LECTURE 17

The Theory of a Single Endomorphism

Recall that an *endomorphism* is a map $T : V \longrightarrow V$ is a linear transformation between a vector space V and itself.

Let $L(V,V)$ denote the set of endomorphisms of a finite dimensional vector space V. This set has the natural structure of a vector space with addition and scalar multiplication being defined as usual for a set of functions between two vectors spaces:

$$\begin{aligned} T, T' \in L(V,V) &\Rightarrow T + T' \in L(V,V) \text{ is defined by } (T+T')(v) = T(v) + T'(v) \\ \lambda \in \mathbb{F} \;,\; T' \in L(V,V) &\Rightarrow \lambda T \in L(V,V) \text{ is defined by } (\lambda T)(v) = \lambda T(v) \;. \end{aligned}$$

Recall that once we choose a basis $B = \{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ for V, every such linear transformation can be represented as an $n \times n$ matrix

$$\mathbf{T}_{B,B} = \left(\left[T(\mathbf{v}_1)_B, \cdots, T(\mathbf{v}_n)_B\right]\right)$$

(that is, is to say we form an $n \times n$ matrices by using the coordinate vector $T(\mathbf{v}_i)$ with respect to the basis B as the i^{th} column). In fact, once we fix a basis we have via this correspondence an isomorphism between $L(V,V)$ and $M_{n\times n}(\mathbb{F}) \cong \mathbb{F}^{n^2}$. This argument shows that $L(V,V)$ is n^2-dimensional if V is n-dimensional.

Now let fix our attention on a single endomorphism $T : V \rightarrow V$. The composition $T^2 = T \circ T$ of T with itself is another endomorphism of V, as are the compositions

$$\begin{aligned} T^3 &= T \circ T \circ T \\ T^4 &: = T \circ T \circ T \circ T \\ &\vdots \end{aligned}$$

Each power of T; $T^2, T^3, T^4, \ldots$ is thus another "vector" in the vector space $L(V,V)$. Since $L(V,V)$ is n^2-dimensional, it must be that the $n^2 + 1$ linear transformations

$$\mathbf{1}, T \;,\; T^2 \;,\; \ldots \;,\; T^{n^2}$$

are linearly dependent. (Here $\mathbf{1} : V \rightarrow V$ is the identity transformation of V defined by $\mathbf{1}(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$.) Thus, there must be some equation of the form

$$\alpha_0 \mathbf{1} + \alpha_1 T + \alpha_2 T^2 + \cdots + \alpha_{n^2} T^{n^2} = \mathbf{0}_{L(V,V)} \quad .$$

(*)

This looks a lot like a polynomial equation. And in fact, under a suitable definition of what a polynomial is be it is exactly a polynomial equation.

1. Digression: Polynomials

Definition 17.1. *An* ***indeterminant*** *is symbol x that generates a list of symbols x^0, x^1, x^2,via the multiplication rule*

$$x^i x^j = x^j x^i = x^{i+j}$$

It is most common to denote x^0 by 1 and to denote x^1 by x.

DEFINITION 17.2. *Let $\mathbb{F}$ be a field and x an indeterminate. The set of polynomials over $\mathbb{F}$ is the set $\mathbb{F}[x]$ of formal expressions of the form*

$$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \quad , \quad a_1, \ldots, a_n \in \mathbb{F} \quad , \quad n < \infty \quad .$$

Two polynomials

$$\begin{gathered} a_0 + a_1 x + \cdots + a_n x^n \\ b_0 + b_1 x + \cdots + b_m x^m \end{gathered}$$

say, with $m \leq n$, are regarded the same if $a_i = b_i$ for all i between 0 and m, and $a_j = 0$ for all j such that $m < j \leq n$.

This set is given the structure of a vector space over $\mathbb{F}$ via the following definitions of scalar multiplication and vector addition.

$$\lambda \cdot (a_0 + a_1 x) + \cdots + a_n x^2 = (\lambda a_0) + (\lambda a_1)\, x + \cdots + (\lambda a_n)\, x^n \quad , \quad \forall\ \lambda \in \mathbb{F} \quad .$$

$$\left(a_0 + a_1 x + \cdots + a_n x^2\right) + \left(b_0 + b_1 x + \cdots + b_n x^n\right) = (a_0 + b_0) + (a_1 + b_1)\, x + \cdots + (a_n + b_n)\, x^n$$

(where we may have added terms of the form $0_{\mathbb{F}} \cdot x^j$ to one or both of these polynomials to express it as a sum of $n+1$ terms). In addition, the set $\mathbb{F}[x]$ is given a multiplicative structure via

$$\left(a_0 + a_1 x + \cdots + a_n x^2\right) \cdot \left(b_0 + b_1 x + \cdots + b_m x^m\right) = \sum_{i=0}^{n} \sum_{j=0}^{m} a_i b_j x^{i+j} \quad .$$

REMARK 17.3. It is also possible to define polynomials without indeterminants as in

DEFINITION 17.4. *The set of polynomials over a field $\mathbb{F}$ is the set of finite sequences $[a_0, a_1, \ldots, a_n]$ of elements of $\mathbb{F}$ modulo equivalence relations of the form*

$$[a_0, a_1, \ldots, a_n, 0_{\mathbb{F}}, \ldots, 0_{\mathbb{F}}] = [a_0, a_1, \ldots, a_n]$$

The place of the last nonzero entry of a polynomial is called the ***degree*** *of the polynomial. The set of polynomials over $\mathbb{F}$ is naturally an (infinite-dimensional) vector space over $\mathbb{F}$ with scalar multiplication defined by*

$$\lambda\, [a_0, a_1, \ldots, a_n] = [\lambda a_0, \lambda a_1, \ldots, \lambda a_n]$$

and vector addition defined by

$$[a_0, a_1, \ldots, a_n] + [b_0, b_1 + \cdots + b_m] = \begin{cases} [a_0 + b_0, a_1 + b_1, \ldots, a_m + b_m, a_{m+1}, \ldots, a_n] & ; \quad \text{if } n \geq m \\ [a_0 + b_0, a_1 + b_1, \ldots, a_n + b_n, b_{n+1}, \ldots, b_m] & ; \quad \text{if } m \geq n \end{cases}$$

The product of two polynomials $[a_0, \ldots, a_n]$ and $[b_0, \ldots, b_m]$ is the ordered list of $m + n + 1$ elements of $\mathbb{F}$ whose i element is

$$\sum_{0 \leq k, \leq i} a_{k+i} b_i$$

The rationale for introducing the notion of an indeterminant in the original defintion is two-fold. First of all, it allows us to write down the rule for multiplying polynomials in a readily remembered and useable form. That is, we can utilize the symbols $1, x, x^2, \ldots$ as the standard basis elements of the vector space of polynomials and then define multiplication of polynomials via

$$\left(\sum_{i=0}^{n} a_i x^i\right) \left(\sum_{j=0}^{m} b_j x^j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} (a_i b_j)\, x^{i+j}$$

Secondly, the notion of an indeterminate allows us to think of *evaluating* polynomials by replacing x by something more specific. However, we **are not** restricted to think of only replacing x by an element of the field $\mathbb{F}$. We just need to make sure that whatever we use to replace x by, that there are also suitable replacements for $x^2, x^3, \ldots$, consistent with the rule $x^i x^j = x^{i+j}$, and that we the multiplications λx^i where $\lambda \in \mathbb{F}$ are also well-defined. So in particular, we could think of replacing x by an $n \times n$ matrix, or even an endomorphism of a vector space as in (*). Thus, given a polynomial $f \in \mathbb{R}[x]$, we can get

- a function $f : \mathbb{R} \to \mathbb{R}$ by replacing the indeterminate x with a real reparameter
- a function $f : \mathbb{C} \to \mathbb{C}$ by replacing the indetermnate x with a complex parameter.
- a function $f : M_{2\times 2}(\mathbb{R}) \to M_{2\times 2}(\mathbb{R})$ by replacing the indeterminate x by a real 2×2 matrix.

Here's a striking application of this ability for replace indeterminates by different types of parameters.

Theorem 17.5 (Cayley-Hamilton Theorem). *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix and let*

$$P(\lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$$

where $\mathbf{A} - \lambda\mathbf{I}$ *is the* $n \times n$ *matrix formed from* $\mathbf{A}$ *by subtracting an indeterminate* λ *from its diagonal elements. Then*

$$P(\mathbf{A}) = \mathbf{0}_{Mat_{n\times n}}$$

(The proof of this theorem won't come until we reach §24 of the text.)

Example 17.6. Consider the case of a 2×2 matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \Rightarrow \quad \mathbf{A} - \lambda\mathbf{I} = \begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix}$$

and so

$$\Rightarrow \quad P(\lambda) = \det(\mathbf{A} - \lambda\mathbf{I}) = (a-\lambda)(d-\lambda) - bc = (ad-bc)\lambda^0 + (-a-d)\lambda - d\lambda + \lambda^2$$

Now if we replace λ^0 by $\mathbf{A}^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, λ by $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and λ^2 by $\mathbf{A}^2 = \begin{pmatrix} a^2+bc & ab+bd \\ ac+cd & d^2+bc \end{pmatrix}$, we find

$$\begin{aligned} P(\mathbf{A}) &= (ad-bc)\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (-a-d)\begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} a^2+bc & ab+bd \\ ac+cd & d^2+bc \end{pmatrix} \\ &= \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix} + \begin{pmatrix} -a^2-ad & -ab-db \\ -ac-dc & -ad-d^2 \end{pmatrix} + \begin{pmatrix} a^2+bc & ab+bd \\ ac+cd & d^2+bc \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \end{aligned}$$

2. Back to Endomorphisms

OK, we have just discovered that each endomorphism $T \in L(V,V)$ satisfies some polynomial equation; and what this means is that there is a polynomial $p \in \mathbb{F}[x]$ such that when replace the indeterminate x with T and interprete the resulting algebraic expression as another endomorphism of V, then $f(T)$ is the zero endomorphism $\mathbf{0}_{L(v,v)}(v) = \mathbf{0}_V$ for all $v \in V$.

In fact, we can in a similar fashion "evaluate" any polynomial $f \in \mathbb{F}[x]$ at $x = T$. The following rules of evaluation follow directly from the two defining properties of an endomorphism (i.e., compatibility with scalar multiplication and vector addition):

Lemma 17.7. *Let* $T \in L(V,V)$ *and let* $f, g \in \mathbb{F}[x]$. *Then*

(a) $f(T)T = Tf(T)$
(b) $(f \pm g)(T) = f(T) \pm g(T)$
(c) $(fg)(T) = f(T)g(T)$

Ok. Let's now tighten up our observation that every endomorphism satsifies a polynomial equation.

THEOREM 17.8. *Let $T \in L(V,V)$. Then there exists a polynomial $m(x) \in F[x]$ of the form*

$$m(x) = x^r + \xi_{r-1}x^{r-1} + \cdots + \xi_1 x + \xi_0$$

with the properties

1. $m(T) = 0$
2. *If $f(x)$ is any polynomial in $F[x]$ such that $f(T) = \mathbf{0}$, then $m(x)$ divides $f(x)$ in $F[x]$ (meaning $f(x) = q(x)\,m(x)$ for some polynomial $q(x) \in \mathbb{F}[x]$).*

Proof. We have already seen that the endomorphisms $\mathbf{1}, T, T^2, \ldots, T^{n^2}$ must be linearly dependent (since $\dim L(V,V) = n^2$). Let r be the smallest integer such that

$$\mathbf{1}, T, T^2, \ldots, T^{r-1} \qquad \text{are linearly independent}$$

while

$$\mathbf{1}, T, T^2, \ldots, T^{r-1}, T^r \qquad \text{are linearly dependent} \quad .$$

Then there must be a dependence relation of the form

$$\xi_0\mathbf{1} + \xi_1 T + \cdots + \xi_{r-1}T^{r-1} + T^r = \mathbf{0}$$

So setting

$$m(x) = x^r + \xi_{r-1}x^{r-1} + \cdots + \xi_1 x + \xi_0$$

will have the stated form and will satisfy 1.

Now suppose f is any other polynomial such that $f(T)$ equal $\mathbf{0}$, Because $\mathbf{1}, T, \ldots, T^{r-1}$ are linearly independent, there can not exist a polynomial of lower degree $< r$ that satisfies $f(T) = \mathbf{0}$. By the Division Algorithm for polynomials, there exist unique polynomials $q(x)$ and $r(x)$ such that

$$f(x) = q(x)\,m(x) + r(x) \quad \text{and either } r(x) = 0 \text{ or } \deg(r) < \deg(m)$$

Suppose $f(x) = q(x)\,m(x) + r(x)$ and $\deg(r) < \det(m)$. Then

$$\mathbf{0} = f(T) = q(T)\,m(T) + r(T) = q(T)\cdot\mathbf{0} + r(T) \quad \Rightarrow \quad r(T) = \mathbf{0}$$

which contradicts the condidition that m is the polynomial of lowest degree that evaluates to $\mathbf{0}$ when substitute T for x. Thus, it must be the case that $r(x) = 0$. Hence,

$$f(x) = q(x)\,m(x).$$

DEFINITION 17.9. *Let $T \in L(V,V)$. The polynomial $m(x) \in \mathbb{F}[x]$ defined in the preceding theorem is called the* ***minimal polynomial*** *of T: $m(x)$ is characterized as the non-zero polynomial of minimal degree such that $m(T) = \mathbf{0}$. If we normalize the leading coefficient to $1_{\mathbb{F}}$, then $m(x)$ is unique.*

EXAMPLES 17.10. (1) Let $\mathbf{T} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Then

$$\mathbf{T}^2 = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{T}^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

So if $f(x) = x^3$. We have $f(T) = 0$. In fact, $f(x) = x^3$ is the minimal polynomial of $\mathbf{T}$. This is because the minimal polynomial $m_{\mathbf{T}}(x)$ of $\mathbf{T}$ has to divide $f(x)$. But the only nontrivial factors of $f(x) = x^3$ are x and x^2 and neither of these vanish when evaluated at $\mathbf{T}$. Thus, $m_T(x) = x^3$

(2) Let $\mathbf{T} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. Then

$$\mathbf{T}^2 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix} = 2 \cdot \mathbf{T}$$

so

$$\mathbf{T}^2 - 2\mathbf{T} = 0$$

The minimal polynomial of $\mathbf{T}$ must therefore divide $x^2 - 2x$. There are three possibilities:

$$x^2 - 2x = x\,(x-2)$$

So $m_{\mathbf{T}}$ could be x, $x-2$, or $x\,(x-2)$. It can't be x since $\mathbf{T} \neq \mathbf{0}$ However,

$$\mathbf{T} - 2\mathbf{I} = \mathbf{0}$$

So $m_{\mathbf{T}}\,(x)$ must be the minimal polynomial of $\mathbf{T}$.

3. Eigenvectors and Eigenvalues

It will take us a while before we begin to identify the minimal polynomial of a given endomorphism.

We shall start with a simpler problem. Rather than try to identify the "smallest" polynomial $f\,(x)$ such that $f\,(T) \in L\,(V,V)$ sends each vector $v \in V$ to $\mathbf{0}_V$, we shall look for the simplest sort of polynomial that can send a particular non-zero vector to $\mathbf{0}_V$. Now such a polynomial can not be a polynomial of degree 0; because that would mean $f\,(x) = a \in \mathbb{F}$, with $a \neq 0_{\mathbb{F}}$. And so

$$f\,(T)\,v = (a \cdot \mathbf{1})\,(v) = av = 0 \text{ only if } v = \mathbf{0}_V.$$

So the simplest possibility would be a linear polynomial of the form $f\,(x) = x - \lambda$.

DEFINITION 17.11. *A non-zero vector v is an* ***eigenvector*** *of an endomorphism $T : V \to V$ if there exists a $\lambda \in \mathbb{F}$ such that $(T - \lambda\mathbf{1})\,v = \mathbf{0}_V$. The corresponding λ is called the* ***eigenvalue*** *of T corresponding to v.*

THEOREM 17.12. *Suppose $(v_1, \lambda_1), (v_2, \lambda_2), \ldots, (v_k, \lambda_k)$ are eigenvector/eigenvalue pairs for a linear transformation $T : V \to V$ and that the eigenvalues $\lambda_1, \ldots, \lambda_k$ are all distinct. Then the corresponding eigenvectors are linearly independent.*

Proof. Suppose first that $k = 1$. Then $\{v_1\}$ is a linearly independent set of eigenvectors and so the statement is true. Now suppose the statement is true for all $k < n$. Suppose there was a dependence relation amongst the eigenvectors $v_1, \ldots, v_n$. Then we'd have

$$a_1 v_1 + \cdots + a_n v_n = \mathbf{0}_V \tag{*}$$

Suppose that some coefficient $a_i \neq 0_{\mathbb{F}}$. We shall show that this contradicts the induction hypothesis. Applying T to both sides of (*) we get

$$a_1 \lambda_1 v_1 + \cdots + a_i \lambda_i v_i + \cdots + a_n \lambda_n v_n = \mathbf{0}_V \tag{**}$$

Multiplying (*) by λ_i and subtracting it from (**) yields

$$a_1\,(\lambda_1 - \lambda_i)\,v_1 + \cdots + a_i\,(\lambda_i - \lambda_i)\,v_i + \cdots + a_n\,(\lambda_n - \lambda_i)\,v_n = \mathbf{0}_V$$

Since the i^{th} term on the left hand side of this last expression vanishes identically, there are at most $n-1$ non-zero terms on the left hand side. By the induction hypothesis, the vectors $v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n$ must be linearly independent. So we must have

$$a_1\,(\lambda_1 - \lambda_i) = 0_{\mathbb{F}} \quad , \quad \ldots \quad , \quad a_{i-1}\,(\lambda_{i-1} - \lambda_i) = 0_{\mathbb{F}} \quad , \quad a_{i+1}\,(\lambda_{i+1} - \lambda_i) = 0_{\mathbb{F}} \quad , \quad \ldots \quad , \quad a_n\,(\lambda_n - \lambda_i) = 0_{\mathbb{F}}$$

Since the $\lambda_i \neq \lambda_j$ for $i \neq j$, we must have

$$a_1 = 0_{\mathbb{F}} \quad , \quad \ldots \quad , \quad a_{i-1} = 0_{\mathbb{F}} \quad , \quad a_{i+1} = 0_{\mathbb{F}} \quad , \quad \ldots \quad , \quad a_n = 0_{\mathbb{F}}$$

But then the original expression (*) says

$$\mathbf{0}_V + \mathbf{0}_v + \cdots + \cdots + \mathbf{0}_V + a_i v_i + \mathbf{0}_V + \cdots + \mathbf{0}_V = \mathbf{0}_V \quad \Rightarrow \quad a_i = 0_{\mathbb{F}}$$

Which contradicts our hypothesis that $a_i \neq 0_{\mathbb{F}}$. Thus, there can be no dependence relation amongst the vectors $v_1, \dots, v_n$. $\square$

Application

Here is a nice application of this theorem. Let $\{r_1, \dots, r_k\}$ be a list of distinct real numbers and consider the corresponding exponential functions $e^{r_1 x}, e^{r_2 x}, \dots, e^{r_k x}$. Show that these exponential functions regarded as vectors in the vector space $C^\infty(\mathbb{R})$ of smooth differentiable functions on $\mathbb{R}$ are linearly independent.

Well, we have seen that the differential operator $\frac{d}{dx} : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ is an endomorphism of $C^\infty(\mathbb{R})$. Now note that each exponential function $e^{r_i x}$ is an eigenvector for $\frac{d}{dx}$ with eigenvalue r_i.

$$\frac{d}{dx} e^{r_i x} = r_i e^{r_i x} \quad .$$

Since the functions $e^{r_1 x}, \dots, e^{r_k x}$ have distinct eigenvalues with respect to $\frac{d}{dx}$ they must be linearly independent. $\square$

4. Digression: Determinants of Endomorphisms

The determinant function as we defined it in Lecture 16 is only applicable to square matrices. On the other hand, once we have a basis $B = \{v_1, \dots, v_n\}$ for V we can attach to any endomorphism $T : V \to V$ a square matrix

$$\mathbf{A}_{T,B} = \begin{pmatrix} | & & | \\ T(v_1)_B & \cdots & T(v_n)_B \\ | & & | \end{pmatrix}$$

Here $T(v_i)_B$, $i = 1, \dots n$, is the coordinate vector of the vector $T(v_i)$ with respect to the basis B. This furnishes us with an $n \times n$ matrix that will allow us to define **a** determinant function for T; we set

$$Det_B(T) = \det(\mathbf{A}_{T,B})$$

If we had adopted another basis B' for V, we could end up with a very different matrix $\mathbf{A}_{T,B'}$, and so it could happen that our notion of a determinant for T depends not only on T but on the basis B. However, it turns out that $Det_B(T)$ is independent of B.

To see this, recall the Change-Of-Basis theory developed in Lecture 15. Suppose we have two bases $B = \{v_1, \dots, v_n\}$ and $B' = \{v'_1, \dots, v'_n\}$ for V. Then by expressing the vectors v'_i as as linear combinations of the vectors v_j and reading off the coefficients we can form a change-of-basis matrix $\mathbf{C}_{B,B'}$ that maps coordinate vectors with respect to B directly to the coordinate vectors with respect to B' :

$$\begin{tikzcd}
& \mathbf{v}_B \arrow[dd, "\mathbf{C}_{B,B'}"] \\
V \ni v \arrow[ur, "i_B"] \arrow[dr, "i_{B'}"'] & \\
& \mathbf{v}_{B'} = \mathbf{C}_{B,B'}\mathbf{v}_B
\end{tikzcd}$$

We can use this change of basis matrix to relate the matrices $\mathbf{A}_{T,B}$ and $\mathbf{A}_{T,B'}$. Let v be an arbitrary vector in V. The whole point of the matrices $\mathbf{A}_{T,B}$ and $\mathbf{A}_{T,B'}$ is that they allow us to implement the linear transformation T directly on coordinate vectors:

$$\begin{aligned} T(v)_B &= \mathbf{A}_{T,B}\mathbf{v}_B && \text{((i))} \\ T(v)_{B'} &= \mathbf{A}_{T,B'}\mathbf{v}_{B'} && \text{((ii))} \end{aligned}$$

So, on the one hand,

$$\mathbf{A}_{T,B}\mathbf{v}_B = \mathbf{A}_{T,B}(\mathbf{C}_{B'B}\mathbf{v}_{B'}) = \mathbf{A}_{T,B}\mathbf{C}_{B',B}\mathbf{v}_{B'}$$

and on the other

$$\begin{aligned} T(v)_{B'} &= \mathbf{C}_{B,B'} T(v)_B \\ &= \mathbf{C}_{B,B'} \mathbf{A}_{T,B} \mathbf{v}_B \\ &= \mathbf{C}_{B,B'} \mathbf{A}_{T,B} \mathbf{C}_{B',B} \mathbf{v}_{B'} \end{aligned}$$

But then (ii) implies

$$\mathbf{A}_{T,B'} = \mathbf{C}_{B,B'} \mathbf{A}_{T,B} \mathbf{C}_{B',B} \mathbf{v}_{B'}$$

Now let's compute the determinant of both sides

$$\begin{aligned} \det(\mathbf{A}_{T,B'}) &= \det(\mathbf{C}_{B,B'} \mathbf{A}_{T,B} \mathbf{C}_{B',B} \mathbf{v}_{B'}) \\ &= \det\left(\mathbf{C}_{B,B/}\right) \det(\mathbf{A}_{T,B}) \det(\mathbf{C}_{B',B}) \end{aligned}$$

But

$$\mathbf{C}_{B,B'} = (\mathbf{C}_{B',B})^{-1}$$

and so

$$\det(\mathbf{C}_{B,B'}) = \frac{1}{\det(\mathbf{C}_{B',B})}$$

whence

$$\det(\mathbf{A}_{T,B'}) = \det(\mathbf{A}_{T,B})$$

We conclude:

THEOREM 17.13. *The function $Det : L(V,V) \to \mathbb{F}$ defined by*

$$Det(T) = \det(\mathbf{A}_{T,B}) \qquad \textit{for some basis } B \textit{ of } V$$

is independent of the choice of B.

5. Back to Eigenvectors and Eigenvalues

THEOREM 17.14. *Let V be a finite-dimensional vector space and let $T \in L(V,V)$ and let $\lambda \in \mathbb{F}$. Then λ is an eigenvalue of T if and only if the determinant of $T - \lambda\mathbf{1}$ is equal to $0_{\mathbb{F}}$.*

Proof.

First, suppose λ is an eigenvalue of T. So there is a corresponding $v \in V$ such that $T(v) = \lambda v$. Then the vector v lies in the kernel of the operator $T - \lambda\mathbf{1}$; for

$$(T - \lambda\mathbf{1})(v) = T(v) - \lambda v = \lambda v - \lambda v = \mathbf{0}_V \quad .$$

Since $T - \lambda\mathbf{1}$ has a non-trivial kernel, it must have determinant $0_{\mathbb{F}}$.

Conversely, suppose $T - \lambda\mathbf{1}$ has determinant $0_{\mathbb{F}}$. Then then it has a non-trivial kernel. So we can find a non-zero vector $v \in \ker(T - \lambda\mathbf{1})$. But then

$$\mathbf{0}_V = (T - \lambda\mathbf{1})(v) = T(v) - \lambda v \quad \Rightarrow \quad T(v) = \lambda v$$

and so λ is an eigenvalue of T.

5.1. Examples. We'll start with some matrix examples as that is anyway the arena where more general examples are to be calculated.

The basic procedure for finding the eigenvalues and eigenvectors of a matrix $\mathbf{A}$ is as follows:

- Find the roots of the polynomial equation $\det(\mathbf{A} - \lambda\mathbf{1}) = 0$.

- Each λ satisfying $\det(\mathbf{A}-\lambda\mathbf{1})$ will be an eigenvalue. To find the corresponding eigenvectors we look for solutions of the linear system
$$(\mathbf{A}-\lambda\mathbf{I})\,\mathbf{x}=\mathbf{0}$$
This is a homogeneous linear system. It is (hopefully, by now) straight-forward to find a basis for solution space of such a homogeneous linear system. The vectors in this basis can be used as a complete set of linearly independent eigenvectors with eigenvalue λ. However, if you have multiple eigenvalues, you will have to carry out such a computation for each eigenvalue.

EXAMPLE 17.15. Find the eigenvalues and eigenvectors of $\mathbf{A}=\begin{pmatrix}3&0&0\\0&2&0\\0&0&2\end{pmatrix}$.

- We have
$$\det(\mathbf{A}-\lambda\mathbf{1})=\det\begin{pmatrix}3-\lambda&0&0\\0&2-\lambda&0\\0&0&2-\lambda\end{pmatrix}=(3-\lambda)(2-\lambda)^2$$
So the possible eigenvalues are the roots of $(3-\lambda)(2-\lambda)^2=0$; that is, $\lambda=3,2$.

For the eigenvalue $\lambda=3$ we find the corresponding eigenvectors by solving $(\mathbf{A}-3\mathbf{1})\,\mathbf{x}=\mathbf{0}$; or
$$\begin{pmatrix}0&0&0\\0&-1&0\\0&0&-1\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}\quad\Rightarrow\quad\begin{cases}0=0\\-x_2=0\\-x_3=0\end{cases}\quad\Rightarrow\quad x_2=0=x_3.$$
Thus, a solution, must have $x_2=x_3=0$. The component x_1, however, is free. And so, any vector of the form
$$\mathbf{x}=\begin{pmatrix}x_1\\0\\0\end{pmatrix}=x_1\begin{pmatrix}1\\0\\0\end{pmatrix}$$
will be an eigenvector of $\mathbf{A}$ with eigenvalue 3. We typically state the answer though in terms of the basis vector $\begin{pmatrix}1\\0\\0\end{pmatrix}$.

For the eigenvalue $\lambda=2$. We proceed in the same way,
$$(\mathbf{A}-2\mathbf{1})\,\mathbf{x}=\mathbf{0}\quad\Rightarrow\quad\begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}\quad\Rightarrow\quad x_1=0$$
The components x_2 and x_3 are free, however, Thus, the general solution will be
$$\mathbf{x}=\begin{pmatrix}0\\x_2\\x_3\end{pmatrix}=x_2\begin{pmatrix}0\\1\\0\end{pmatrix}+x_3\begin{pmatrix}0\\0\\1\end{pmatrix}$$
As before we usually just state the basis vectors for this solution space as our eigenvectors. So we have two (linearly independent) eigenvectors corresponding to the eigenvalue $=2$. In such a situation, we say that the $\lambda=2$ *eigenspace* is 2-dimensional. Note however that any vector of the form
$$s\begin{pmatrix}0\\1\\0\end{pmatrix}+r\begin{pmatrix}0\\0\\1\end{pmatrix}$$
will be an eigenvector of $\mathbf{A}$ with eigenvalue 2. □

REMARK 17.16. In fact, it is more accurate to think in terms of eigenvalues and eigenspaces, rather than eigenvalues and eigenvectors. For it is easy to see that if v_1 and v_2 are two eigenvectors with the same eigenvalue λ, then any linear combination of v_1 and v_1 will be an eigenvector with eigenvalue λ.

EXAMPLE 17.17. Find the eigenvalues and eigenvectors of $\mathbf{A} = \begin{pmatrix} 3 & -2 & 5 \\ 1 & 0 & 7 \\ 0 & 0 & 2 \end{pmatrix}$., characteristic polynomial: $X^3 - 5X^2 + 8X - 4 = (X-1)(X-2)^2$

- We have
$$\det(\mathbf{A} - \lambda\mathbf{1}) = \det \begin{pmatrix} 3-\lambda & -2 & 5 \\ 1 & -\lambda & 7 \\ 0 & 0 & 2-\lambda \end{pmatrix} = -\lambda^3 + 5\lambda^2 - 8\lambda + 4 = (1-\lambda)(2-\lambda)^2$$
So we apparently we have two eigenvalues $\lambda = 1$ and $\lambda = 2$.
- $\lambda = 1$
  We need to solve $(\mathbf{A} - \mathbf{1})\,\mathbf{x} = \mathbf{0}$.
$$\begin{pmatrix} 2 & -2 & 5 \\ 1 & -1 & 7 \\ 0 & 0 & 2 \end{pmatrix} \xrightarrow{\text{row reduce}} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$
From the reduced row echelon form of $(\mathbf{A} - \mathbf{1})$ we see that vectors in the solution solution will satisfy
$$x_1 = x_2 \quad , \quad x_3 = 0$$
Thus, the eigenvectors for $\lambda = 1$ will be of the form
$$\mathbf{x} = x_2 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$
- $\lambda = 2$
  We need to solve $(\mathbf{A} - 2 \cdot \mathbf{1})\,\mathbf{x} = \mathbf{0}$. Proceeding as we did for $\lambda = 1$:
$$\begin{pmatrix} 1 & -2 & 5 \\ 1 & -2 & 7 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{\text{row reduce}} \begin{pmatrix} 1 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$
Thus, our eigenvectors for $\lambda = 2$ will be vectors $[x_1, x_2, x_3]$ satisfying
$$x_1 = 2x_2 \quad , \quad x_3 = 0$$
- In summary,
$$\begin{aligned} \text{eigenvalue } \lambda &= 1 \qquad \text{with eigenspace } span\left(\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}\right) \\ \text{eigenvalue } \lambda &= 2 \qquad \text{with eigenspace } span\left(\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}\right) \end{aligned}$$

□

5.2. Algebraic and Geometric Multiplicities.

Let's compare the preceding two examples.

In the first example, the characteristic polynomial was $(3-\lambda)(2-\lambda)^2$, and we found one (linearly independent) eigenvector for the eigenvalue $\lambda = 3$ and two (linearly independent) eigenvectors for the eigenvalue $\lambda = 2$.

In the second example, the characteristic polynomial was $(1-\lambda)(2-\lambda)^2$ and we found one (linearly independent) eigenvector for the eigenvalue $\lambda = 1$ and only one (linearly independent) eigenvector for the eigenvalue $\lambda = 2$.

This raises the question, how many eigenvectors should we expect for a given eigenvalue?

To answer this question (albeit incompletely), we need a little more terminology.

If r is an eigenvalue of an $n \times n$ matrix $\mathbf{A}$, then it is a solution of the polynomial equation $\det(\mathbf{A} - \lambda\mathbf{1}) = 0$. By the Fundamental Theorem of Algebra, this means that $\lambda - r$ is a factor of $\det(\mathbf{A} - \lambda\mathbf{1})$. This same factor may appear multiple times, however, in a complete factorization of $\det(\mathbf{A} - \lambda\mathbf{1})$.

DEFINITION 17.18. *The* ***algebraic multiplicity*** *of an eigenvalue r of a matrix is the largest integer m such that $(\lambda - r)^m$ divides* $\det(\mathbf{A} - \lambda\mathbf{1})$.

Thus, in the first example, where $\det(\mathbf{A} - \lambda\mathbf{1}) = (3 - \lambda)(2 - \lambda)^2$, the eigenvalue 3 has algebraic multiplicity 1 and the eigenvalue 2 has algebraic multiplicity 2.

DEFINITION 17.19. *The* ***geometric multiplicity*** *of an eigenvalue r of a matrix is the dimension of the corresponding eigenspace.*

Thus, in the first example, since the $\lambda = 2$ eigenspace has two linearly independent eigenvectors, we say the eigenvalue $\lambda = 2$ has geometric multiplicity 2.

In the second example, the $\lambda = 2$ eigenspace has only one linearly independent eigenvector, so its geometric multiplicity is 1. (Note, the algebraic multiplicity of the $\lambda = 2$ eigenspace remains 2).

In general, so for any eigenvalue λ of an $n \times n$ matrix we have the following inequalities

$$1 \leq \text{geometric multiplicity} \leq \text{algebraic multiplicity} \leq n$$

REMARK 17.20. When the geometric multiplicity of an eigenvalue λ is > 1 we often refer to λ as a **degenerate eigenvalue** or this circumstance as a **degeneracy**.

5.3. Complex Eigenvalues and Eigenvectors. The next example will show that matrices do not always have eigenvalues and eigenvectors.

EXAMPLE 17.21. Consider the matrix $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. The characteristic polynomial for this matrix is

$$\det(\mathbf{A} - \lambda\mathbf{1}) = \det\begin{pmatrix} -\lambda & 1 \\ -1 & -\lambda \end{pmatrix} = \lambda^2 + 1 \neq 0 \quad \text{for any real value of } \lambda$$

So if we were to insist on thinking of $\mathbf{A}$ as acting on the real vector space $\mathbb{R}^2$, we are not going to have any eigenvectors or eigenvalues.

On the other hand, we could instead relax our restriction that the underlying field consists only of real numbers. Doing so, we'd have two eigenvalues

$$\lambda^2 + 1 = 0 \quad \Rightarrow \quad \lambda = \pm\sqrt{-1} \quad \Rightarrow \quad \lambda = i, -i$$

Can we find the corresponding eigenvectors?

Certainly, just be following the precedure as above.

- $\lambda = i$

  First we find the solution set of $(\mathbf{A} - i\mathbf{1}) = 0$

$$\begin{pmatrix} -i & 1 \\ -1 & -i \end{pmatrix} \xrightarrow{R_1 \to iR_1} \begin{pmatrix} 1 & i \\ -1 & -i \end{pmatrix} \xrightarrow{R_2 \to R_2 + R_1} \begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix}$$

  So we've reached a matrix in reduced row echelon form. The corresponding equations will be

$$x_1 = -ix_2$$

  or

$$\mathbf{x} = x_2 \begin{pmatrix} -i \\ 1 \end{pmatrix}$$

Note that this eigenvector is not interpretable as a vector in $\mathbb{R}^2$, but it is interpretable as a vector in $\mathbb{C}^2$.

- $\lambda = -i$

Proceding in the same way,

$$\mathbf{A} - (-i)\,\mathbf{1} = \begin{pmatrix} i & 1 \\ -1 & i \end{pmatrix} \xrightarrow{R_1 \to -iR_1} \begin{pmatrix} 1 & -i \\ -1 & i \end{pmatrix} \xrightarrow{R_2 \to R_2 + R_1} \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix}$$

And so

$$x_1 = ix_2 \quad \Rightarrow \quad \mathbf{x} = x_2 \begin{pmatrix} i \\ 1 \end{pmatrix}$$

Thus, we have a complex eigenvalue $\lambda = i$ with complex eigenvector $\begin{pmatrix} -i \\ 1 \end{pmatrix}$ and a complex eigenvalue $\lambda = -i$ with complex eigenvector $\begin{pmatrix} i \\ 1 \end{pmatrix}$.

Remark 17.22. In many physical applications, the eigenvalues of a particular matrix have direct interpretation as a real measurable quantity (for example, the moment of inertia of a body about a natural axis or rotation). In such situations, the occurence of complex eigenvectors might put the physical interpretation of the matrix in jeopardy. What usually saves the day in such situations is that the matrix for which eigenvalues is **symmetric**, i.e. $\mathbf{A} = \mathbf{A}^t$. Maybe latter in the course we will prove

Theorem 17.23. *If $\mathbf{A}$ is a symmetric $n \times n$ matrix, then*

- *all the roots of* $\det(\mathbf{A} - \lambda\mathbf{1}) = 0$ *are real (and so $\mathbf{A}$ has only real eigenvalues).*
- *$\mathbf{A}$ has a total of n linearly independent eigenvectors. (Put another way, the sum of the geometric multiplicities of the eigenvalues of $\mathbf{A}$ is n).*