1 Eigenvectors and Eigenvalues

We begin linear algebra two by asking a very fundamental question: when do matrices (equivalently, linear transformations) act like scalars? As in, when does multiplying a vector by a matrix result in a scaled vector. Formally, how can we find a vector $v \in V$ such that there exists a $\lambda \in \mathbb{F}$ such that:

$$Av = \lambda v$$

Definition 1.1:

An eigenvector of a linear transformation $T : V \to V$ (equivalently, a square matrix) is a non-zero vector $v \in V$ such that there exists $\lambda \in \mathbb{F}$ where:
$$Tv = \lambda v$$

λ is called a eigenvalue of T.

Definition 1.2:

Given an eigenvalue λ of T, we define the eigenspace of λ to be:

$$V_\lambda := \{v \in V \mid Tv = \lambda v\}$$

Definition 1.3:

The spectrum, denoted spec, of a linear transformation is the set of all of its eigenvalues. Formally:

$$\text{spec}(T) = \{\lambda \in \mathbb{F} | \exists 0 \neq v \in V : Tv = \lambda v\}$$

Theorem 1.1:

Every eigenspace is a subspace of V

Proof:

Let λ be the eigenvalue.

There are two criteria for a subspace:

- $0 \in V_\lambda$: Since $T(0) = 0 = \lambda \cdot 0 \implies 0 \in V_\lambda$.
- Given $v, u \in V_\lambda$ and $\alpha \in \mathbb{F}$, $v + \alpha u \in V_\lambda$: we know $v + \alpha u \in V_\lambda \iff T(v + \alpha u) = \lambda(v + \alpha u)$. And we know $T(v + \alpha u) = Tv + \alpha Tu = \lambda v + \alpha \lambda u = \lambda(v + \alpha u)$.

∎

Notice that if λ isn't an eigenvalue of T, then $V_\lambda = \{0\}$.

Theorem 1.2:

Given a matrix A, λ is an eigenvalue of A if and only if $\det(\lambda I - A) = 0$

Proof:

We know λ is an eigenvalue of A if and only if there exists a $v \in V$ such that: $Av = \lambda v = \lambda I v \iff (\lambda I - A)v = 0$ this is if and only if the nullspace of $\lambda I - A$ is non-zero, which is if and only if $\lambda I - A$ is singular (not invertible) which is if and only if $\det(\lambda I - A) = 0$ ∎

Definition 1.4:

We denote the characteristic polynomial of a matrix A as $p_A(x)$ and define it to be: $p_A(x) := \det(xI - A)$.

Theorem 1.3:

λ is an eigenvalue of A if and only if λ is a root of $p_A(x)$

Proof:

λ is a root of $p_A(x) \iff p_A(\lambda) = 0 \iff \det(\lambda I - A) = 0 \iff \lambda$ is an eigenvalue of A ∎

Notice that *by definition* $V_\lambda = \mathrm{N}(\lambda I - A)$. This is trivial, so I won't waste the energy or space proving it.

Let's try an example:

Given the real matrix $A = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix}$ find all eigenvalues and eigenvectors.

First, let's find A's characteristic polynomial:

$$p_A(x) = \det \begin{pmatrix} x-1 & 1 \\ -2 & x-4 \end{pmatrix} = (x-1)(x-4) + 2 = x^2 - 5x + 6 = (x-3)(x-2)$$

So A has two eigenvalues: 2 and 3. So we need to find the respective nullspaces.

For $\mathrm{N}(2I - A)$:

$$\begin{pmatrix} 1 & 1 \\ -2 & -2 \end{pmatrix} \overset{-\frac{1}{2}R_2}{\longleftrightarrow} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \overset{R_2 \mapsto R_2 - R_1}{\longleftrightarrow} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$$

So $V_2 = \mathrm{span} \left\{ \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$

Similarly for 3, we get:

$$\begin{pmatrix} 2 & 1 \\ -2 & -1 \end{pmatrix} \longmapsto \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}$$

So $V_3 = \mathrm{span} \begin{pmatrix} 1 \\ -2 \end{pmatrix}$

And we see that in fact, these are eigenvectors (check for yourself!)

Note: an invertible matrix doesn't necessarily have an eigenvalue, for instance the rotational matrix:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

But all singular matrices have at least one eigenvalue: 0, since their nullspace is non-trivial.

A characteristic polynomial is in the form: $p_A(x) = x^n + a_1 x^{n-1} + \cdots + a_n$ where $a_n = -\operatorname{trace}(A), a_0 = \pm \det(A)$

Proof:

Suppose:

$$A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \cdots & \alpha_{nn} \end{pmatrix}$$

And:

$$p_A(x) = \det(xI - A)$$

Therefore: $p_A(0) = \det(0I - A) = \det(-A) = (-1)^n \det(A)$. And we know that $p_A(0) = a_n \implies a_n = \pm \det(A)$.

To understand the rest, we must focus on the permutation definition of the determinant.

To get the coefficients for x^n and x^{n-1} we have to consider every permutation that multiplies by at least n and $n-1$ elements on $xI - A$'s diagonal (elements in the form $x - \alpha$). The only permutation that does this is the identity permutation.

This is obvious for x^n, the reasoning for x^{n-1} is that any permutation that for at least $n-1$ values, $\sigma(i) = i$ (it gives $n-1$ values on the diagonal). But since a permutation is bijective, this means that the "last" permutation must remain the same. So $\sigma = I$.

So that means it suffices to just focus on the identity permutation.

So the coefficient for x^n, a_0 if you will, is the coefficient of x^n in:

$$\prod_{i=1}^{n}(x - \alpha_{ii})$$

But the coefficient for x^n in this polynomial is 1.

The coefficient for x^{n-1} is going to be:

$$-\alpha_{11}x^{n-1} - \alpha_{22}x^{n-1} - \cdots - \alpha_{nn}x^{n-1} = -x^{n-1}\sum_{i=1}^{n}\alpha_{ii} = -\operatorname{trace}(A)x^{n-1}$$

So $a_n = -\operatorname{trace}(A)$ ∎

Theorem 1.4:

Eigenvectors of different eigenvalues are linearly independent

Proof:

We will prove this by induction. Our invariant will be $P(n) = $ "n eigenvectors with different eigenvalues are linearly independent".

Base case: $n = 1$. This is trivial.

Inductive step: assume $P(n)$. Suppose we have $v_1 \ldots v_{n+1}$ eigenvectors with respective $\lambda_1 \ldots \lambda_{n+1}$ (distinct) eigenvalues.

Suppose there exists a zero linear combination of $n + 1$ eigenvectors of T:

$$\alpha_1 v_1 + \ldots \alpha_{n+1} v_{n+1} = 0$$

If we then take T on both sides:

$$\alpha_1 \lambda_1 v_1 + \cdots + \alpha_{n+1} \lambda_{n+1} v_{n+1} = 0$$

Since all of the eigenvalues are distinct, there exists a $\lambda_i \neq 0$. Without loss of generality, assume $\lambda_{n+1} \neq 0$. We can then multiply both sides of the first equation by λ_{n+1}:

$$\begin{cases} \alpha_1 \lambda_{n+1} v_1 + \cdots + \alpha_{n+1} \lambda_{n+1} v_{n+1} = 0 \\ \alpha_1 \lambda_1 v_1 + \cdots + \alpha_{n+1} \lambda_{n+1} v_{n+1} = 0 \end{cases}$$

We can then subtract the second equation from the first and get:

$$\alpha_1 (\lambda_{n+1} - \lambda_1) v_1 + \cdots + \alpha_n (\lambda_{n+1} - \lambda_n) v_n = 0$$

By our inductive assumption $\alpha_i (\lambda_{n+1} - \lambda_i) = 0$ (as $v_1 \ldots v_n$ are linearly independent). Since the eigenvalues are distinct, this means $\alpha_i = 0$ (for $1 \leq i \leq n$. So from our original zero sum:

$$\alpha_{n+1} v_{n+1} = 0$$

So $\alpha_{n+1} = 0$, so all $\alpha_i = 0$, so $v_1 \ldots v_{n+1} = 0 \implies P(n+1)$ ∎

2 Matrix Forms

2.1 Similarity

Definition 2.1:

Two matrices, M_1 and M_2 are similar (denoted $M_1 \sim M_2$) if they represent the same linear transformation. In other words, M_1 is similar to M_2 if and only if there exists a linear transformation T and two bases B and C such that:

$$M_1 = [T]_B$$
$$M_2 = [T]_C$$

Lemma 2.1.1:

Given an invertible matrix A and a basis B, there exists a basis C such that $A = [I]_C^B$

Proof:

Suppose $B = (v_1 \ldots v_n)$ we need to prove that there exists a basis C such that $\forall i : [v_i]_C = C_i(A)$.

And suppose $A = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & \ddots & \vdots \\ \alpha_{n1} & \ldots & \alpha_{nn} \end{pmatrix}$

Let C_M be the matrix form of C (C_M's columns are C's vectors).

So we need to prove there exists a basis $C = \{u_1 \ldots u_n\}$, such that $v_i = \alpha_{1i} u_1 + \cdots + \alpha_{ni} u_n = C_M \cdot C_i(A) = C_i(C_M \cdot A)$.

This means that $B_M = C_M \cdot A$, so $C_M = B_M \cdot A^{-1}$. Since B_M and A are invertible, so is C_M, therefore it represents a basis ∎

Theorem 2.1:

$A \sim B \iff$ there exists an invertible matrix P such that $A = P^{-1}BP$

Proof:

$\implies$ Suppose B and C are our bases such that $A = [T]_B$ and $B = [T]_C$. We know that:

$$[I]_B^C [T]_C [I]_C^B = [T]_B$$

And:

$$[I]_C^{B^{-1}} = [I]_B^C$$

So let:

$$P := [I]_C^B$$
$$\implies A = P^{-1}BP$$

As required.

$\impliedby$ Since B is a matrix, it represents some linear transformation. Suppose $B = [T]_C^C$
Since P is invertible, we know by the previous lemma that there exists a basis D such that $P = [I]_D^C$
(since P^{-1} is invertible, we apply the lemma to it and the inverse, P, is just $[I]_C^D$) so:

$$A = P^{-1}BP = [I]_D^C [T]_C^C [I]_C^D = [T]_D^D$$

So A represents T as well, so $A \sim B$ ∎

Theorem 2.2:

Similar matrices have the same characteristic polynomial

Proof:

This makes sense because they represent the same linear transformation, so they have the same eigenvalues.
Suppose $A \sim B$, so $A = P^{-1}BP$. So:

$$p_A(x) = \det(xI - A) = \det(xP^{-1}P - P^{-1}BP) = \det(P^{-1}(xI - B)P) =$$

$$= \det(P^{-1})\det(xI - B)\det(P) = \det(xI - B) = p_B(x) \quad ∎$$

This is also a rigorous proof to what I said at the beginning of the proof. Since similar matrices share a characteristic polynomial, they share the same eigenvalues (since eigenvalues are roots to the characteristic polynomial).

Statement:

$A \sim B \implies \operatorname{trace}(A) = \operatorname{trace}(B)$

Proof:

Since $A \sim B$, there exists an invertible matrix such that $A = P^{-1}BP$ so:

$$\operatorname{trace}(A) = \operatorname{trace}(P^{-1}BP)$$

We know that for two matrices, $\operatorname{trace}(AB) = \operatorname{trace}(BA)$ so:

$$\operatorname{trace}(P^{-1}BP) = \operatorname{trace}(BPP^{-1}) = \operatorname{trace}(B) \quad ∎$$

2.2   Diagonalization

Definition 3.1:

>   A matrix A is called diagonalizable if it is similar to a diagonal matrix.

Definition 3.2:

>   A linear transformation is diagonalizable if one of its representations is a diagonal matrix.

By the definition of similarity (definition 2.1) a linear transformation is diagonalizable if and only if a representation of it is diagonalizable (which is equivalent to every representation being diagonalizable).

Lemma 3.1.1:

>   A diagonal matrix's eigenvalues are the values on its diagonal, and its eigenvectors are e_i

Proof:

>   Let $A = \begin{pmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{pmatrix}$
>   Therefore $p_A(x) = (x - \alpha_1) \cdots (x - \alpha_n)$, so its eigenvalues are α_i.
>   Also note that $Ae_i = \alpha_i e_i$, so e_i are eigenvectors (and every other eigenvector is some linear combination of them). ∎

Theorem 3.1:

>   A linear transformation is diagonalizable if and only if there exists a basis of its eigenvectors

Proof:

> $\implies$ Since T is diagonalizable, it has a representation, $[T]_B^B$ that is a diagonal matrix. Since it's diagonal it has a eigenvectors $e_1 \ldots e_n$. So T has eigenvalues of $v_1 \ldots v_n$ where $[v_i]_B = e_i$, and since $\{e_i\}_{i=1}^n$ are linearly independent, so is v_i (in fact, these eigenvectors are the vectors that make up B), so they form a basis.
>
> $\impliedby$ Let $B = (v_1 \ldots v_n)$ be the basis of eigenvectors. This means that $Tv_i = \lambda v_i$. So $C_i([T]_B) = [Tv_i]_B = [\lambda v_i]_B = \lambda e_i$. So $[T]_B$ is a diagonal matrix ∎

Corollary 1:

>   A matrix is diagonalizable if and only if there exists a basis of eigenvectors

Proof:

> A is diagonalizable $\iff$ $A \sim D$ where D is a diagonal matrix $\iff$ A and D are both represented by the same linear transformation T $\iff$ T is diagonalizable $\iff$ there exists a basis of eigenvectors of T $\iff$ there exists a basis of eigenvectors for A ∎

Corollary 2:

>   A linear transformation (and by extension, a matrix), T, is diagonalizable iff $\displaystyle\sum_{\lambda \in \mathrm{spec}(T)} \dim V_\lambda = \dim V$

Proof:

> $\implies$: T is diagonalizable so there exists a basis of eigenvectors, let it be B. We can partition B into subsets of eigenvectors with the same eigenvalue. Suppose:
>
> $$B = B_1 \sqcup \cdots \sqcup B_t$$
>
> B_i must be a basis for V_i since it's linearly independent and it must span V_i since if not, we can add the vector not in its span to the basis B and get a linearly independent set that's larger than B which would be a contradiction.
>   So $|B| = \displaystyle\sum_{i=1}^t |B_i| \implies \dim V = \sum_{\lambda \in \mathrm{spec}} \dim V_\lambda$

$\impliedby$: Since eigenvectors with different eigenvalues are linearly independent, that means that $\bigoplus_{\lambda \in \text{spec}(T)} V_\lambda$.

So:

$$\dim\left(\bigoplus_{\lambda \in \text{spec}(T)} V_\lambda\right) = \sum_{\lambda \in \text{spec}(T)} \dim V_\lambda = \dim V$$

So any basis of $\bigoplus_{\lambda \in \text{spec}(T)} V_\lambda$ is a basis of V and since any basis of $\bigoplus_{\lambda \in \text{spec}(T)} V_\lambda$ is a basis of eigenvectors, that means V has a basis of eigenvectors

$\blacksquare$

Notice that this gives us a handy algorithm for finding the matrix that diagonalizes a matrix. Given the matrix A with a basis $B = (v_1 \ldots v_n)$ of eigenvectors, we define:

$$P = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix}$$

Now notice that:

$$C_i(P^{-1}AP) = P^{-1}A \cdot C_i(P) = P^{-1}Av_i = \lambda_i P^{-1} v_i = \lambda_i P^{-1} C_i(P) = \lambda_i C_i(P^{-1}P) = \lambda_i C_i(I)$$

So

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Which is a diagonal matrix.

How did we get to this? Well, notice that if A is the representation of T on the basis C ($A = [T]_C$) and since T is diagonalizable, then there exists a basis of eigenvectors B such that $[T]_B$ is a diagonal matrix.

Now remember that:

$$[T]_B = [I]_B^C [T]_C^C [I]_C^B$$

So what we really want to find is $[I]_C^B$. So:

$$C_i([I]_C^B) = [b_i]_C$$

Which is just an eigenvector of T written relative to the basis of A. So $[b_i]_C$ is an eigenvector of A (this is a simple proof: $A[b_i]_C = [T]_C^C [b_i]_C = [Tb_i]_C = [\lambda b_i]_C = \lambda [b_i]_C$).

So we see that $[I]_C^B$ is just a matrix whose columns are $A's$ eigenvectors. So we define $P = [I]_C^B$.

Let's now refocus on characteristic polynomials.

Definition 3.3:

The algebraic multiplicity of an eigenvalue is its multiplicity in the characteristic polynomial. We denote the algebraic multiplicity of λ, an eigenvalue of T as: $\mu_T(\lambda)$ (I may also use similar notation, like μ_λ or the like). Formally:

$$\mu_T(\lambda) := \max\left\{k \mid (x-\lambda)^k | p_T(x)\right\}$$

Definition 3.4:

The geometric multiplicity of an eigenvalue is the dimension of its eigenspace. It is denoted as $\gamma_A(\lambda)$. Formally:

$$\gamma_T(\lambda) = \dim(V_\lambda) = \dim \text{N}(T - \lambda I)$$

The following lemma should be something you're familiar with from linear algebra 1, but in order to make sure you're up to date and whatnot, there's no harm in having it here.

Suppose A, B are two square matrices (not necessarily of the same size) and C is some other (not necessarily square) matrix. Then $\det \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array} \right) = \det(A) \cdot \det(B)$

Proof:

Suppose $A \in \mathbb{F}^{n \times n}, B \in \mathbb{F}^{m \times m}$ $(C \in \mathbb{F}^{n \times m})$.
And let:
$$M := \left(\begin{array}{c|c} A & C \\ \hline 0 & B \end{array} \right) \in \mathbb{F}^{n+m \times n+m}$$

So by the permutation definition of the determinant:

$$\det(M) = \sum_{\sigma \in S_{n+m}} \text{sign}(\sigma) \cdot \prod_{i=1}^{n+m} [M]_{i\sigma(i)}$$

The idea is to prove that for any $\sigma \in S_{n+m}$ where if $\exists n < i \leq n+m$ such that $1 \leq \sigma(i) \leq n$ then there exists a $1 \leq j \leq n$ such that $n < \sigma(j)$ (so $[M]_{j\sigma(j)} = 0$ and the product of that permutation is 0).
So assume for the sake of a contradiction the contrary:

$$\forall 1 \leq j \leq n : 1 \leq \sigma(j) \leq n$$

This means that $\forall k > n : \sigma(k) > n$ since σ is bijective and therefore injective. (Think of it like this: all of A is taken, so everything past it must be in B.) But this directly contradicts that $n < i \leq n+m$ and $\sigma(i) \leq n$ ⚡ ∎

Theorem 3.2:

$$1 \leq \gamma_\lambda \leq \mu_\lambda \leq n$$

Proof:

Since λ is an eigenvalue, it has a corresponding eigenvector, say v, so $v \in V_\lambda$, so $\gamma_\lambda = \dim V_\lambda \geq 1$.
Now, suppose $\gamma_\lambda = k$, so there exists k linearly independent eigenevectors $v_1 \ldots v_k$. If we expand this to a basis $(v_1 \ldots v_k, \hat{v}_{k+1} \ldots \hat{v}_n)$. So:

$$P := \begin{pmatrix} | & & | \\ v_1 & \cdots & \hat{v}_n \\ | & & | \end{pmatrix}$$

Is invertible. And:

$$AP = \begin{pmatrix} | & & | & & | \\ \lambda v_1 & \cdots & \lambda v_k & \cdots & A\hat{v}_n \\ | & & | & & | \end{pmatrix}$$

Also notice that $P^{-1}v_i = P^{-1}C_i(P) = C_i(P^{-1}P) = C_i(I) = e_i$. So:

$$P^{-1}AP = \begin{pmatrix} | & & | & & | \\ \lambda e_1 & \cdots & \lambda e_k & \cdots & P^{-1}A\hat{v}_n \\ | & & | & & | \end{pmatrix} = \left(\begin{array}{c|c} \lambda I_k & * \\ \hline 0 & * \end{array} \right)$$

Let this matrix be M. So $A \sim M \implies p_A(x) = p_M(x)$.
And we know that:

$$p_M(x) = \det(xI - M) = \det \left(\begin{array}{c|c} (x-\lambda)I_k & * \\ \hline 0 & * \end{array} \right) = \det((x-\lambda)I_k)\det(*) = (x-\lambda)^k s(x)$$

For some polynomial $s(x)$.
So the multiplicity of λ in $p_M(x)$ and therefore $p_A(x)$ is at least k (since we know almost nothing about $s(x)$).
So $\mu_\lambda \geq \gamma_\lambda$ ∎

Theorem 3.3:

A linear transformation, T, is diagonalizable iff the algebraic multiplicity of every eigenvalue is equal to its geometric multiplicity and $p_T(x)$ can be fully factorized

Proof:

$\implies$: We know that T is diagonalizable if and only if:

$$\sum_{\lambda \in \text{spec}(T)} \gamma_T(\lambda) = \dim V$$

We also know that the degree of the characteristic polynomial is $\dim V$ and the sum of the algebraic multiplicities must be less than or equal to the degree. So:

$$\sum_{\lambda \in \text{spec}(T)} \mu_T(\lambda) \leq \dim V$$

By the previous theorem, theorem 3.2:

$$\dim V = \sum_{\lambda \in \text{spec}(T)} \gamma_T(\lambda) \leq \sum_{\lambda \in \text{spec}(T)} \mu_T(\lambda) \leq \dim V$$

Which means that:

$$\sum_{\lambda \in \text{spec}(T)} \mu_T(\lambda) = \sum_{\lambda \in \text{spec}(T)} \gamma_T(\lambda) = \dim V$$

Which can only be true if $\forall \lambda \in \text{spec}(T) : \gamma_T(\lambda) = \mu_T(\lambda)$ (since otherwise, the sum of geoemtric multiplicities would be less than the sum of algebraic multiplicities).
This also means that $p_T(x)$ is fully factorizable, since the sum of the alegbraic multiplicities equals the degree of $p_T(x)$.

$\impliedby$: Since $p_T(x)$ is fully factorizable:

$$\sum_{\lambda \in \text{spec}(T)} \mu_T(\lambda) = \deg(p_T(x)) = \dim V$$

So:

$$\sum_{\lambda \in \text{spec}(T)} \gamma_T(\lambda) = \dim V$$

Which means T is diagonalizable (by the corollary of theorem 3.1)

$\blacksquare$

2.3 Triangularization

Notice that if $p_T(x)$ is fully factorizable, it doesn't mean that T is diagonalizable. But does it have any other significance?

It turns out that it does.

Definition 4.1:

    A matrix A is triangularizable if it is similar to an upper triangle matrix

Theorem 4.1:

    A linear transformation T is triangularizable $\iff$ $p_T(x)$ is fully factorizable

Proof:

$\implies$: Since T is triangularizable, there exists a basis B such that:

$$[T]_B = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

So:

$$p_T(x) = \det \begin{pmatrix} x - \lambda_1 & & * \\ & \ddots & \\ 0 & & x - \lambda_n \end{pmatrix} = (x - \lambda_1) \cdots (x - \lambda_n)$$

Which is fully factorized.

$\impliedby$: We will prove this by induction on the dimension of V.

Let:

$P(n) = $ "Given a vector space V where $\dim V = n$, and a linear transformation T on V, if $p_T(x)$ is fully factorizable, then T is triangularizable"

Base case: $n = 1$: This means that given any basis B, $[T]_B = (\lambda)$ which is an upper triangular matrix, so its characteristic polynomial is fully factorizable.

Inductive step: Assume $P(n)$. Let V be a vector space such that $\dim V = n + 1$.

Since $p_T(x)$ is fully factorizable, $\exists \lambda_1 \in \mathbb{F} : p_T(\lambda_1) = 0$. Let v_1 be an eigenvector of eigenvalue λ_1. Let $B = (v_1, \hat{v}_1 \ldots \hat{v}_n)$ the completion of (v_1) to a basis. So:

$$[T]_B = \begin{pmatrix} \lambda_1 & \text{---}\,*\,\text{---} \\ | & \\ 0 & A \\ | & \end{pmatrix}$$

Where $A \in \mathbb{F}^{n \times n}$.

Since $p_T(x) = \det(xI - [T]_B) = (x - \lambda_1)\det(xI - A) = (x - \lambda_1)p_A(x)$. Since $p_T(x)$ is fully factorizable, $p_A(x)$ is fully factorizable. By $P(n)$ this means it is triangularizable. So there exists a matrix P' such that $P'^{-1}AP' = U'$ where U' is an upper triangle matrix.

Let $P := (1) \oplus P'$. This means that $P^{-1} = (1) \oplus P'^{-1}$. So:

$$P^{-1}[T]_B P = \begin{pmatrix} 1 & \text{---}\,0\,\text{---} \\ | & \\ 0 & P'^{-1} \\ | & \end{pmatrix} \begin{pmatrix} \lambda_1 & \text{---}\,*\,\text{---} \\ | & \\ 0 & A \\ | & \end{pmatrix} \begin{pmatrix} 1 & \text{---}\,0\,\text{---} \\ | & \\ 0 & P' \\ | & \end{pmatrix} = \begin{pmatrix} \lambda_1 & \text{---}\,*\,\text{---} \\ | & \\ 0 & P'^{-1}AP \\ | & \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 & \text{---}\,*\,\text{---} \\ | & \\ 0 & U' \\ | & \end{pmatrix}$$

Which is an upper triangle matrix, let's denote it as U. So $[T]_B \sim U$, which means T is triangularizable

∎

Any linear transformation above $\mathbb{C}$ is triangularizable

Proof:

Let be T be a linear transformation above $\mathbb{C}$. Since all polynomials above $\mathbb{C}$ are fully factorizable, $p_T(x)$ is fully factorizable, which means T is triangularizable.

Statement:

Notice that this gives us a recursive algprithm for triangularizing matrices. Say you're given a matrix A:

1. Find an eigenvector to A, let it be v.

2. Complete v to a basis $(v, u_2 \ldots u_n)$ and define the matrix M columns are vectors in B (where v is first).

3. Notice that:

$$M^{-1}AM = \begin{pmatrix} \lambda & \overline{}\ *\ \overline{} \\ | & \\ 0 & A' \\ | & \end{pmatrix}$$

Where $A' \in \mathbb{F}^{n-1 \times n-1}$

4. Execute the algorithm on A' to get a matrix P' that triangularizes A'

5. $P := M \cdot ((1) \oplus P')$ will be the matrix that triangularizes A

Proof:

The idea behind this algorithm comes from the recursive nature of the proof of the above theorem.
We will quickly prove that this algorithm works. First off, it obviously terminates since each step gives you a smaller matrix.
Notice that:

$$P^{-1} = ((1) \oplus P'^{-1}) \cdot M^{-1}$$

$$\implies P^{-1}AP = ((1) \oplus P'^{-1}) \cdot M^{-1} \cdot A \cdot M \cdot ((1) \oplus P') =$$

$$= ((1) \oplus P'^{-1}) \cdot \begin{pmatrix} \lambda & \overline{}\ *\ \overline{} \\ | & \\ 0 & A' \\ | & \end{pmatrix} \cdot ((1) \oplus P') = \begin{pmatrix} \lambda & \overline{}\ *\ \overline{} \\ | & \\ 0 & P^{-1}A'P \\ | & \end{pmatrix}$$

Since $P'^{-1}A'P'$ is an upper triangle matrix, $P^{-1}AP$ is an upper triangle matrix.

2.4 Zeroing Polynomials

Definition 5.1:

Given a polynomial $p(x) = \sum_{i=0}^{d} a_i x^{d-i}$ we define $p(A) := \sum_{i=0}^{d} a_i A^{d-i}$ where $A^0 = I$

Lemma 5.1.1:

For every matrix, there exists a polynomial $0 \neq p(x)$ such that $p(A) = 0$

Proof:

If we define $S = \left\{ A^i \mid 0 \leq i \leq n^2 \right\}$ over the vector space $\mathbb{F}^{n \times n}$, it must be linearly dependent since $|S| = n^2 + 1 > \dim \mathbb{F}^{n \times n}$. So there exists scalars $\alpha_0 \ldots \alpha_{n^2}$ (where at least one is non-zero) such that:

$$\sum_{i=0}^{n^2} \alpha_i A^{n^2 - i} = 0$$

So if we define:

$$p(x) = \sum_{i=0}^{n^2} \alpha_i x^{n^2 - i}$$

We get that $p(A) = 0$ $\blacksquare$

13

$$p_A(A) = 0$$

Proof:

Suppose $A \in \mathbb{F}^{n \times n}$. This means:

$$(xI - A) \cdot \mathrm{adj}(xI - A) = \det(xI - A)I = p_A(x)I$$

Let $B := \mathrm{adj}(xI - A)$. Since $[B]_{ij} = (-1)^{i+j} \det((xI - A)_{ji})$, B is a matrix of polynomials up to degree $n - 1$ (since removing a row and column from $xI - A$ and taking the determinant means that the max polynomial degree is one less than the max polynomial degree of $xI - A$). Suppose:

$$B = x^{n-1}B_{n-1} + \cdots + xB_1 + B_0 = \sum_{i=0}^{n-1} B_i x^i$$

So:

$$(xI - A)B = (xI - A) \sum_{i=0}^{n-1} B_i x^i = \sum_{i=0}^{n-1} B_i x^{i+1} - \sum_{i=0}^{n-1} AB_i x^i = B_{n-1}x^n + \sum_{i=1}^{n-1}(B_{i-1} - AB_i)x^i - AB_0$$

And suppose:

$$p_A(x) = \sum_{i=0}^{n-1} c_i x^i + x^n$$

So:

$$p_A(A) = \sum_{i=0}^{n-1} c_i A^i + A^n$$

This means that:

$$B_{n-1}x^n + \sum_{i=1}^{n-1}(B_{i-1} - AB_i)x^i - AB_0 p_A(x) = x^n I + \sum_{i=0}^{n-1} c_i x^i I$$

So:

$$B_{n-1} = I$$

$$B_{i-1} - AB_i = c_i$$

$$c_0 I = -AB_0$$

$$\implies p_A(A) = A^n + \sum_{i=1}^{n-1}(B_{i-1} - AB_i)A^i - AB_0$$

$$\sum_{i=1}^{n-1}(B_{i-1} - AB_i)A^i = \sum_{i=1}^{n-1} B_{i-1}A^i - \sum_{i=1}^{n-1} B_i A^{i+1} = \sum_{i=1}^{n-1} B_{i-1}A^i - \sum_{i=2}^{n} B_{i-1}A^i = B_0 A - B_{n-1}A^n$$

Since $B_{n-1} = I$, we see that this equals $B_0 A - A^n$

$$\implies p_A(A) = A^n + AB_0 - A^n - AB_0 = 0 \quad \blacksquare$$

Definition 5.2:

A simplified polynomial is a polynomial whose leading coefficient is 1

Definition 5.3:

An irreducible polynomial is a simplified polynomial that cannot be written as the product of two polynomials of smaller degree.

Every simplified polynomial can be written as a unique product of irreducible polynomials (up to order)

Proof:

- Existence: We will prove this by strong induction. Let: $P(n) = $ "A polynomial of degree n can be written as a product of irreducible polynomials"

  Base case: $n = 1$: This means $P(x) = x - \alpha$, which is irreducible $\implies P(1)$

  Inductive step assume $P(1) \ldots P(n)$ we must prove $P(n+1)$. Let $P(x)$ be a simplified polynomial of degree $n + 1$. If $P(x)$ is irreducible, then it is a product of irreducible polynomials (itself). Otherwise, it is the product of two simplified polynomials of a smaller degree, let them be $q(x), r(x)$. By our inductive assumption, $r(x)$ and $q(x)$ are products of irreducible polynomials $\implies P(x)$ is a product of irreducible polynomials.

- Uniqueness: Assume there exists polynomials which have two factorizations of irreducible polynomials. This means there exists a minimum degree where there exists a polynomial with two distinct factorizations. Let this degree be d. Let $P(x)$ be a polynomial of degree d with two distinct factorizations. Suppose:

$$P(x) = \begin{cases} f_1 \cdots f_t \\ g_1 \cdots g_s \end{cases}$$

Where f_i, g_i are irreducible.

Since these two factorizations are distinct, $\forall i, j : f_i \neq g_j$, since otherwise, $\frac{P(x)}{f_i} = \frac{P(x)}{g_j}$ has two distinct factorizations, while their degree is less than $P(x)$'s.

Let:

$$Q = f_2 \cdots f_t$$

$$R = g_2 \cdots g_s$$

$$\implies f_1 Q = g_1 R \implies (g_1 - f_1) Q = f_1 (P - Q)$$

Since f_1 doesn't divide $g_1 - f_1$ (otherwise, $\frac{g_1}{f_1} - 1$ would be a polynomial, so $g_1 = f_1$ in contradiction)

So f_1 must divide Q, so f_1 appears in the factorization of Q, which is a contradiction since $q_2 \ldots q_s \neq f_1$ ⨍

Note that this proof is almost the same as the proof for prime factorization.

Lemma 5.2.2:

If f is a zero polynomial of A (ie. $f(A) = 0$), then $p_A(x) \mid f(x)^n$

Proof:

Let $d := \deg(f)$. We want a matrix polynomial:

$$B(x) = \sum_{i=0}^{d-1} x^i B_i$$

Such that $(xI - A)B(x) = f(x)I$ (this implies that $p_A(x) \det B(x) = \det(f(x)I) = f(x)^n$ so $p_A(x)$ divides $f(x)^n$)

This means that:

$$(xI - A)\sum_{i=0}^{d-1} x^i B_i = \sum_{i=0}^{d-1} x^{i+1} B_i - \sum_{i=0}^{d-1} x^i A B_i = \sum_{i=1}^{d} x^i B_{i-1} - \sum_{i=0}^{d-1} x^i A B_i$$

$$= x^d B_{d-1} + \sum_{i=1}^{d-1} x^i B_{i-1} - \sum_{i=1}^{d-1} x^i A B_i - A B_0 = x^d B_{d-1} + \sum_{i=1}^{d-1} x^i (B_{i-1} - A B_i) - A B_0$$

Suppose:

$$f(x)I = x^d I + \sum_{i=0}^{d-1} \beta_i x^i I$$

This is true only if:

$$B_{d-1} = I$$

$$\forall 1 \le i \le d - 1 : B_{i-1} - A B_i = \beta_i I$$

$$-A B_0 = \beta_0 I$$

So we define:

$$B_{d-1} = I$$

$$B_{i-1} = \beta_i I + A B_i$$

But this doesn't necessarily mean $-A B_0 = \beta I$.

Since $f(A) = 0$:

$$0 = A^d + \sum_{i=0}^{d-1} \beta_i A^i = A^d + \sum_{i=1}^{d-1} A^i (B_{i-1} - A B_i) + \beta_0 I = A^d + \sum_{i=1}^{d-1} A^i B_{i-1} - \sum_{i=1}^{d-1} A^{i+1} B_i + \beta_0 I =$$

$$= A^d - \sum_{i=2}^{d} A^i B_{i-1} + \sum_{i=1}^{d-1} A^i B_{i-1} + \beta_0 I =$$

$$= A^d - A^d B_{d-1} - \sum_{i=2}^{d-1} A^i B_{i-1} + \sum_{i=2}^{d-1} A^i B_{i-1} + B_0 A + \beta_0 I = A^d (I - B_{d-1}) + A B_0 + \beta_0 I = A B_0 + \beta_0 I$$

Which means that:

$$0 = A B_0 + \beta_0 I \implies -A B_0 = -\beta_0 I \quad \blacksquare$$

Definition 5.4:

The minimal polynomial of a matrix is a simplified zeroing polynomial with the minimum possible degree. We denote it as $m_A(x)$

So, for example, given a nilpotent matrix, A with nilpotent degree of k (ie. $A^k = 0$ and $A^{k-1} \ne 0$)), the minimal polynomial of A is x^k

Lemma 5.2.3:

The minimal polynomial is unique

Proof:

Assume that $m'_A(x)$ is also a minimal polynomial. So it has the same degree as $m_A(x)$ Since both $m_A(x), m'_A(x)$ are simplfied, $\deg(m_A(x) - m'_A(x)) < m_A(x)$. Let $p(x)$ be the simplified form of $m_A(x) - m'_A(x)$ (divide it by its leading coefficient). So:

$$p(A) = m_A(A) - m'_A(A) = 0 - 0 = 0$$

So $p(x)$ is also a zeroing polynomial. Since $\deg(p) < \deg(m_A)$, it must be the zero polynomial. So $m_A - m'_A = 0 \implies m_A = m'_A$ ∎

Lemma 5.2.4:

Given a zeroing polynomial $p(x)$, $m_A \mid p$

Proof:

We know that there exists some polynomial $q(x)$ and $r(x)$ such that $\deg(r) < \deg(m_A)$ such that:

$$p(x) = m_A(x)q(x) + r(x)$$

This means that $(p - m_A \cdot q)(x) = r(x) \implies r(A) = 0$, which can only be true if $r(x) = 0$. So $m_A \mid p$ ∎

Theorem 5.2:

$m_A(x)$ and $p_A(x)$ have the same factorization up to multiplicity

Proof:

Suppose $m_A(x) = f_1^{k_1} \cdots f_t^{k_t}$, where f_i is an irreducible polynomial.
Suppose $p_A(x) = g_1^{l_1} \cdots g_s^{l_s}$.
We know:

$$m_A(x) \mid p_A(x) \mid m_A(x)^n$$

So every f_i must appear in the factorization of $p_A(x)$. Furthermore, if we suppose $f_i = g_i$, then $k_i \leq l_i$). And $k \leq s$.
And every g_i must appear in the factorization of $m_A(x)^n$, so for every g_i, there exists an f_j equal to it. So:

$$\forall i \exists j : f_i = g_j \wedge g_i = f_j$$

So the factorizations are the same (just the multiplicity of every irreducible polynomial in $p_A(x)$'s factorization is greater than or equal to its multiplicity in $m_A(x)$'s.

Lemma 5.3.1:

$A \sim B \implies (f(A) = 0 \implies f(B) = 0)$

Proof:

Let $f(x)$ be a zeroing polynomial of A. Suppose:

$$f(x) = \sum_{i=0}^{d} \alpha_i x^i$$

So:

$$\sum_{i=0}^{d} \alpha_i A^i$$

And suppose: $B = P^{-1}AP$. So:

$$f(B) = \sum_{i=0}^{d} \alpha_i P^{-1}AP = P^{-1}\left(\sum_{i=0}^{d} \alpha_i A\right)P = P^{-1} \cdot 0P = 0 \quad ∎$$

Theorem 5.3:

The minimal polynomials of two similar matrices are equal

Proof:

Suppose $A \sim B$. So by the previous lemma: $m_A(B) = 0, m_B(A) = 0$, so $m_B \mid m_A \mid m_B \implies m_A = m_B$ ∎

Definition 5.5:

The mimimal polynomial of a linear transformation is the minimal polynomial of one of its matrix representations.

By the previous theorem, this is well-defined, since the minimal polynomial of similar matrices are the same.

2.5 Invariant Subspaces

Definition 6.1:

> A subspace $U \leq V$ is invariant under a linear transformation T if $T(U) \leq U$

So if U is invariant under T, $T\mid_U : U \to U$ defined like so:

$$T\mid_U (u) = Tu$$

(T reduced to U) is a valid linear transformation

> Note that eigenspaces are invariant.

Lemma 6.1.1:

> If $V_1 \dots V_t$ are invariant subspaces such that $V = V_1 \oplus \dots V_t$ then given any linear transformation T over V, $\operatorname{Im} T = \operatorname{Im} T_1 \oplus \dots \operatorname{Im} T_t$ and $\ker T = \ker T_1 \oplus \cdots \oplus \ker T_t$ where $T_i := T\mid_{V_i}$

Proof:

> Let B_i be a basis of V_i.
> Since $\operatorname{Im} T_{V_i}, \ker T_{V_i} \subseteq V_i$, since V_i are invariant, and $\operatorname{Im} T_{V_i} \cap \operatorname{Im} T_{V_j}, \ker T_{V_i} \cap \ker T_{V_j} = \{0\}$ for $i \neq j$.
>
> - Suppose $v \in \ker T$, so there exists $v_i \in V_i$ such that $v = v_1 + \cdots + v_t$. So $Tv = Tv_1 + \cdots + Tv_t = 0$, assume for the sake of a contradiction that $Tv_1 \dots Tv_k \neq 0$, so that means that $Tv_k = -Tv_1 - \cdots - Tv_{k-1}$ since $Tv_i \in V_i$ this means that $Tv_k \in V_1 \oplus \cdots \oplus V_{k-1}$ which means that V_k and $V_1 \oplus \cdots \oplus V_{k-1}$ aren't disjoint, in contradiction.
>
> - Suppose $Tv \in \operatorname{Im} T$, and so $\exists v_i \in V_i : Tv = v_1 + \cdots + v_t$.
>   Since $v \in V_1 \oplus + \cdots + V_t$ there exist $u_i \in V_i$ such that $v = u_1 + \cdots + u_t$, which means $Tv = Tu_1 + \cdots + Tu_t$. Since the sum is direct, every sum is unique, which means that $v_i = Tu_i \implies v_i \operatorname{Im} T \cap V_i \implies v_i \operatorname{Im} T_i$

∎

Lemma 6.1.2:

> $V_1 \dots V_t$ are invariant subspaces with bases B_i, such that $V = V_1 \oplus \cdots \oplus V_t$. If we define the basis $B := B_1 \sqcup \cdots \sqcup B_t$, then:
> $$[T]_B = \begin{pmatrix} \boxed{A_1} & & \\ & \ddots & \\ & & \boxed{A_t} \end{pmatrix}$$

Proof:

> Let v_j be the j-th vector in B.
> Let $a_0 = 0, a_i := \dim V_i$. So the elements between (inclusive) the indexes $a_{i-1} + 1 \dots a_i$ in B are the vectors in B_i. Notice that:
> $$\forall a_{i-1} < j \leq a_i : Tv_j \in V_i \implies Tv_j \in \operatorname{span}(B_i) = \operatorname{span}\left(v_{a_{i-1}+1} \cdots v_{a_i}\right) \implies [Tv_j]_B \in \operatorname{span}\left(e_{a_{i-1}+1} \cdots e_{a_i}\right)$$
> So these vectors create a block.

Our goal in this section is to find invariant subspaces $V_1 \dots V_t$ such that:

$$V = V_1 \oplus \cdots \oplus V_t$$

> Note that if T is diagonalizable, then its eigenspaces satisfy this.
> If T isn't diagonalizable, then its V_λs wont be sufficiently "large" enough to "cover" V. But remember that:

$$V_\lambda = \operatorname{N}(\lambda I - T)$$

And this is a subspace of $\operatorname{N}((\lambda I - T)^k)$ forall natural k. So perhaps if we take this to the extreme, with $\operatorname{N}((\lambda I - T)^n)$ where $n := \dim V$, it will suffice? It turns out it does.

Given a linear transformation with an eigenvalue λ, we define the generalized eigenspace of λ to be:

$$K_\lambda := \mathrm{N}\left((\lambda I - T)^n\right)$$

Where $n := \dim V$

Our goal is to now prove that these subspaces satisfy our need.

Definition 6.3:

Given T a linear transformation, $0 \neq v \in V$, we define v's path to be:

$$E_v = \left(T^{m-1}v, T^{m-2}v, \cdots Tv, v\right)$$

Where $\forall 0 \leq t < m : T^t v \neq 0$ and $T^m v = 0$
We say E_v has a length of m.

Note that not every vector (or even linear transformation) has a path. For example, any isomorphic linear transformation doesn't have a path (since $T^m v$ will never be 0).

Lemma 6.1.3:

A path is linearly independent

Proof:

Assume $E_v = \left(T^{m-1}v \ldots Tv, v\right)$ is a path of length m.
Suppose:

$$\sum_{i=0}^{m} \alpha_i T^i v = 0$$

$$\implies \sum_{i=0}^{m} \alpha_i T^{m-1+i} v = 0$$

But $\forall t \geq m : T^t v = 0$, as the length of E_v is m. So:

$$\alpha_0 T^{m-1} v = 0$$

Since $T^{m-1}v \neq 0 \implies \alpha_0 = 0$
If we take T^{m-2} we see:

$$\sum_{i=1}^{m} \alpha_i T^{m-2+i} v = 0$$

$$\implies \alpha_1 T^{m-1} v = 0 \implies \alpha_1 = 0$$

If we continue doing this, we see that $\forall i : \alpha_i = 0$ which means E_v is linearly indpendent ∎

Lemma 6.1.4:

$\forall t \in \mathbb{N} : (T - \lambda I)^t v = 0 \implies (T - \lambda I)^n v = 0$

Proof:

This is trivial if $v = 0$, otherwise:

- If $t \leq n$ then this is trivial. This means $n - t \geq 0$, so $(T - \lambda I)^n = (T - \lambda I)^t (T - \lambda I)^{n-t} = 0(T - \lambda I)^{n-t} = 0$

- Else if $t > n$, so there exists some $m \leq l : (T - \lambda I)^m v = 0, (T - \lambda I)^{m-1} v \neq 0$ (this requires $v \neq 0$). So we form a path:
$$E_v = \left((T - \lambda I)^{m-1} \ldots (T - \lambda I)v\right)$$

  Since E_v is linearly independent by our previous lemma, that means $m \leq n$ and $(T - \lambda I)^m = 0 \implies (T - \lambda I)^n = 0$ ∎

Lemma 6.1.5:

If E_v is a path, span(E_v) is invariant under T

Proof:

Let $W := \text{span}(E_v)$, and m be the length of E_v.

$$\implies T(W) = \text{span}(T(E_v)) = \text{span}(T^m v, T^{m-1} v \dots T v) = \text{span}(T^{m-1} v \dots T v) =$$

$$= \text{span}(E_v \setminus \{v\}) \subseteq \text{span}(E_v) = W \quad \blacksquare$$

Lemma 6.1.6:

Given a linear transformatio, T, and a polynomial $p(x)$, $T \circ f(T) = f(T) \circ T$

Proof:

Suppose:

$$p(x) = \sum_{i=0}^{d} \alpha_i x^i$$

$$\implies p(T) = \sum_{i=0}^{d} \alpha_i T^i$$

$$\implies T(p(T)) = T(\sum_{i=0}^{d} \alpha_i T^i) = \sum_{i=0}^{d} \alpha_i T^{i+1}$$

$$p(T)(T) = (\sum_{i=0}^{d} \alpha_i T^i)(T) = \sum_{i=0}^{d} \alpha_i T^{i+1}$$

$$\implies p(T) \circ T = T \circ p(T) \quad \blacksquare$$

Corollary 1:

K_λ is invariant under any polynomial of T

Proof:

First, we'll prove it's invariant under T. Suppose $v \in K_\lambda$:

$$\implies (\lambda I - T)^n T v = T(\lambda I - T)^n v = T0 = 0$$

So $Tv \in K_\lambda$ as required.
We will prove the corollary by induction. $P(m) = "K_\lambda$ is invariant under any polynomial of degree m".
Base case: $m = 1$, so $p(T) = \alpha T - \beta$. Suppose $v \in K_\lambda$ $p(T)v = \alpha Tv - \beta v \in K_\lambda$ as required.
Inductive step: Assume $P(m)$, and $p(T) = \sum_{i=0}^{m+1} \alpha_i T^i$. Suppose $v \in K_\lambda$.

$$\implies p(T)v = \sum_{i=0}^{m+1} \alpha_i T^i v = \sum_{i=1}^{m+1} \alpha_i T^i v + \alpha_0 v = T(\sum_{i=0}^{m} \alpha_{i+1} T^i v) + \alpha_0 v$$

Since $\sum_{i=0}^{m} \alpha_{i+1} T^i$ is a polynomial of degree m, by $P(m)$, $\sum_{i=0}^{m} \alpha_{i+1} T^i v \in K_\lambda$, and by $P(1) \implies T(\sum_{i=0}^{m} \alpha_{i+1} T^i) \in K_\lambda$. So:

$$T(\sum_{i=0}^{m} \alpha_{i+1} T^i v) + \alpha_0 v \in K_\lambda \quad \blacksquare$$

Lemma 6.1.7:

Suppose $0 \neq v \in K_\lambda$, and $\mu \neq \lambda$, so $(\mu I - T)v \in K_\lambda$ and $(\mu I - T)v \neq 0$ (which means $v \notin V_\mu$)

Proof:

Since $\mu I - T$ is a polynomial of Tm by our previous lemma, K_λ is closed under $\mu I - T$, so $(\mu I - T)v \in K_\lambda$.
Assume for the sake of a contradiction that $(\mu I - T)v = 0 \implies Tv = \mu v$.
Since $v \in K_\lambda$:

$$(\lambda I - T)^n v = 0 \implies (\lambda I - T)^{n-1}(\lambda I - T)v = 0 \implies (\lambda I - T)^{n-1}(\lambda - \mu)v = 0$$

Since $\mu \neq \lambda$, $\lambda - \mu \neq 0$, so:
$$(\lambda I - T)^{n-1}v = 0$$

If we do this revursively, we see:
$$(\lambda I - T)v = 0 \implies Tv = \lambda v$$

But we know $Tv = \mu v$, so $\lambda = \mu \notin$ ∎

Lemma 6.1.8:
Suppose $\lambda \neq \mu$, then $K_\lambda \cap K_\mu = \{0\}$
Proof:
Suppose $v \in K_\lambda \cap K_\mu$ and assume for the sake of a contradiction that $v \neq 0$.
By the previous lemma:
$$0 \neq (\mu I - T)v \in K_\lambda$$

Let:
$$v_1 := (\mu I - T)v$$
$$\forall i > 1 : v_i := (\mu I - T)v_{i-1} = (\mu I - T)^i v$$

Inductively, we see $0 \neq v_i \in K_\lambda$, which necessarily means $v_n \neq 0$, but $v_n = (\mu I - T)^n v$, which contradicts that $v \in K_\mu$ ∉, so $v = 0$ ∎

Definition 6.4:

$$I_\lambda := \operatorname{Im}\left((\lambda I - T)^n\right)$$

Lemma 6.1.9:
I_λ is invariant under T
Proof:
Suppose $v \in I_\lambda$, so $\exists u \in V : v = (\lambda I - T)^n u$. So:
$$Tv = T(\lambda I - T)^n u = (\lambda I - T)^n Tu \in I_\lambda \quad \blacksquare$$

Lemma 6.1.10:
$V = K_\lambda \oplus I_\lambda$
Proof:

- $K_\lambda \cap I_\lambda = \{0\}$.
  Suppose $v \in K_\lambda \cap I_\lambda$ This means that:
  $$(\lambda I - T)^n v = 0$$
  $$\exists u \in V : (\lambda I - T)^n u = v$$

  So:
  $$\implies (\lambda I - T)^{2n} u = 0$$

  By a previous lemma this means that
  $$(\lambda I - T)^n u = 0 \implies v = 0$$

  So $K_\lambda \cap I_\lambda = \{0\}$.

- $K_\lambda + I_\lambda = V$:

  We know:
  $$\dim(K_\lambda + I_\lambda) = \underbrace{\dim K_\lambda + \dim I_\lambda}_{\text{By the rank-nullity theorem}} - \underbrace{\dim(K_\lambda \cap I_\lambda)}_{=\dim\{0\}=0} = \dim V$$
  $$= \dim V$$

  So $K_\lambda + I_\lambda = V$

  $$\implies V = K_\lambda \oplus I_\lambda \quad \blacksquare$$

Lemma 6.1.11:

If T is nilpotent, then:

- $p_T(x) = x^n$
- T's only eigenvalue is 0

Proof:

This is trivial in the case that $T = 0$, otherwise:

Since T is nilpotent, there exists an m such that $T^m = 0$ while $T^{m-1} \neq 0$. So $m_T(x) \mid x^m \implies m_T(x) = x^k$ for some $1 \leq k \leq m$. (In fact $k = m$: if $k < m$, that means $T^k \neq 0$ (as $T^{m-1} \neq 0$), so $k = m$)

Since the characteristic polynomial and the minimal polynomial have the same irreducible factorization, and x is irreducible:

$$p_T(x) = x^n$$

(Since the degree of p_T is n)

So the only root of $p_T(x)$ is 0 $\quad \blacksquare$

Lemma 6.1.12:

Let T be a linear transformation, $\lambda \in \mathbb{F}$. Let $T_0 := T \mid_{K_\lambda}$. So $p_{T_0}(x) = (x - \lambda)^{\dim K_\lambda}$

Proof:

Let:

$$L = T - \lambda I : K_\lambda \longrightarrow K_\lambda$$

L is nilpotent since $\forall v \in K_\lambda : L^n v = (T - \lambda I)^n v = (-1)^n (\lambda I - T)^n v = 0$.

So by the previous lemma:

$$p_L(x) = x^{\dim K_\lambda}$$

But at the same time:

$$p_L(x) = \det(xI - L) = \det(xI + \lambda I - T) = \det((x + \lambda)I - T) = p_T(x + \lambda)$$

Which means that:

$$p_T(x + \lambda) = p_L(x) \implies p_T(x) = p_L(x - \lambda) = (x - \lambda)^{\dim K_\lambda} \quad \blacksquare$$

$\dim K_\lambda = \mu_\lambda$

Proof:

Let B' be a basis for K_λ, B'' a basis for I_λ. So if we define:

$$B := B' \sqcup B''$$

B is a basis for V.

Let $A := [T]_B$.

By lemma 6.1.2 A is in the form:

$$A = \left(\begin{array}{c|c} M_K & \\ \hline & M_I \end{array} \right)$$

So $p_T(x) = p_A(x) = p_{M_K}(x) p_{I_K}(x)$

We know that $p_{M_k}(x) = p_{T|_{K_\lambda}}(x) = (x - \lambda)^{\dim K_\lambda}$ by our previous lemma.

This means that the multiplicity of λ in $p_T(x)$ is at least $\dim K_\lambda$, so $\mu_\lambda \geq \dim K_\lambda$.

But $p_{M_I}(\lambda) \neq 0$ since if this were true, there would exist a

$$0 \neq v \in I_\lambda : Tv = \lambda v \implies v \in V_\lambda \implies v \in K_\lambda \implies v \in K_\lambda \cap I_\lambda \implies v = 0 \; \text{⨍}$$

So this means the multiplicity of λ in $p_T(x)$ is equal to the multiplicity of λ in $p_{M_K}(x) = \dim K_\lambda$ $\blacksquare$

Spectral Factorization Theorem (6.1):

$p_T(x)$ is fully factorizable $\iff \displaystyle\bigoplus_{\lambda \in \operatorname{spec}(T)} V_\lambda = V$

Proof:

$p_T(x)$ is fully factorizable iff:

$$\sum_{\lambda \in \operatorname{spec}(T)} \mu_\lambda = \dim V$$

By the previous lemma this is iff:

$$\sum_{\lambda \in \operatorname{spec}(T)} \dim K_\lambda = \dim V$$

Since $\{K_\lambda\}_{\lambda \in \operatorname{spec}(T)}$ are disjoint subspaces:

$$\dim \bigoplus_{\lambda \in \operatorname{spec}(T)} K_\lambda = \sum_{\lambda \in \operatorname{spec}(T)} \dim K_\lambda$$

So $p_T(x)$ is fully factorizable iff:

$$\dim \bigoplus_{\lambda \in \operatorname{spec}(T)} K_\lambda = \dim V$$

$$\iff \bigoplus_{\lambda \in \operatorname{spec}(T)} K_\lambda = V \; \blacksquare$$

2.6 Jordan Normal Form

Definition 7.1:

    A Jordan Block is a matrix denoted as $J_n(\lambda) \in]bF^{n \times n}$ in the form:

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & \cdots & 0 & \lambda \end{pmatrix}$$

Formally:

$$[J_n(\lambda)]_{ij} = \begin{cases} \lambda & i = j \\ 1 & i = j - 1 \\ 0 & \text{else} \end{cases}$$

Statement:

    For any $n > 1$, $J_n(\lambda)$ isn't diagonalizable

Proof:

    Notice that since $xI - J_n(\lambda)$ is an upper triangle matrix:

$$p_{J_n(\lambda)}(x) = \det(xI - J_n(\lambda)) = (x - \lambda)^n$$

So its only eigenvalue is λ. But:

$$N(\lambda I - J_n(\lambda)) = N \begin{pmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \\ 0 & \cdots & \cdots & 0 \end{pmatrix}$$

    Which has a dimension of 1 (since its rank is $n-1$), which means that $\mu_\lambda \neq \gamma_\lambda \implies J_n(\lambda)$ isn't diagonalizable

    ∎

Lemma 6.2.1:

    Let T be a nilpotent linear transformation and let $B \subseteq V$ be a basis. $[T]_B = \begin{pmatrix} J_{m_1}(0) & & \\ & \ddots & \\ & & J_{m_t}(0) \end{pmatrix} \iff$

    there exists disjoint paths $\{E_i\}_{i=1}^t$ such that $B = \bigsqcup_{i=1}^t E_i$

Proof:

    $\implies$: Since $J_{m_i}(0)$ is a block of $[T]_B$, that means there exists a subset (in order) E_i of B such that $J_{m_i}(0) = [T_i]_{E_i}$ where $T_i := T\mid_{E_i}$.

    Suppose $E_i = (v_1 \ldots v_{m_i})$. This means that $Tv_1 = 0$ and $Tv_i = v_{i-1} \implies v_i = T^{n-i}v_n$. Let $v := v_n$

    So that means:

$$E_i = \left(T^{n-1}v \ldots Tv, v\right)$$

    And since $T^n v = Tv_1 = 0$, this is a path (in other words, jordan matrices represent paths).

    These paths are disjoint because their union forms a basis.

    $\impliedby$: Let $V_i := \text{span}(E_i)$. We know that V_i is invariant under T. Let $T_i := T\mid_{V_i}$. By using what we proved previously, $[T_i]_{E_i} = J_{m_i}(0)$ (where $m_i := |E_i|$).

    Since $B = \bigoplus_{i=1}^t E_i \implies [T]_B = \bigoplus_{i=1}^t [T_i]_{E_i} = \bigoplus_{i=1}^t J_{m_i}(0)$ as required.

    ∎

Definition 7.2:

The degree of nilpotency of a nilpotentic linear transformation T is the minimum k such that $T^k = 0$.

Lemma 6.2.2:

If T is a nilpotentic linear transformation of degree k, its maximum path is of length k

Proof:

It's obvious that all paths of T are at most of length k, since $\forall v \in V : T^k v = 0$ so the last (really the first) element in v's path will be $T^i v$ for some $i < k$, so the path is of length k or less.

Assume for the sake of a contradiction that all paths have a length of strictly less than k. Let E_v represent the path starting at v (ie. $E_v = (T^m v \cdots T v, v)$). This means that $\forall v \in V : |E_v| < k \implies \forall v \in V \exists i < k : T^i v = 0$ in contradiction with the fact that k is T's degree of nilpotency.

Lemma 6.2.3:

Suppose T is a nilpotentic linear transformation. Therefore there exists a basis which is a disjoint union of paths of T

Proof:

Let k be the degree of nilpotency of T.

Remember that $\operatorname{Im} T^2 \subseteq \operatorname{Im} T$, and $\operatorname{Im} T^k = \{0\}$. So let's focus on the chains:

$$\{0\} \subset \operatorname{Im} T^{k-1} \subseteq \cdots \subseteq \operatorname{Im} T^2 \subseteq \operatorname{Im} T \subseteq V$$

Notice that if $v \in \operatorname{Im} T^{k-1} \implies \exists u \in V : T^{k-1} u = v \implies T v = T^k u = 0 \implies v \in \ker T$. So $\operatorname{Im} T^{k-1} \subseteq \ker T$. Which means $\operatorname{Im} T^{k-1} \cap \ker T = \operatorname{Im} T^{k-1}$.

So if take the intersection of all these bases with $\ker T$, you get the following chain:

$$\{0\} \subset \operatorname{Im} T^{k-1} \subseteq \cdots \subseteq \ker T \cap \operatorname{Im} T \subseteq \ker T$$

Suppose $T^{k-1} v_1 \ldots T^{k-1} v_{r_1}$ forms a basis of $\operatorname{Im} T^{k-1}$. Let this basis be B_1.

We can extend this basis to a basis of $\ker T \cap \operatorname{Im} T^{k-2}$, etc. recursively. Let:

$$B_i = B_{i-1} \cup \left(T^{k-i} v_{r_{i-1}+1} \ldots T^{k-i} v_{r_i} \right)$$

The extension of B_i to a basis of $\ker T \cap \operatorname{Im} T^{k-i}$.

(Remember that $T^0 = I$, so once we get to B_k, the extension is just vectors in the form v, not Tv)

So B_k is a basis of $\ker T$.

Note that for every element in B_k, it is the end of a path. Suppose $T^{k-i} v_j \in B_k$, since it is also in $\ker T$, $T^{k-i+1} v_j = 0$. So the tuple $\left(T^{k-i} v_j \ldots T v_j, v_j \right)$ is a path. Let this path be E_{v_j}.

Let's define:

$$B := \bigsqcup_{v \in B_k} E_v$$

The union of all theses paths. This is disjoint because the ends of these paths are all different, so the rest of the paths are different.

We will now prove that B is a basis.

- Linear independence: Suppose there exists a zeroing linear combination of B. We can rewrite this zeroing linear combination as a sum of linear combinations of paths. So, suppose v_i is an ordering of elements in B_k (since it's a tuple). It is ordered in the way we defined it, so the vectors of B_1 come first, then $B_2 \setminus B_1$, then $B_3 \setminus B_2$ and so on. (The basis of $\operatorname{Im} T^{k-1}$ comes first, then the remaining vectors in the basis of $\ker T \cap \operatorname{Im} T^{k-2}$ and so on to $\ker T$).

  Let the i-th vector in E_v be $T^{i-1} v$ (technically this reverses the order of E_v, but that's not important right now).

  So that means:

$$\sum_{i=1}^{r_k} \sum_{j=1}^{|E_{v_i}|} \left(\alpha_{ij} T^{j-1} v_i \right) = 0$$

26

(Remember that the size of B_i is r_i, so the size of B_k is r_k)

If we take T^{k-1} of this sum, we get:

$$\sum_{i=1}^{r_k} \sum_{j=1}^{|E_{v_i}|} \left(\alpha_{ij} T^{k-2+j} v_i\right) = 0$$

Notice that given $i > r_1$, the size of E_{v_i} is less than k, so $T^{k-1} E_{v_i} = \{0\}$. And since the degree of nilpotency of T is $k-1$, so forall $j > 1$, $T^{k-2+j} = 0$. So this sum is equal to:

$$\sum_{i=1}^{r_1} \left(\alpha_{i1} T^{k-1} v_i\right) = 0$$

Since $\left(T^{k-1} v_1 \ldots T^{k-1} v_{r_1}\right) = B_1$ which is the basis of $\operatorname{Im} T^{k-1}$, and is therefore linearly independent, so $\forall 1 \le i \le r_1 : \alpha_{i1} = 0$

If we take T^{k-2}, we get:

$$\sum_{i=1}^{r_k} \sum_{j=1}^{|E_{v_i}|} \left(\alpha_{ij} T^{k-3+j} v_i\right) = 0$$

Forall $i > r_2$, $|E_{v_2}| < k-1 \implies T^{k-2} E_{v_2} = \{0\}$. And $\forall j > 2 : T^{k-3+j} = 0$, so:

$$\sum_{i=1}^{r_2} \sum_{j=1}^{2} \left(\alpha_{ij} T^{k-3+j} v_i\right) = 0$$

$$\sum_{i=1}^{r_2} \left(\alpha_{i1} T^{k-2} + \alpha_{i2} T^{k-1} v_i\right) = 0$$

Expanding this to two sums, we see:

$$0 = \sum_{i=1}^{r_2} \left(\alpha_{i1} T^{k-2} v_i\right) + \sum_{i=1}^{r_2} \left(\alpha_{i2} T^{k-1} v_i\right) = \sum_{i=r_1+1}^{r_2} \left(\alpha_{i1} T^{k-2} v_i\right) + \sum_{i=1}^{r_1} \left(\alpha_{i2} T^{k-1} v_i\right)$$

But this is a linear combination of B_2, so $\forall 1 \le i \le r_1 : \alpha_{i2} = 0$ and $\forall 1 \le i \le r_2 : \alpha_{i1} = 0$.

If we continue to do this process inductively, we will keep getting zero spans of the B_i, so in the end, we get that $\forall i, j : \alpha_{ij} = 0$, which means B is linearly independent.

- Spanning: First, we will prove a statement to help us prove this.

  Statement:

  If $T^m v \in T^m(\operatorname{span}(B)) \implies T^{m-1} v \in T^{m-1}(\operatorname{span}(B))$

  Proof:

  This means there exists a $u \in \operatorname{span}(B)$ such that:

  $$T^m v = T^m u \implies T^m v - T^m u = 0 \implies T(T^{m-1} v - T^{m-1} u) = 0 \implies T^{m-1} v - T^{m-1} u \in \ker T$$

  But at the same time:

  $$T^{m-1} v - T^{m-1} u \in \operatorname{Im} T^{m-1} \implies T^{m-1} v - T^{m-1} u \in \ker T \cap \operatorname{Im} T^{m-1}$$

  This means that $T^{m-1} v - T^{m-1} u \in \operatorname{span}(B_{m-1})$, since B_{m-1} is a basis of $\ker T \cap \operatorname{Im} T^{m-1}$. These vectors have the form of $T^{m-1} w$ (some are $T^i w'$ for $i > m-1$, but that is still $T^{m-1} w$ for $w = T^{i+1-m} w'$). So suppose:

  $$T^{m-1} v - T^{m-1} u = \sum_{i=1}^{r_{m-1}} \alpha_i T^{m-1} v_i = T^{m-1} \sum_{i=1}^{r_{m-1}} \alpha_i v_i$$

  Since the paths of v_i are in B, that means that v_i are also in B, so $\sum_{i=1}^{r_{m-1}} \alpha_i v_i \in \operatorname{span}(B) \implies T^{m-1} v - T^{m-1} u \in T^{m-1}(\operatorname{span}(B))$.

So there exists a $w \in \text{span}(B)$ where:

$$T^{m-1}v - T^{m-1}u = T^{m-1}w \implies T^{m-1}v = T^{m-1}(u+w)$$

And since $u, w \in \text{span}(B) \implies u + w \in \text{span}(B)$, so

$$T^{m-1}v \in T^{m-1}(\text{span}(B)) \quad \blacksquare$$

Suppose $v \in V$. Since $B_1 = \left(T^{k-1}v_1 \ldots T^{k-1}v_{r_1}\right) = T^{k-1}\left(v_1 \ldots v_{r_1}\right)$ is a basis for $\text{Im}\, T^{k-1}$ therefore $\left(v_1 \ldots v_{r_1}\right) \subseteq B$, $T^{k-1}v \in T^{k-1}\text{span}\left(v_1 \ldots v_{r_1}\right) \implies T^{k-1}v \in T^{k-1}\text{span}(B)$.

By our previous statement, that means that $T^{k-2}v \in T^{k-2}\text{span}(B)$ and inductively: $\cdots \implies Tv \in T(\text{span}(B)) \implies v \in \text{span}(B)$ $\blacksquare$

Lemma 6.2.4:

Let E be a path of length m and $T_0 := T \mid_{\text{span}(E)}$, then:

$$\dim(\ker T_0 \cap \text{Im}\, T_0^j) = \begin{cases} 0 & j \geq m \\ 1 & j < m \end{cases}$$

Proof:

If $j \geq m$: Since $T_0^m v = 0$ and remember that $E = \left(T_0^{m-1}v \cdots T_0 v, v\right)$, so $T_0^j(\text{span}(E)) = \{0\}$. Since T_0 is defined over $\text{span}(E)$, this means that $\text{Im}\, T_0^j = \{0\} \implies \ker T_0 \cap \text{Im}\, T_0^j = \{0\} \implies \dim(\ker T_0 \cap \text{Im}\, T_0^j) = 0$

Else if $j < m$: Notice that since $T_0^{m-1}v \ldots T_0 v \in \text{Im}\, T_0 \implies \dim \text{Im}\, T_0 \geq m-1$. And by the rank-nullity theorem:

$$\dim \ker T_0 + \dim \text{Im}\, T_0 = \dim \text{span}(E) = m \implies \dim \ker T_0 = m - \dim \text{Im}\, T_0 \leq 1$$

Which means that:

$$\dim(\ker T_0 \cap \text{Im}\, T_0^j) \leq 1$$

But $T_0^{m-1}v \in \ker T_0 \implies T_0^{m-1}v \in \ker T_0 \cap \text{Im}\, T_0^j$ which means that the dimension is non-zero, so:

$$\dim(\ker T_0 \cap \text{Im}\, T_0^j) = 1 \quad \blacksquare$$

Lemma 6.2.5:

Let T be a nilpotent linear transformation of degree k. The number of paths of length greater than j (where $1 \leq j \leq k$) in B, the basis which is a union paths of T, is equal to $\dim(\ker T \cap \text{Im}\, T^j)$

Proof:

Forall paths E_i in B, let $V_i := \text{span}(E_i)$. And let $T_i := T \mid_{V_i}$.

Since:

$$B = B_1 \sqcup \cdots \sqcup E_t$$

That means:

$$V = V_1 \oplus \cdots \oplus V_t$$

Which in turn means:

$$\text{Im}\, T^j = \text{Im}\, T_1^j \oplus \cdots \oplus \text{Im}\, T_t^j$$
$$\ker T = \ker T_1 \oplus \cdots \oplus \ker T_t$$

Which means that:

$$\ker T \cap \text{Im}\, T^j = \bigoplus_{i=1}^{t} \ker T_i \cap \text{Im}\, T_i^j \implies \dim(\ker T \cap \text{Im}\, T^j) = \sum_{i=1}^{t} \dim(\ker T_i \cap \text{Im}\, T_i^j)$$

By our previous lemma $\dim(\ker T_i \cap \text{Im}\, T_i^j) = 1$ if $j < |E_i|$, and 0 otherwise, this must be equal to the number of paths of length greater than j $\blacksquare$

The number of paths of some length in two different bases of paths is the same

Proof:

The number of paths of *exactly* length j is going to be the number of paths longer than j minus the number of paths longer than $j-1$, so the number of paths of exactly length j in B is going to be:

$$\dim(\ker T \cap \operatorname{Im} T^j) - \dim(\ker T \cap \operatorname{Im} T^{j-1})$$

Which doesn't rely on which paths we choose for B ∎

Definition 7.3:

A Jordan Normal Form of a linear transformation is a matrix representation in the form:

$$\begin{pmatrix} \boxed{J_{m_1}(\lambda_1)} & & \\ & \ddots & \\ & & \boxed{J_{m_t}(\lambda_t)} \end{pmatrix}$$

(Similarly the Jordan Normal Form of a matrix is a matrix of the above form that is similar to the given matrix)

The Nilpotent Jordan Normal Form (6.2):

Suppose T is a nilpotent linear transformation. T has a Jordan Normal Form which is unique up to the order of its blocks

Proof:

- Existence: We proved in a previous lemma that a nilpotent linear transformation has a Jordan Normal form iff there exists a basis which is a disjoint union of paths. In lemma 6.2.3 we showed that any nilpotent matrix has a basis of paths. Together this means that every nilpotent linear transformation has a Jordan Normal Form.

- Uniqueness: Suppose $[T]_{B_1}$ and $[T]_{B_2}$ are both Jordan Normal Forms of T. B_1 and B_2 must both be disjoint unions of paths since block jordan matrices represent paths (as proven in a previous lemma).

  The number of block jordan matrices in $[T]_{B_i}$ of some size m is the number of paths of length m in B_i. By our previous lemma, this number is doesn't depend on B_i. So the number of block jordan matrices of length m in $[T]_{B_1}$ is equal to the number in $[T]_{B_2}$ meaning that their only difference is the order of the block jordan matrices in the Jordan Normal Form (as all of the block jordans have 0s on their diagonal, so the only distinguishing trait of each block jordan is their size) ∎

Suppose T is a linear transformation. T has a Jordan Normal Form iff $p_T(x)$ is fully factorizable

Proof:

$\implies$: This is pretty trivial. Suppose J is T's jordan normal form, so $p_T(x) = p_J(x)$. Since J is an upper triangle matrix, its characteristic polynomial is fully factorizable (product of elements on the diagonal).

$\impliedby$: Suppose:

$$p_T(x) = (x - \lambda_1) \cdots (x - \lambda_t)$$

We know by the Spectral Factorization Theorem that there exists invariant subspaces $\{K_{\lambda_i}\}_{i=1}^t$ such that:

$$V = \bigoplus_{i=1}^t K_{\lambda_i}$$

Let $T_i := T \mid_{K_{\lambda_i}}$. We know by lemma 6.1.12 that $p_{T_i}(x) = (x - \lambda_i)^{\mu(\lambda_i)}$.

If we then define:

$$\widetilde{T_i} := T_i - \lambda_i I$$

$\widetilde{T_i}$ is nilpotent: $p_{\widetilde{T}}(x) = \det(xI - [T_i]_B + \lambda_i I) = \det((x + \lambda_i)I - [T_i]_B) = p_T(x + \lambda_i) = x^{\mu(\lambda_i)}$.

So $\widetilde{T_i}$ has a unique (up to order) Jordan Normal Form. Suppose B is a basis such that $[\widetilde{T_i}]_B$ is in jordan normal form, $J_1(0) \oplus \cdots \oplus J_t(0)$. So:

$$[T_i - \lambda_i I]_B = J_1(0) \oplus \cdots \oplus J_t(0) \implies [T_i]_{E_i} - \lambda_i[I]_{E_i} = J_1(0) \oplus \cdots \oplus J_t(0)$$

$$\implies [T_i]_B = J_1(0) \oplus \cdots \oplus J_t(0) + \lambda_i I = J_1(\lambda_i) \oplus \cdots \oplus J_t(\lambda_i)$$

Which is a jordan normal form.

And this is unique up to order (otherwise, $\widetilde{T_i}$'s jordan normal form wouldn't be unique either). ∎

Now might be a good time to stop and ponder why the hell we just went through all the trouble to prove something seemingly useless. I'll give you a spoiler: it's because it's not useless. You can use the Jordan Normal Form to prove if two matrices are similar or not, regardless of whether or not they're diagonalizable. And taking the power of a jordan normal form is much easier than that of another matrix (other than a diagonal matrix).

Theorem 6.4:

The following traits of the Jordan Normal Form are true:

1. The number of block jordan matrices with λ on their diagonal in T's Jordan Normal Form is γ_λ.

2. The sum of the sizes of the λ block jordans in the Jordan Normal Form is μ_λ.

3. The size of the maximum λ block jordan in the Jordan Normal Form is the multiplicity of $x - \lambda$ in $m_T(x)$.

Proof:

Let A_λ be the block of λ block jordan matrices in the Jordan Normal Form, ie:

$$A_\lambda = \bigoplus_{i=1}^t J_{m_i}(\lambda)$$

1. We need to prove here that $t = \gamma_\lambda$. Note that A_λ is $[T_i]_B$ from our previous proof (where i corresponds to λ obviously). Since A_λ is the representation of a reduced T over K_λ (as per our previous proof), $\gamma_{A_\lambda}(\lambda) = \gamma_\lambda$ since $V_\lambda \subseteq K_\lambda$.

   Also notice that $\gamma_{J_{m_i}}(\lambda) = 1$, and $\gamma_{A_\lambda} = \sum_{i=1}^t \gamma_{J_{m_i}}(\lambda)$, which means:

   $$\gamma_\lambda = \sum_{i=1}^t 1 = t$$

2. We know the size of A_λ is $\dim K_\lambda = \mu_\lambda$. And the size of A_λ is the sum of the sizes of all λ jordan blocks (since A_λ is their direct sum).

3. Notice that since A_λ is a direct sum of jordan blocks, which have a minimal polynomial of $(x - \lambda)^{m_i}$, the minimal polynomial of A_λ will be $(x - \lambda)^{m_\lambda}$ where m_λ is the maximum m_i, so the maximum size of the jordan block.

   Since all of the A_λs have different roots to their minimal polynomial, the minimal polynomial of T will be:
   $$m_T(x) = \prod_{\lambda \in \text{spec}(T)} m_{A_\lambda}(x) = \prod_{\lambda \in \text{spec}(T)} (x - \lambda)^{m_\lambda}$$

$\blacksquare$

Theorem 6.5:

T is diagonalizable $\iff$ $m_T(x)$ fully factorizes and the multiplicity of each factor is 1, ie. $m_T(x) = \prod_{\lambda \in \text{spec}(T)} (x - \lambda)$

Proof:

T is diagonalizable iff there exists a basis such that $[T]_B$ is a diagonal matrix, but a diagonal matrix is a Jordan Normal Form (the blocks are of size 1), so this is iff the maximum multiplicity of λ for each $\lambda \in \text{spec}(T)$ in $m_T(x)$ is at most 1, while at the same time it must be greater or equal to 1, so this is iff $\forall \lambda \in \text{spec}(T)$ the multiplicity of λ in $m_T(x)$ is 1 $\blacksquare$

3 Inner Product Spaces

3.1 The Inner Product

In this section, we must reduce our generalizations. So we will only be discussing the real and complex fields. So $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$ in this section.

Definition 8.1:

If V is a vector space over $\mathbb{F}$ equipped with a function:

$$\langle\ ,\ \rangle : V \times V \longrightarrow \mathbb{F}$$

That satisfies the following properties:
$\alpha, \beta \in \mathbb{F}$ and $v, u, w \in V$

1. $\langle \alpha v + \beta u, w \rangle = \alpha \langle v, w \rangle + \beta \langle u, w \rangle$
2. $\langle v, \alpha u + \beta w \rangle = \overline{\alpha} \langle v, u \rangle + \overline{\beta} \langle v, w \rangle$ where $\overline{\alpha}$ is the complex conjugate
3. $\langle v, u \rangle = \overline{\langle u, v \rangle}$
4. $0 \leq \langle v, v \rangle \in \mathbb{R}$
5. $\langle v, v \rangle = 0 \iff v = 0$

$\langle\ ,\ \rangle$ is called the inner product and V is called the inner product space.

The first and second criteria mean the inner product is *sesquilinear* ($1\frac{1}{2}$ linearity), the third criteria means it is *hermitian*, and the fourth and fifth means its *positive* (technically non-negative),

Statement:

The extended dot product over $\mathbb{F}^n$ defined like so: $\langle v, u \rangle = v^T \overline{u}$ is an inner product

Proof:

Suppose

$$v = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$$

$$u = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix}$$

$$w = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \end{pmatrix}$$

$$\implies \langle v, u \rangle = \sum_{i=1}^{n} \alpha_i \overline{\beta_i}$$

1.
$$\langle \delta v + \varepsilon u, w \rangle = \sum_{i=1}^{n} (\delta \alpha_i + \varepsilon \beta_i) \overline{\gamma_i} = \delta \sum_{i=1}^{n} \alpha_i \overline{\gamma_i} + \varepsilon \sum_{i=1}^{n} \beta_i \overline{\gamma_i} = \delta \langle v, w \rangle + \varepsilon \langle u, w \rangle$$

Which satisfies the first criteria.

2.
$$\langle v, \delta u + \varepsilon w \rangle = \sum_{i=1}^{n} \alpha_i \left(\overline{\delta \beta_i + \varepsilon \gamma_i} \right) = \sum_{i=1}^{n} \alpha_i \left(\overline{\delta \beta_i} \right) + \sum_{i=1}^{n} \alpha_i \left(\overline{\varepsilon \gamma_i} \right) = \overline{\delta} \sum_{i=1}^{n} \alpha_i \overline{\beta_i} + \overline{\varepsilon} \sum_{i=1}^{n} \alpha_i \overline{\gamma_i} =$$
$$= \overline{\delta} \langle v, u \rangle + \overline{\varepsilon} \langle v, w \rangle$$

Which satisfies the second criteria.

3.

$$\overline{\langle u, v \rangle} = \overline{\sum_{i=1}^{n} \beta_i \overline{\alpha_i}} = \sum_{i=1}^{n} \overline{\beta_i \overline{\alpha_i}} = \sum_{i=1}^{n} \overline{\beta_i} \overline{\overline{\alpha_i}} = \sum_{i=1}^{n} \alpha_i \overline{\beta_i} = \langle v, u \rangle$$

Which satisfies the third criteria

4.

$$\langle v, v \rangle = \sum_{i=1}^{n} \alpha_i \overline{\alpha_i} = \sum_{i=1}^{n} |\alpha_i|$$

Since the magnitude of α_i non-negative, ie. $|\alpha_i| \geq 0$, that means $\sum_{i=1}^{n} |\alpha_i| \geq 0$. Which satisfies the fourth criteria

5. Since $\langle v, v \rangle$ is a sum of non-negative values, it equals 0 if and only if all of those values are 0, so iff $|\alpha_i| = 0 \iff \alpha_i = 0 \iff v = 0$, which satisfies the fifth criteria

So all of the criteria are satisfied, which means the standard dot product is an inner product.

Note:

Notice that if $\mathbb{F} = \mathbb{R}$ then $\overline{v} = v$, so the standard dot product over $\mathbb{R}$ is the dot product you'd learn in highschool, ie:

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} \cdot \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_n \end{pmatrix} = \sum_{i=1}^{n} \alpha_i \beta_i$$

Is an inner product over $\mathbb{R}^n$

Definition 8.2:

We define the norm of a vector to be a function:

$$|| \ || : V \longrightarrow \mathbb{F}$$

That satisfies the following the following criteria:

1. Positiveness: $0 \leq ||v|| \in \mathbb{R}$ and $||v|| = 0 \iff v = 0$
2. Homogeneity: $||\alpha v|| = |\alpha| \cdot ||v||$
3. The Triangular Inequality: $||v + u|| \leq ||v|| + ||u||$

The purposes of the norm is to represent the "length" or "magnitude" of a vector.
A vector space equipped with a norm is called a normed vector space.

Note:

We will only be focusing on one kind of norm right now, the inner product norm, where:

$$||v|| := \sqrt{\langle v, v \rangle}$$

This comes from the geometric definition of the inner/dot product.
The proof that this is a norm (ie. it satisfies all criteria) is not trivial, so our next step is to prove that it is indeed a norm.

Definition 8.3:

Two vectors v and u are called orthogonal if: $\langle v, u \rangle = 0$. This is denoted as $v \perp u$ (notice that since the inner product is hermitian, $\langle v, u \rangle = 0 \iff \langle u, v \rangle = 0$, so orthogonality is symmetric)

Lemma 7.1.1:

The following are characteristics of orthogonality:

- $\langle v, 0 \rangle = 0$
- $v \perp u \iff u \perp v$
- $v \perp u \implies \alpha v \perp \beta u$

Proof:

-
$$\langle v, 0 \rangle = \langle v, 0 \cdot v \rangle = \overline{0} \langle v, v \rangle = 0 \langle v, v \rangle = 0$$

- Suppose $v \perp u \implies \langle v, u \rangle = 0$. So $\langle u, v \rangle = \overline{\langle v, u \rangle} = \overline{0} = 0$ which means that $u \perp v$ (this also proves the other direction. This is because you can switch the names of u and v)

-
$$\langle \alpha v, \beta u \rangle = \alpha \cdot \overline{\beta} \langle v, u \rangle = \alpha \cdot \overline{\beta} \cdot 0 = 0 \implies \alpha v \perp \beta u$$

■

Lemma 7.1.2:

$v \perp u \implies ||v + u||^2 = ||v||^2 + ||u||^2$

Proof:

Since $v \perp u \implies \langle v, u \rangle = \langle u, v \rangle = 0$

$$||v + u||^2 = \langle v + u, v + u \rangle = \langle v, v \rangle + \langle v, u \rangle + \langle u, v \rangle + \langle u, u \rangle = \langle v, v \rangle + \langle u, u \rangle = ||v||^2 + ||u||^2 \quad ■$$

Lemma 7.1.3:

$0 \geq ||v|| \in \mathbb{R}$ and $||\alpha v|| = |\alpha| \, ||v||$

Proof:

1.
$$||v|| = \sqrt{\langle v, v \rangle}$$

which by the criteria for the inner product is a non-negative real number

2.
$$||\alpha v|| = \sqrt{\langle \alpha v, \alpha, v \rangle} = \sqrt{\alpha \overline{\alpha} \langle v, v \rangle} = \sqrt{|\alpha|^2 \langle v, v \rangle} = |\alpha| \sqrt{\langle v, v \rangle} = |\alpha| \, ||v||$$

■

Cauchy-Bunyakovsky-Schwarz Inequality(7.1.4):

1. $|\langle v, u \rangle| \leq ||v|| \cdot ||u||$
2. $|\langle v, u \rangle| = ||v|| \cdot ||u|| \iff \{v, u\}$ is linearly dependent

Proof:

1. **If** $u = 0$:
$$\langle v, u \rangle = |\langle v, 0 \rangle| = 0$$
$$||v|| \cdot ||u|| = ||v|| \cdot 0 = 0$$

So $|\langle v, u \rangle| = ||v|| \cdot ||u||$ and $\{v, u\} = \{v, 0\}$ which is linearly dependent, so the lemma holds.

Else $u \neq 0$ Let:
$$z := v - \frac{\langle v, u \rangle}{\langle u, u \rangle} u$$

Notice:
$$\langle z, u \rangle = \left\langle v - \frac{\langle v, u \rangle}{\langle u, u \rangle} u, u \right\rangle = \langle v, u \rangle - \frac{\langle v, u \rangle}{\langle u, u \rangle} \langle u, u \rangle = \langle v, u \rangle - \langle v, u \rangle = 0$$

So $z \perp u$, which means that $z \perp \dfrac{\langle v, u \rangle}{\langle u, u \rangle} u$ so by lemma 7.1.2:

$$||v||^2 = \left|\left| z + \frac{\langle v, u \rangle}{\langle u, u \rangle} u \right|\right|^2 = ||z||^2 + \left|\left| \frac{\langle v, u \rangle}{\langle u, u \rangle} u \right|\right|^2 = ||z||^2 + \left(\frac{\langle v, u \rangle}{\langle u, u \rangle} \right)^2 ||u||^2 =$$

$$= ||z||^2 + \left(\frac{\langle v, u \rangle}{||u||} \right)^2$$

$$\implies \langle v, u \rangle^2 = ||v||^2 ||u||^2 - ||z||^2 ||u||^2$$

Since $-||z||^2 ||u||^2 \leq 0$ which means that:

$$\langle v, u \rangle^2 \leq ||v||^2 ||u||^2 \implies |\langle v, u \rangle| \leq ||v|| \cdot ||u||$$

Which satisfies the first part of the lemma

2. We saw that if $u = 0$, the lemma is satsified. So we now must prove the second part of the lemma for $u \neq 0$. So we need to prove that $|\langle v, u \rangle| = ||v|| \cdot ||u|| \iff \{v, u\}$ is linearly dependent:

$\implies$:

$$|\langle v, u \rangle| = ||v|| \cdot ||u|| \implies \langle v, u \rangle^2 = ||v||^2 ||u||^2 \implies ||z||^2 ||u||^2 = 0$$

Since $u \neq 0 \implies ||u|| \neq 0$ so $||z|| = 0 \implies z = 0$. So:

$$v - \frac{\langle v, u \rangle}{\langle u, u \rangle} u = 0 \implies v = \frac{\langle v, u \rangle}{\langle u, u \rangle} u$$

So $\{v, u\}$ is linearly dependent

$\impliedby$: Since $\{v, u\}$ is linearly dependent, $v = \alpha u$. So:

$$|\langle v, u \rangle| = |\langle \alpha u, u \rangle| = |\alpha| \, |\langle u, u \rangle| = |\alpha| \, ||u||^2$$

And $||v|| \cdot ||u|| = ||\alpha u|| \cdot ||u|| = |\alpha| \, ||u|| \cdot ||u|| = |\alpha| \, ||u||^2 = |\langle v, u \rangle|$ which satisfies the lemma.

∎

Theorem 7.1:

The inner product norm is a norm function

Proof:

We saw in lemma 7.1.3 that the inner product norm satisfies the first two criteria.
Let's now check the third criterion:

$$||v + u|| = \sqrt{\langle v + u, v + u \rangle} = \sqrt{\langle v, v \rangle + \langle v, u \rangle + \langle u, v \rangle + \langle u, u \rangle} = \sqrt{||v||^2 + \langle v, u \rangle + \langle u, v \rangle + ||u||^2}$$

Since $\langle u, v \rangle = \overline{\langle v, u \rangle} \implies \langle v, u \rangle + \langle u, v \rangle = 2\,\mathrm{Re}(\langle v, u \rangle)$
We know that $\mathrm{Re}(z) \leq |z|$ so $2\,\mathrm{Re}(\langle v, u \rangle) \leq 2\,|\langle v, u \rangle|$, and by the Cauchy-Bunyakovsky-Schwarz Inequality, we know that $2\,|\langle v, u \rangle| \leq 2\,||v|| \cdot ||u||$, so $\langle v, u \rangle + \langle u, v \rangle \leq 2\,||v|| \cdot ||u||$, which means:

$$||v + u|| \leq \sqrt{||v||^2 + 2\,||v|| \, ||u|| + ||u||^2} = ||v|| + ||u|| \quad \blacksquare$$

3.2   Orthogonality

Definition 9.1:

Given a set $B = (v_1 \ldots v_n)$, we define the Gram Matrix, denoted G_B to be a matrix where:

$$[G_B]_{ij} = \langle v_i, v_j \rangle$$

So:

$$G_B = \begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle & \cdots & \langle v_1, v_n \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle & \cdots & \langle v_2, v_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle v_n, v_1 \rangle & \langle v_n, v_2 \rangle & \cdots & \langle v_n, v_n \rangle \end{pmatrix}$$

Statement:

If B is a basis, $\langle u, w \rangle = [u]_B^T G_B \overline{[w]_B}$

Proof:

Suppose $B = (v_1 \ldots v_n)$, and $u, w \in V$ such that $u = \alpha_1 v_1 + \cdots + \alpha_n v_n$ and $w = \beta_1 v_1 + \cdots + \beta_n v_n$. So:

$$\langle u, w \rangle = \left\langle \sum_{i=1}^{n} \alpha_i v_i, \sum_{j=1}^{n} \beta_j v_j \right\rangle = \sum_{i=1}^{n} \alpha_i \cdot \sum_{j=1}^{n} \left(\overline{\beta_j} \cdot \langle v_i, v_j \rangle \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \overline{\beta_j} \langle v_i, v_j \rangle$$

Since $[u]_B^T \in \mathbb{F}^{1 \times n}, G_B \in \mathbb{F}^{n \times n}, \overline{[w]_B} \in \mathbb{F}^{n \times 1}$, this means that $[u]_B^T G_B \overline{[w]_B} \in \mathbb{F}^{1 \times 1}$, so it's a scalar. Which means we can simplify calculations by calculating directly:

$$[u]_B^T G_B \overline{[w]_B} = \sum_{i=1}^{n} \alpha_i \cdot [G_B \overline{[w]_B}]_{i1} = \sum_{i=1}^{n} \alpha_i \cdot \sum_{j=1}^{n} [G_B]_{ij} \overline{\beta_j} = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \overline{\beta_j} \langle v_i, v_j \rangle$$

So $\langle u, w \rangle = [u]_B^T G_B \overline{[w]_B}$ ∎ (While technically I'm equating a matrix and a scalar, the matrix is a 1×1 matrix, so the equation is fine)

Theorem 8.1:

B is linearly independent $\iff G_B$ is invertible

Proof:

$\implies$: We will prove that G_B's columns are linearly independent. Suppose there exist $\alpha_1 \ldots \alpha_n$ such that:

$$\sum_{i=1}^{n} \alpha_i \cdot C_i(G_B) = 0$$

This means that (since $\sum_{i=1}^{n} \alpha_i \cdot C_i(G_B)$ is a column vector. Remember that $[v]_t$ is the t-th element in the column/row vector v):

$$\forall 1 \leq j \leq n : [\sum_{i=1}^{n} \alpha_i \cdot C_i(G_B)]_j = 0 \implies \sum_{i=1}^{n} \alpha_i \langle v_j, v_i \rangle = 0 \implies \left\langle v_j, \sum_{i=1}^{n} \overline{\alpha_i} v_i \right\rangle = 0$$

Let $u := \sum_{i=1}^{n} \overline{\alpha_i} v_i$. So $\forall 1 \leq n : \langle u, v_i \rangle$. Notice that:

$$\langle u, u \rangle = \left\langle u, \sum_{i=1}^{n} \overline{\alpha_i} v_i \right\rangle = \sum_{i=1}^{n} \alpha_i \langle u, v_i \rangle = \sum_{i=1}^{n} 0 = 0$$

Which means that $u = 0$.

So $\alpha_1 v_1 + \ldots \alpha_n v_n = 0$, since $v_1 \ldots v_n$ are linearly independent, this means that $\alpha_1 = \cdots = \alpha_n = 0$.

So the columns of G_B are linearly independent, which means that G_B is invertible.

$\Longleftarrow$: Suppose that there exists α_is such that:
$$\sum_{i=1}^{n} \alpha_i v_i = 0$$

Now notice that given some j:
$$[\sum_{i=1}^{n} \alpha_i C_i(G_B)]_j = \sum_{i=1}^{n} \alpha_i \langle v_j, v_i \rangle = \left\langle v_j, \sum_{i=1}^{n} \alpha_i v_i \right\rangle = \langle v_j, 0 \rangle = 0$$

So that means that:
$$\sum_{i=1}^{n} \alpha_i C_i(G_B) = 0$$

But since G_B is invertible and thus its columns are linearly independent, this means that $\alpha_i = 0$.
So B is linearly independent.

∎

Definition 9.2:

- $S \subset V$ is an orthogonal set if $\forall v \neq u \in S : v \perp u$
- $S \subset V$ is an orthonormal set if S is orthogonal and $\forall v \in S : ||v|| = 1$
- $S \subset V$ is a non-trivial orthogonal set if S is orthogonal and $0 \notin S$
- $S \subset V$ is an orthogonal basis if S is orthogonal and a basis of V
- $S \subset V$ is an orthonormal basis if S is orthonormal and a basis of V

(Notice that $||0|| = 0 \neq 1$, so any orthonormal set is a non-trivial orthogonal set.)

Corollary 1:

Any non-trivial orthogonal set is linearly independent

Proof:

Suppose $S = (v_1 \ldots v_n)$ is a non-trivial orthogonal set (ignore the fact that its quite obviously a tuple: the difference between sets and tuples is almost non-existent sometimes in linear algebra.)
So $[G_S]_{ii} = \langle v_i, v_i \rangle$, since $v_i \neq 0$ this means $\langle v_i, v_i \rangle \neq 0$, so the diagonal of G_S is non-zero.
And for $i \neq j$, notice that $[G_S]_{ij} = \langle v_i, v_j \rangle = 0$ since v_i, v_j are orthogonal. So G_S is a diagonal matrix whose diagonal is non-zero, so G_S is invertible, which means S is linearly independent.

Lemma 8.2.1:

Suppose $E = (e_1 \ldots e_n)$ is an orthonormal basis. Let $v \in V$, suppose $v = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$, then:

1. $\alpha_i = \langle v, e_i \rangle$
2. $||v||^2 = \sum_{i=1}^{n} |\alpha_i|^2$ (Extended Pythagorean Theorem)

Proof:

1.
$$v = \sum_{i=1}^{n} \alpha_i e_i$$

So:
$$\langle v, e_i \rangle = \left\langle \sum_{j=1}^{n} \alpha_j e_j, e_i \right\rangle = \sum_{j=1}^{n} \alpha_j \langle e_j, e_i \rangle$$

Since E is orthogonal, $\langle e_j, e_i \rangle = 0$ when $j \neq i$ and 1 when $j = i$, so
$$= \alpha_i \langle e_i, e_i \rangle \alpha_i$$

2.

$$||v||^2 = \langle v, v \rangle = \left\langle \sum_{i=1}^{n} \alpha_i e_i, \sum_{j=1}^{n} e_j \right\rangle = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \overline{\alpha_j} \langle e_i, e_j \rangle$$

For the same reasoning, this is equal to:

$$\sum_{i=1}^{n} \alpha_i \overline{\alpha_i} = \sum_{i=1}^{n} |\alpha_i|^2$$

∎

Definition 9.3:

Suppose $0 \neq v \in V$, the normed form of the vector v (this is a bit of a mouthfull, so I'll just call it v-normed) is defined to be:

$$\hat{v} := \frac{1}{||v||} v$$

(Obviously this has a norm of 1)

Gram-Schmidt Theorem (8.2):

Every inner product space V has an orthonormal basis

Proof:

We only need to find an orthogonal basis, as we can then norm the basis and that will be orthonormal.

The idea is to do this recursively, given a set of orthogonal vectors, we find a vector that's orthogonal to all of them. The way to do this allows us to analyze some more characteristics of the inner product in the future. Remember that geometrically, to find the normal to a plane (equivalent here to finding an orthogonal vector to a subspace) we could take some vector not on the plane, and project it onto the plane. Say v is the vector and $\tilde{v}$ is its projection. Since $v = \tilde{v} + n$, where n is a normal vector to the plane, $v - \tilde{v}$ gives us our desired normal vector. (It's not important we prove this, because we're proving this for a more general situation. This geometric example is something that should've been proven in high school, but it really doesn't matter.)

So the idea is that we work backwards: we find $\tilde{v}$ and from that we find n. Since we know that $\tilde{v}$ is in the plane (which is the span of all prior normal vectors found as this is a recursive algorithm), it is a linear combination of some known vectors. So we just need to find the proper linear combination and we have our desired vector. Afterwards, we just need to verify that the vectors form a basis, and then we have an orthogonal basis.

If you don't understand this explanation, that's fine. You can understand the proof without understanding the intuition behind it.

Suppose $B = (v_1 \dots v_n)$ is some basis of V, our goal is to construct from this an orthogonal basis.

Let $E = (w_1 \dots w_n)$ be constructed as follows:

$$w_1 := v_1$$

$$w_n := v_n - \sum_{i=1}^{n-1} \alpha_{in} w_i$$

So every vector is v_n minus some linear combination of all prior vectors in E. This is where the intuition above comes in. We want to find the linear combination of $w_1 \dots w_{n-1}$ that gives the projection of w_n onto span $w_1 \dots w_{n-1}$, and from that we find the normal (orthogonal vector) to the subspace. So we need to find the αs that make E an orthogonal basis still, though. Intuition only goes so far.

Let's begin with finding the αs that make this orthogonal.

We want $\forall i < n : \langle w_n, w_i \rangle = 0$. If we assume that $\forall m < n$ this holds (ie. $\forall m < n \forall i < m : w_m \perp w_i$), which we can do as we are building this recursively we see:

$$\langle w_n, w_i \rangle = \left\langle v_n - \sum_{t=1}^{n-1} \alpha_{tn} w_t, w_i \right\rangle = \langle v_n, w_i \rangle - \sum_{t=1}^{n-1} \alpha_{tn} \langle w_t, w_i \rangle$$

By our assumption, $\forall t \neq i : \langle w_t, w_i \rangle = 0$, so this equals:

$$= \langle v_n, w_i \rangle - \alpha_{in} \langle w_i, w_i \rangle$$

Since we want this to be 0, this is iff:

$$\langle v_n, w_i \rangle = \alpha_{in} \langle w_i, w_i \rangle \iff \alpha_{in} = \frac{\langle v_n, w_i \rangle}{\langle w_i, w_i \rangle}$$

So $\forall i < n$ we have found α_{in}, and this is all the αs we need to find (as when $i \geq n$, α_{in} isn't used anywhere). So we have found an orthogonal set with n elements. We just need to finish and prove that it's a basis. We will prove that it spans V, and since there are n elements, it must be a basis.

Statement:

$\forall 1 \leq k \leq n : \text{span} \{v_1 \ldots v_k\} = \text{span} \{w_1 \ldots w_k\}$

Proof:

We will prove this by induction on k.
Base case: $k = 1$, in this case $w_1 = v_1$, so their spans are obviously the same
Inductive step: Assume $P(k-1)$. We know:

$$w_k = v_k - \gamma_1 w_1 - \cdots - \gamma_{k-1} w_{k-1}$$

For some γs, which means that $v_k \in \text{span} \{w_1 \ldots w_k\}$. On the flip side, $w_k \in \text{span} \{v_k, w_1 \ldots w_{k-1}\} = \text{span} \{v_1 \ldots v_k\}$
Suppose $v \in \text{span} \{v_1 \ldots v_k\}$ this means:

$$v = \alpha_1 v_1 + \cdots + \alpha_k v_k$$

Since $\alpha_1 v_1 + \cdots + \alpha_{k-1} v_{k-1} \in \text{span} \{v_1 \ldots v_{k-1}\} = \text{span} \{w_1 \ldots w_{k-1}\}$, and $v_k \in \text{span} \{w_1 \ldots w_k\}$, so $v \in \text{span} \{w_1 \ldots w_k\}$
Suppose $v \in \text{span} \{w_1 \ldots w_k\}$, since $\text{span} \{w_1 \ldots w_{k-1}\} = \text{span} \{v_1 \ldots v_{k-1}\}$ and $w_k \in \text{span} \{v_1 \ldots v_k\}$, that means $v \in \text{span} \{v_1 \ldots v_k\}$.
So $\text{span} \{v_1 \ldots v_k\} = \text{span} \, w_1 \ldots w_k$ ∎

This means that when $k = n$, $\text{span}(E) = \text{span}(B) = V$, so E spans and therefore is an orthogonal basis. So its norm is an orthonormal basis.

Definition 9.4:

Given a set $S \subseteq V$ we define S's orthogonal complement to be:

$$W^\perp = \{v \in V \mid \forall w \in S : v \perp w\}$$

Lemma 8.3.1:

$S^\perp$ is a subspace

Proof:

- $0 \in S^\perp$: since $\forall v \in S : \langle 0, v \rangle = 0 \implies 0 \perp v \implies 0 \in S^\perp$

- Suppose $v, u \in S$, $\alpha \in \mathbb{F}$, $v + \alpha u \in S^\perp$: Let $w \in S$, we need to show $v + \alpha u \perp w$.

$$\langle v + \alpha u, w \rangle = \langle v, w \rangle + \alpha \langle u, w \rangle = 0$$

So $v + \alpha u \in S^\perp$

∎

Lemma 8.3.2:

$S^\perp = \text{span}(S)^\perp$

Proof:

Suppose $S = \{v_1 \ldots v_k\}$

$\subseteq$: Let $v \in S^\perp$, we need to prove that $v \in \text{span}(S)^\perp$, so that $\forall u \in \text{span}(S) : v \perp u$. We know that $\forall i : \langle v_i, v \rangle = 0$.

Let $u \in \text{span}(S)$. Suppose $u = \sum_{i=1}^{k} \alpha_i v_i$, so:

$$\langle u, v \rangle = \left\langle \sum_{i=1}^{k} \alpha_i v_i, v \right\rangle = \sum_{i=1}^{k} \langle v_i, v \rangle = 0$$

So $\forall u \in \text{span}(S) : v \perp u$, so $v \in \text{span}(S)^\perp$

$\supseteq$: Let $u \in \text{span}(S)^\perp$. Since $\forall v \in S$, $v \in \text{span}(S) \implies v \perp u$, so $u \in S^\perp$

$$\implies S^\perp = \text{span}(S)^\perp \quad \blacksquare$$

Lemma 8.3.3:

Given a non-trivial orthogonal set B_0 it can be extended to an orthogonal basis B

Proof:

Suppose $B_0 = (v_1 \ldots v_k)$. We can extend this to a basis $B' = (v_1 \ldots v_k, v_{k+1} \ldots v_n)$. If we execute the process in the Gram-Schmidt Theorem on $(v_{k+1} \ldots v_n)$ we get a non-trivial orthogonal set that spans the same subspace (ie. it is also a basis), let that be B_0', so if we let $B = B_0 \cup B_0'$, B is a basis (as the process doesn't change the span) $\blacksquare$

Theorem 8.3:

Given any subspace W, $V = W \oplus W^\perp$

Proof:

- $W \cap W^\perp = \{0\}$: Suppose $v \in W \cap W^\perp$. Since $v \in W^\perp$, it is perpendicular to all vectors in W, including itself. So $\langle v, v \rangle = 0 \implies v = 0$

- Suppose the orthogonal basis of W is $B_0 = \{v_1 \ldots v_k\}$, and its extension to an orthogonal basis of V is $B = \{v_1 \ldots v_n\}$.

  Let's look at $B_1 = \{v_{k+1} \ldots v_n\}$. Since $\forall v \in B_1 \forall u \in B_0 : v \perp u$ since $B = B_0 \cup B_1$ is an orthogonal basis. This means that $\text{span}(B_1) \perp \text{span}(B_0) = W$. This should already be obvious, but I will prove this formally:

  Let $v \in \text{span}(B_1)$, suppose $v = \sum_{i=k+1}^{n} \alpha_i v_i$. Let $u \in \text{span}(B_0)$, suppose $u = \sum_{i=1}^{k} \alpha_i v_i$. So:

$$\langle v, u \rangle = \left\langle \sum_{i=k+1}^{n} \alpha_i v_i, \sum_{i=1}^{k} \alpha_i v_i \right\rangle = \sum_{i=k+1}^{n} \sum_{j=1}^{k} \alpha_i \overline{\alpha_j} \langle v_i, v_j \rangle$$

  Since $i \neq j$ and B is orthogonal, $\langle v_i, v_j \rangle = 0$, so:

$$\langle v, u \rangle = 0$$

  So $\text{span}(B_1) \subseteq W^\perp$. This means that $\dim \text{span}(B_1) \leq W^\perp$, so:

$$\underbrace{\dim W}_{=\text{span}(B_0)} + \dim \text{span}(B_1) \leq \dim W + \dim W^\perp \implies \dim V \leq \dim W + \dim W^\perp$$

  Since $\dim(W + W^\perp) = \dim W + \dim W^\perp - \dim(W \cap W^\perp) = \dim W + \dim W^\perp$, this means that:

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

So $V = W \oplus W^\perp$ $\blacksquare$

3.3 Projection

Remember in our proof for the Gram-Schmidt Theorem we discussed (during my attempt to explain the intuition behind the proof) how w_n is the normal from span $\{w_1 \ldots w_{n-1}\}$ to v_n, and $\sum_{i=1}^{n-1} \frac{\langle v_n, w_i \rangle}{\langle w_i, w_i \rangle} w_i$ is the projection of v_n onto span $\{w_1 \ldots w_{n-1}\}$. We're now going to expand on this idea.

Definition 10.1:

Given a vector v and a subspace W with an orthogonal basis $E = \{w_1 \ldots w_k\}$, we define the projection of v onto W as:

$$\underset{W}{\pi}(v) := \sum_{i=1}^{k} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i$$

Note:

While I called this "the" projection onto W, the definition is very much relative to the choice of basis. Despite this, we will soon prove that it doesn't matter which basis we choose.

Lemma 9.1.1:

The following are characteristics of projection:

1.
$$\underset{W}{\pi}(v) \in W$$

2.
$$\underset{W}{\pi}(v) = v \iff v \in W$$

3.
$$\underset{W}{\pi} : V \longrightarrow W$$

Is a linear transformation

4.
$$\underset{W}{\pi}(v) \mid_W = I_W$$

5.
$$\forall w \in W : v - \underset{W}{\pi}(v) \perp w$$

6.
$$v \perp W \implies \underset{W}{\pi}(v) = 0$$

Proof:

Suppose $E = \{w_1 \ldots w_k\}$ is an orthogonal basis of W

1. By definition:
$$\underset{W}{\pi}(v) \in \text{span}\,\{w_1 \ldots w_n\} = W$$

2. We will prove both directions:

$\implies$: Suppose $\underset{W}{\pi}(v) = v$, since $\underset{W}{\pi}(v) \in W$ this means that $v \in W$.

$\impliedby$: Suppose $v = \sum_{i=1}^{k} \alpha_i w_i$.

$$\underset{W}{\pi}(v) = \sum_{i=1}^{k} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i = \sum_{i=1}^{k} \frac{\left\langle \sum_{j=1}^{k} \alpha_j w_j, w_i \right\rangle}{\langle w_i, w_i \rangle} w_i = \sum_{i=1}^{k} \frac{\sum_{j=1}^{k} \alpha_j \langle w_j, w_i \rangle}{\langle w_i, w_i \rangle} w_i =$$

$$= \sum_{i=1}^{k} \frac{\alpha_i \langle w_i, w_i \rangle}{\langle w_i, w_i \rangle} w_i = \sum_{i=1}^{k} \alpha_i w_i = v$$

3. Suppose $v, u \in V$ and $\alpha \in \mathbb{F}$, so:

$$\underset{W}{\pi}(v + \alpha u) = \sum_{i=1}^{k} \frac{\langle v + \alpha u, w_i \rangle}{\langle w_i, w_i \rangle} w_i = \sum_{i=1}^{k} \frac{\langle v, w_i \rangle + \alpha \langle u, w_i \rangle}{\langle w_i, w_i \rangle} w_i =$$

$$\sum_{i=1}^{k} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i + \alpha \sum_{i=1}^{k} \frac{\langle u, w_i \rangle}{\langle w_i, w_i \rangle} w_i = \underset{W}{\pi}(v) + \alpha \underset{W}{\pi}(u)$$

Which means $\underset{W}{\pi}$ is a linear transformation

4. We've already seen that $\forall w \in W : \underset{W}{\pi}(w) = w$, so:

$$\underset{W}{\pi}(W) = W \subseteq W$$

So W is $\underset{W}{\pi}$ invariant. And $\underset{W}{\pi}$ doesn't change vectors in W, so its reduced function is just the identity on W

5. Suppose $w = \sum_{i=1}^{k} \alpha_i w_i$.

$$\left\langle v - \underset{W}{\pi}(v), w \right\rangle = \langle v, w \rangle - \left\langle \sum_{i=1}^{k} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i, \sum_{j=1}^{k} \alpha_j w_j \right\rangle = \langle v, w \rangle - \sum_{i=1}^{k} \sum_{j=1}^{k} \overline{\alpha_j} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} \langle w_i, w_j \rangle$$

Since E is orthogonal, $\langle w_i, w_j \rangle = 0$ when $i \neq j$, so:

$$= \langle v, w \rangle - \sum_{i=1}^{k} \overline{\alpha_i} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} \langle w_i, w_i \rangle = \langle v, w \rangle - \sum_{i=1}^{k} \overline{\alpha_i} \langle v, w_i \rangle = \langle v, w \rangle - \left\langle v, \sum_{i=1}^{k} \alpha_j w_j \right\rangle = \langle v, w \rangle - \langle v, w \rangle = 0$$

Which means that $v - \underset{W}{\pi}(v) \perp w$

6. Since $v \perp W \implies \forall i : v \perp w_i$. So:

$$\underset{W}{\pi}(v) = \sum_{i=1}^{k} \frac{\langle v, w_i \rangle}{\langle w_i, w_i \rangle} w_i = \sum_{i=1}^{k} \frac{0}{\langle w_i, w_i \rangle} w_i = 0$$

∎

Theorem 9.1:

$\underset{W}{\pi}$ is independent of the choice for the orthogonal basis of W

Proof:

Let B, B' be two bases for W, and $v \in V$. I will denote $\underset{B}{\pi}$ as the projection relative to B and $\underset{B'}{\pi}$ as the projection relative to B'.

By the prior lemma:

$$v - \underset{B}{\pi}(v), v - \underset{B'}{\pi}(v) \perp W \implies v - \underset{B}{\pi}(v), v - \underset{B'}{\pi}(v) \in W^{\perp}$$

Since:

$$v = \underbrace{\underset{B}{\pi}(v)}_{\in W} + \underbrace{v - \underset{B}{\pi}(v)}_{\in W^{\perp}} = \underbrace{\underset{B'}{\pi}(v)}_{\in W} + \underbrace{v - \underset{B'}{\pi}(v)}_{\in W^{\perp}}$$

By theorem 8.3, $V = W \oplus W^{\perp}$, so v has a single, unique representation as a sum of an element in W and $W^{\perp}$. So:

$$v - \underset{B}{\pi}(v) = v - \underset{B'}{\pi}(v) \implies \underset{B}{\pi}(v) = \underset{B'}{\pi}(v) \quad \blacksquare$$