; W_{i^*-1})(T)v_{i+1} = g(T)r_{i^*}(T)v_{i^*} + \sum_{j=1}^{i^*-1} g(T)r_j(T)v_j$$

where the terms for $j > i^*$ can be ignored since $r_j(T)v_j = 0$ for $j > i^*$. Notice that $g(T)r_{i^*}(T)v_{i^*}$ must be inside of W_{i^*-1}. We conclude that $s_T(v_{i+1}; W_{i^*-1})(x)$ divides $g(x)r_{i^*}(x)$. However

$$\begin{aligned}
\deg g(x)r_{i^*}(x) &= \deg g(x) + \deg r_{i^*}(x) \\
&= \deg s_T(v_{i+1}; W_{i^*-1})(x) - \deg s_T(v_{i+1}; W_i)(x) + \deg r_{i^*}(x) \\
&< \deg s_T(v_{i+1}; W_{i^*-1})(x) - \deg s_T(v_{i+1}; W_i)(x) + \deg s_T(v_{i+1}; W_i)(x) \\
&= \deg s_T(v_{i+1}; W_{i^*-1})(x)
\end{aligned}$$

This contradiction shows that all $r_j(x) = 0$, and hence

$$s_T(v_{i+1}; W_i)(T)v_{i+1} = 0$$

We conclude that (i) is satisfied. Of course (ii) was satisfied by choice of v_{i+1}. Lastly, if we let $m_{i+1}(x)$ be the minimal polynomial of v_{i+1} and consider the equation

$$m_{i+1}(T)v_{i+1} = 0 = m_1(T)v_1 + \cdots + m_i(T)v_i$$

where each term equals zero. Our arguments above show that $m_{i+1}(x)$ divides $m_j(x)$ for $1 \leqslant j \leqslant i$. We now proceed inductively on i until we have spanned all of V.

■

It can be shown that r and $m_1(x), \ldots, m_r(x)$ in the cyclic decomposition are uniquely determined by conditions (i) and (ii). Also, if

$$m(x) = x^n + \cdots + a_1 x + a_0$$

then with respect to the basis $\{v, Tv, \ldots, T^{n-1}v\}$ the operator T restricted to $Z(v;T)$ has basis given by the companion matrix

$$\begin{pmatrix} 0 & 0 & 0 & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & -a_1 \\ 0 & 1 & 0 & \cdots & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -a_{n-1} \end{pmatrix}$$

There are two conditions that can prevent T from being diagonalizable:

(a) Its minimal polynomial $m(x)$ does not split over k, or

(b) $m(x)$ does split but it has a repeated root

We suppose $m(x)$ does split over k, and we will analyze the situation for when $m(x)$ has multiple roots. Say $(x-a)^e$ is the highest power of $x-a$ that occurs in the prime factorization of $m(x)$ and that $e \geqslant 1$. By the Primary Decomposition theorem we know that $\mathrm{Ker}((T-aI)^e)$ occurs in a direct decomposition of V in terms of T-invariant subspaces and that $(x-a)^e$ is the minimal polynomial of T acting on $\mathrm{Ker}((T-a)^e)$. This means that there is v in V with

$$(T-aI)^e v = 0 \quad \text{and} \quad (T-aI)^{e-1} v \neq 0$$

We call such a v a **generalized eigenvector** of T with eigenvalue a and **rank** e. Notice that ordinary eigenvectors are the generalized eigenvectors with rank 1. The cyclic subspace $Z(v;T)$ generated by v has basis

$$\mathcal{B} = \{v, (T - aI)v, (T - aI)^2 v, \ldots, (T - aI)^{e-1} v\}$$

Let $\iota : Z(v; T) \to V$ denote the inclusion map. Then

$$[T \circ \iota]_{\mathcal{B}}^{\mathcal{B}} = \begin{pmatrix} a & 0 & 0 & \cdots & 0 \\ 1 & a & 0 & \cdots & 0 \\ 0 & 1 & a & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a \end{pmatrix}$$

with every diagonal entry equal to a, every subdiagonal entry equal to 1, and all other entries equal to 0. This follows from the fact that

$$T(T - aI)^i v = a(T - aI)^i v + (T - aI)^{i+1} v$$

We use the notation $J_e(a)$ for the matrix $[T \circ \iota]_{\mathcal{B}}^{\mathcal{B}}$, and we call $J_e(a)$ a **Jordan block**. Notice that the only (ordinary) eigenvectors of T in $Z(v; T)$ are non-zero scalar multiples of v. So the fact that $m(x)$ has a multiple root corresponds to the fact that T has 'too few' eigenvectors. More specifically, we can define the **algebraic multiplicity** d of a as the multiplicity of a as a root of the characteristic polynomial $p(x)$ and the **geometric multiplicity** g of a as the maximal size of a linearly independent set of eigenvectors with eigenvalue a. In other words, $g = \dim(\mathrm{Ker}(T - aI))$. So we have the inequalities

$$\frac{d}{e} \leqslant g \leqslant d - (e - 1)$$

To summarize, the operator $T \circ \iota$ is diagonalizable if and only $g = d$. This can only occur when $e = 1$, but if $e = 1$ then it is guaranteed to occur. The next result will show that as long as $m(x)$ splits over k then $\mathrm{Ker}((T - aI)^e)$ will have a basis of generalized eigenvectors.

Theorem 9 (Jordan Canonical Form)**.** Given a linear operator T acting on V and a a root of $m(x)$ that occurs with multiplicity e, the T-invariant subspace $\mathrm{Ker}((T - aI)^e)$ has a basis of generalized eigenvectors.

Proof. We may suppose $V = \mathrm{Ker}((T-aI)^e)$ for simplicity. Define $N = T-aI$ so that N is a nilpotent operator on V. It suffices to show that V has a basis consisting of generalized eigenvectors for N. For this we proceed by induction on $\dim V$, where for the base case $\dim V = 1$ we can simply take any non-zero vector in V as a basis. So suppose the result is true for vector spaces with dimension less than $\dim V$ and note that $\dim NV < \dim V$. By our inductive hypothesis there are generalized eigenvectors $v_1', \ldots, v_s'$ and integers $d_1, \ldots, d_s \geqslant 1$ so that NV has basis

$$\{v_1', Nv_1', \ldots, N^{d_1-1}v_1', \ldots, v_s', Nv_s', \ldots, N^{d_s-1}v_s'\}$$

Choose $v_1, \ldots, v_s$ in V with $Nv_i = v_i'$. Then

$$\{v_1, Nv_1, \ldots, N^{d_1}v_1, \ldots, v_s, Nv_s, \ldots, N^{d_s}v_s\}$$

is a linearly independent subset of V. After all, if

$$\sum_{i=1}^{s}\sum_{j=0}^{d_i} a_{ij}N^j v_i = 0$$

then applying N to this summation produces

$$\sum_{i=1}^{s}\sum_{j=0}^{d_i-1} a_{ij}N^j v_i' = 0$$

so that $a_{ij} = 0$ for $1 \leqslant i \leqslant s$ and $0 \leqslant j \leqslant d_i - 1$. But then

$$0 = \sum_{i=1}^{s} a_{id_i}N^{d_i}v_i = \sum_{i=1}^{s} a_{id_i}N^{d_i-1}v_i'$$

so that $a_{id_i} = 0$ for all i as well. For an arbitrary v in V we can choose scalars a_{ij} for which

$$Nv = \sum_{i=1}^{s}\sum_{j=0}^{d_i-1} a_{ij}N^j v_i' = N\sum_{i=1}^{s}\sum_{j=0}^{d_i-1} a_{ij}N^j v_i$$

This means there is $\tilde{v}$ in $\mathrm{Ker}(N)$ for which

$$v = \tilde{v} + \sum_{i=1}^{s}\sum_{j=0}^{d_i-1} a_{ij}N^j v_i$$

So $\mathrm{Ker}(N)$ and $\{N^j v_i\}$ spans all of V (though we won't get a trivial intersection!) So if $\{N^j v_i\}$ does not already span V, we choose $v_{s+1}, \ldots, v_r$ in $\mathrm{Ker}(N)$ to complement the span of $\{N^j v_i\}$. Set $d_i = 1$ for $s < i \leqslant r$. We now have the appropriate basis for V, thus completing the proof. ∎

In the notation of the theorem, the action of T on $\mathrm{Ker}((T - aI)^e)$ is represented by the block diagonal matrix whose entries are Jordan blocks $J_{d_1}(a), \ldots, J_{d_r}(a)$. The integers $d_1, \ldots, d_r$ are unique up to a reordering since they correspond to the dimensions of the cyclic subspaces $Z(v_i; T)$ in a cyclic decomposition of $\mathrm{Ker}((T - aI)^e)$.

There are further decompositions if we specialize to the situation of complex numbers, including polar decomposition, singular value decomposition, QR decomposition, Cholesky decomposition, etc. It is an open question the extent to which these decompositions translate to fields different from $\mathbb{C}$.

Research Problem. Analyze the singular value decomposition of a matrix with entries in a finite field.

0.4 Bilinear Forms

Recall that we write $B : V \times V \to k$ for a bilinear form of V. We say that B is **non-degenerate** if $B(u, v) = 0$ for all u implies that $v = 0$. We say B is **symmetric** provided $B(v, u) = B(u, v)$ for all u, v, and we say B is **alternate** provided $B(v, u) = -B(u, v)$ for all u, v.

Suppose $\mathcal{B} = \{v_1, \ldots, v_n\}$ is a basis of V and define $b_{ij} = B(v_i, v_j)$. For u and v in V write $u = \sum a_i v_i$ and $v = \sum b_j v_j$. Then

$$B(u, v) = B\left(\sum a_i v_i, \sum b_j v_j\right) = \sum\sum a_i b_{ij} b_j = [u]_{\mathcal{B}}^t \hat{B}_{\mathcal{B}} [v]_{\mathcal{B}}$$

where $\hat{B} = \hat{B}_{\mathcal{B}}$ is the matrix with $\hat{B}(i, j) = b_{ij}$. We call $\hat{B}$ the **matrix of B relative to $\mathcal{B}$**. If $B(u, v) = 0$ for all u, then $\hat{B}[v]_{\mathcal{B}} = 0$ and hence $[v]_{\mathcal{B}}$ is in the kernel of $\hat{B}$. So B is non-degenerate if and only if $\hat{B}$ is non-singular. Similarly, B is symmetric if and only if $\hat{B}^t = \hat{B}$ and B is alternate if and only if $\hat{B}^t = -\hat{B}$. If $\mathcal{B}'$ is another basis for V then

$$[u]^t_{\mathcal{B}'}\hat{B}_{\mathcal{B}'}[v]_{B'} = B(u,v) = [u]^t_{\mathcal{B}}\hat{B}_{\mathcal{B}}[v]_{\mathcal{B}} = [u]^t_{\mathcal{B}'}([I]^{\mathcal{B}'}_{\mathcal{B}})^t\hat{B}_{\mathcal{B}}[I]^{\mathcal{B}'}_{\mathcal{B}}[v]_{\mathcal{B}'}$$

and so we have the change of basis formula

$$\hat{B}_{\mathcal{B}'} = ([I]^{\mathcal{B}'}_{\mathcal{B}})^t\hat{B}_{\mathcal{B}}[I]^{\mathcal{B}'}_{\mathcal{B}}$$

We say matrices A and B are **congruent** if there is an invertible matrix P for which $B = P^tAP$. So the matrices associated to a bilinear form are all congruent. Note that this is a different relationship from being conjugate. We also say two bilinear forms B_1 and B_2 on V are **equivalent** provided there is a vector space automorphism T of V satisfying $B_2(Tu, Tv) = B_1(u, v)$ for all u and v in V. This is equivalent to the existence of bases for which the matrices representing B_1 and B_2 are equal.

Now the symmetric and alternate forms are united by a common desirable property. Namely, we say u and v are **orthogonal**, denoted as $u \perp v$, provided $B(u, v) = 0$.

Proposition 10. The $\perp$ relation is symmetric iff B is symmetric or alternate.

Proof. It's clear that $\perp$ is symmetric if B is symmetric or alternate, so we assume that $\perp$ is a symmetric relation. For u, v, w in V define

$$z = B(u,v)w - B(u,w)v$$

and consider

$$\begin{aligned} B(u,z) &= B(u, B(u,v)w - B(u,w)v) \\ &= B(u,w)B(u,v) - B(u,w)B(u,v) = 0 \end{aligned}$$

So $u \perp z$ and hence $z \perp u$, which means that

$$\begin{aligned} 0 &= B(z,u) \\ &= B(B(u,v)w - B(u,w)v, u) \\ &= B(u,v)B(w,u) - B(u,w)B(v,u) \end{aligned}$$

So

$$B(u,v)B(w,u) - B(u,w)B(v,u) = 0 \tag{2}$$

for all u, v, w. Now suppose for the sake of contradiction that B is not symmetric or alternate. Then there are u', v', w' in V with $B(u', u') \neq 0$ and $B(w', v') \neq B(v', w')$. Taking $(u, v, w) = (v', v', w')$ in (2) provides

$$B(v', v') = 0$$

Taking $(u, v, w) = (u', v', u')$ provides

$$B(v', u') = B(u', v')$$

Taking $(u, v, w) = (u', u', w')$ provides

$$B(w', u') = B(u', w')$$

Replacing (u, v, w) first with (v', w', u') and then with (w', v', u') produces

$$\begin{aligned} B(v', w')B(u', v') &= B(v', u')B(w', v') \\ B(w', v')B(u', w') &= B(w', u')B(v', w') \end{aligned}$$

Since $B(w', v') \neq B(v', w')$ we get $B(u', v') = 0$ and $B(u', w') = 0$. Hence

$$B(u' + v', w') = B(v', w') \neq B(w', v') = B(w', u' + v')$$

Replacing u and v with $u' + v'$ in (2) produces $B(u' + v', u' + v') = 0$. If we expand this we get the contradiction that

$$0 = B(u' + v', u' + v') = B(u', u') + B(v', u') + B(u', v') + B(v', v') = B(u', u')$$

This contradiction shows that B is either symmetric or alternate. ∎

It is customary to call B **reflexive** if it is symmetric or alternate. From now on we assume that B is reflexive. For W a subspace of V we define the **orthogonal complement** $W^{\perp}$ of W as the set of all v for which $B(v, w) = 0$ for all w in W. It's easy to see that if B is non-degenerate then

$$\dim W + \dim W^{\perp} = \dim V$$

though it need not be the case that $W \cap W^\perp = 0$. In fact, we define the **radical** of W as $\mathrm{Rad}(W) = W \cap W^\perp$ and we say that W is **non-degenerate** relative to B provided $\mathrm{Rad}(W) = 0$. Of course if W is non-degenerate then $V = W \oplus W^\perp$ where the symbol $\oplus$ indicates a direct sum of perpindicular subspaces. We are now in a position to describe all alternate forms.

Theorem 11. If B is alternate on V then there are two-dimensional subspaces $W_1, \ldots, W_r$ of V for which

$$V = W_1 \oplus \cdots \oplus W_r \oplus \mathrm{Rad}(V)$$

and there are bases for each W_i for which

$$\widehat{B|_{W_i}} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

In particular, if B is non-degenerate then $\dim V$ is even.

Proof. If $B = 0$ then $V = \mathrm{Rad}(V)$. Otherwise there are u, v with $B(u, v) \neq 0$. Of course u, v are linearly independent since B is alternate. Define $u_1 = B(u, v)^{-1}u$ and $v_1 = v$ so that $B(u_1, v_1) = 1$ and $B(v_1, u_1) = -1$. Since $B(u_1, u_1) = B(v_1, v_1) = 0$ we see that $\widehat{B|_{W_1}}$ has the appropriate form, where W_1 is the subspace of V spanned by u_1 and v_1. Then $V = W_1 \oplus W_1^\perp$ and we can proceed inductively with the restriction of B to $W_1^\perp$. ∎

We call W a **hyperbolic plane** with **byperbolic pair** if $\widehat{B_W}$ has the form above relative to the basis $\mathcal{B} = \{u, v\}$. So for an alternate sum the vector space decomposes as a sum of hyperbolic planes along with a space for which the bilinear form is trivial.

The situation for symmetric bilinear forms is more complicated. To begin, the **discriminant** $\mathrm{disc}(B)$ of a bilinear form is defined to be $\det(\hat{B})$. If B is degenerate then $\mathrm{disc}(B) = 0$. In case B is nondegenerate we regard $\mathrm{disc}(B)$ as an element of $k^x/(k^x)^2$ since $\det(\hat{B})$ is defined up to the square of a unit (as all matrices representing B are congruent). Two nondegenerate symmetric bilinear forms B_1 and B_2 on vector spaces V_1 and V_2 over a finite field (with characteristic different from 2) are equivalent if and only if $\dim V_1 = \dim V_2$ and $\mathrm{disc}(B_1) = \mathrm{disc}(B_2)$. So up to equivalence there are two nondegenerate

forms. A vector space having a nondegenerate symmetric bilinear form is called a **quadratic space**.

0.5 Real and Complex Vector Spaces

Suppose B is a bilinear form on the real vector space V. We say B is **positive definite** provided that $B(v,v) \geqslant 0$ for all v in V and $B(v,v) = 0$ only for $v = 0$.

Theorem 12. Every real symmetric matrix is diagonalizable.

Proof. TO BE COMPLETED ∎

0.6 Classical Groups

For $T : V \to V$ a linear transformation, the equality $B(Tu, Tv) = B(u, v)$ is equivalent to

$$[T]^t \hat{B}[T] = \hat{B}$$

where we shorten $[T]^{\mathcal{B}}_{\mathcal{B}}$ to $[T]$ for convenience. The classical groups arise by considering different bilinear forms and the groups that preserve them. The following theorem is our general tool in establishing the simplicity of these families of finite groups.

Theorem 13. (Iwasawa) Suppose G acts faithfully and primitively on Ω and that $G' = G$. Write $H = C_G(\omega)$ for a fixed ω in Ω. Suppose there is a solvable subgroup $K \leqslant H$ such that G is generated by the conjugates of K. Then G is simple.

We now provide a sketch of the classical groups over a finite field.

($A_{n-1}(F)$) Define the bilinear form $B(u,v) = 0$ so that $\mathrm{GL}(v)$ is the set of invertible transformations T that satisfy $B(Tu, Tv) = B(u, v)$ for u, v in V. We restrict attention to $\mathrm{SL}(V)$ which consists of all transformations with determinant 1, and then $\mathrm{PSL}(V) = \mathrm{SL}(V)/Z(\mathrm{SL}(v))$ is always a simple group by Iwasawa's theorem except for $\mathrm{PSL}(2,2)$ and $\mathrm{PSL}(2,3)$. We call $\mathrm{PSL}(V)$ the **projective special linear group**.

($C_n(F)$) For B a non-degenerate alternate form on the (even dimensional) vector space V we call the set of all invertible linear transformations T that satisfy $B(Tu, Tv) = B(u, v)$ for u, v in V the **symplectic group** on V, and denote it as $\mathrm{Sp}(V)$. It's possible to show that $\mathrm{Sp}(V)$ has center $\{I, -I\}$ and so it has order 2 if the characteristic of the ground field k is different from 2, and it has order 1 otherwise. We define the **projective symplectic** groups as

$$\mathrm{PSp(n,q)} = \mathrm{Sp}(V)/Z(\mathrm{Sp}(V))$$

where V has dimension n over the field $\mathbb{F}_q$.

Theorem 14. Except for $\mathrm{PSp}(2,2)$, $\mathrm{PSp}(2,3)$, and $\mathrm{PSp}(4,2)$ every projective symplectic group $\mathrm{PSp}(V)$ is simple. Moreover, if $n = 2m$ then

$$|\mathrm{Sp}(n,q)| = q^{m^2} \prod_{i=1}^{m} (q^{2i} - 1)$$

TO BE COMPLETED.

0.7 References

References are listed according to how often I consulted with them in writing this chapter.

(A) *Linear Algebra* by Hoffman and Kunze

(B) *Linear Algebra* by Serge Lang

(C) *Geometric Algebra* by Larry Grove