Polynomials

In order to do what we need to do, it turns out polynomials will be key, so, lets spend a bit of time recalling some *basics*.
Recall that a polynomial (over $\mathbb{R}$ or $\mathbb{C}$) is just an expression of the form:

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

where each of the a_i are numbers (in $\mathbb{R}$ or $\mathbb{C}$).

The **degree** of a polynomial is the largest number n such that $a_n \neq 0$.

A degree n polynomial is **monic** if $a_n = 1$.

The roots of $P(x)$ are the values λ (in $\mathbb{C}$) for which $P(\lambda) = 0$.

$$P(x) = x^2 - 5x + 6 = (x-2)(x-3)$$

is a degree 2 monic polynomial with roots 2 and 3.

Polynomial Long Division

Given two polynomials $N(x)$ and $D(x)$ with $D(x) \neq 0$ there are polynomials $Q(x)$ and $R(x)$ (the quotient and the remainder) such that

$$N(x) = Q(x)D(x) + R(x)$$

and the degree of $R(x)$ is strictly less than the degree of $D(x)$.

We say that $D(x)$ **divides** $N(x)$ if $N(x) = Q(x)D(x)$ or equivalently in the above, if $R(x) = 0$.

Proposition
$(x - \lambda)$ divides $P(x)$ if and only if $P(\lambda) = 0$

Example:
With $N(x) = x^2 + 3x + 3$ and $D(x) = (x + 2)$ we have:

$$x^2 + 3x + 3 = (x + 1)(x + 2) + 1$$

so $Q(x) = (x + 1)$ and $R(x) = 1$.

Polynomial GCD/LCM

The **greatest common divisor** of two polynomials $A(x)$ and $B(x)$ is the largest degree (monic) polynomial $D(x)$ such that

$$D(x)|A(x) \qquad D(x)|B(x)$$

The **least common multiple** of two polynomials $A(x)$ and $B(x)$ is the lowest degree (monic) polynomial $L(x)$ such that

$$A(x)|L(x) \qquad B(x)|L(x)$$

The following gives an alternative characterization of the gcd/lcm:

Theorem
If $P(x)$ is any polynomial that divides $A(x)$ and $B(x)$ then $P(x)$ divides the gcd of $A(x)$ and $B(x)$.
If $P(x)$ is any polynomial that is divisible by both $A(x)$ and $B(x)$ then $P(x)$ is divisible by the lcm of $A(x)$ and $B(x)$.

Example:
With $A(x) = (x+5)^2(x+2)$ and $B(x) = (x+5)(x+3)$ the gcd is $(x+5)$ and the lcm is $(x+5)^2(x+2)(x+3)$.

Euclidean Algorithm (Special case)

We say that two polynomial have **no common factors** if their gcd is 1.

Lemma

If $P_1(x)$ and $P_2(x)$ are polynomials with no common factors then there exists polynomials $S_1(x)$ and $S_2(x)$ so that

$$S_1(x)P_1(x) + S_2(x)P_2(x) = 1$$

Proof We prooceed by induction on the sum of the degrees of $P_1(x)$ and $P_2(x)$.
In the base case, both are degree 0, but if $P_1(x) = 0 = P_2(x)$, they have common factors (everything), so at least one of them is a non-zero constant. Which covers the base case.
For the inductive case, without loss of generality suppose $P_1(x)$ has degree **not smaller than** that of $P_2(x)$.By the division algorithm we can write:

$$P_1(x) = Q(x)P_2(x) + R(x)$$

where $R(x)$ has degree less than $P_1(x)$. We then know that $R(x)$ and $P_2(x)$ also have no common factors, because any common factor would need to be one of $P_1(x)$ aswell.
So by induction there exists $S_2(x)$ and $S_3(x)$ so that

$$S_3(x)P_2(x) + S_1(x)R(x) = 1$$

but then

$$1 = S_3(x)P_2(x) + S_1(x)R(x) = S_1(x)P_1(x) + (S_3(x) - S_1(x)Q(x))P_2(x)$$

which by setting $S_2(x) = S_3(x) - S_1(x)Q(x)$ gives the result.

Evaluating Polynomials at a Linear Transformation/Matrix

If $P(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ is a polynomial, so $a_i \in \mathbb{R}$ (or $\mathbb{C}$) then for any linear transformation $L : V \to V$ (respectively any square matrix) when we write $P(L)$ we mean:

$$P(L) = a_n L^n + a_{n-1}L^{n-1} + \cdots + a_1 L + a_0 \mathrm{Id}_V$$

where we recall that

$$L^\ell = \overbrace{L \circ L \circ \cdots \circ L \circ L}^{\ell} : V \to V$$

Notice that $P(L) : V \to V$ is a linear transformation because it is a linear combination of linear transformations!

$$\begin{aligned} P(L)(\vec{v}) &= a_n L^n(\vec{v}) + a_{n-1}L^{n-1}(\vec{v}) + \cdots + a_1 L + a_0 \mathrm{Id}_V(\vec{v}) \\ &= a_n L(L(L(\cdots(L(\vec{v}))\cdots))) + \cdots + a_1 L(\vec{v}) + a_0 \vec{v} \end{aligned}$$

Notice:

$$L^\ell \circ L^r = L^{\ell+r}$$

and by convention

$$L^0 = \mathrm{Id}_V.$$

Examples

We can do all the calculations with the associated matricies if we prefer.
Consider

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

If we have $P(x) = x^3 + 2x^2 + x + 3$ then

$$\begin{aligned} P(A) &= A^3 + 2A^2 + A + 3\mathrm{Id}_2 \\ &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^3 + 2\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^2 + \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + 3\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 4 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 3 & 0 \\ 0 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 7 & 8 \\ 0 & 7 \end{pmatrix} \end{aligned}$$

and so

$$P(A)((1,2)) = \begin{pmatrix} 7 & 8 \\ 0 & 7 \end{pmatrix}\begin{pmatrix} 1 \\ 2 \end{pmatrix} = (23, 14)$$

In practice we won't ever need to actually evaluate complicated polynomials at linear transformations, they are mostly a theoretical tool for proving things.

We are now going to give a few basic technical results about evaluating polynomials and linear transformations.

Lemma If $P(x)$ and $Q(x)$ are any two polynomials and $R(x) = P(x)Q(x)$ is their product then

$$P(L) \circ Q(L) = R(L)$$

Proof Idea: This follows from the distributive rule for compositions:

$$(a_r L^r + \cdots + a_1 L + a_0 \mathrm{Id}) \circ (b_\ell L^\ell + \cdots + b_1 L + b_0 \mathrm{Id})$$

The way we expand this product is the same as for polynomials.

Lemma
If $P(x)$ and $Q(x)$ are any two polynomials and $L : V \to V$ is any linear transformation then:

$$P(L) \circ Q(L) = Q(L) \circ P(L)$$

Proof Idea: Use the above and that $P(x)Q(x) = R(x) = Q(x)P(x)$.

Corollary
If we factor $P(X) = Q_1(x)Q_2(x)\cdots Q_m(x)$ then

$$P(L) = Q_1(L) \circ \cdots \circ Q_m(L)$$

and the order of factors can be changed.

Proof Idea: This follows from the above by induction

A very important case is if we factor $P(X) = \prod(x - \lambda_i)$ then

$$P(L) = (L - \lambda_1 \mathrm{Id}_n) \circ \cdots \circ (L - \lambda_m \mathrm{Id}_n)$$

and the order in which the roots are presented doesn't matter because these commute.

Lemma
Suppose $L : V \to V$ is any linear transformation.
For any polynomials $P_1(x)$ and $P_2(x)$ with $P_1(x)|P_2(x)$ we have

$$\mathrm{Ker}\,(P_1(L)) \subset \mathrm{Ker}\,(P_2(L)) \qquad \mathrm{Im}\,(P_2(L)) \subset \mathrm{Im}\,(P_1(L))$$

Proof Sketch If we write $P_2(x) = Q(x)P_1(x)$ then $P_2(L) = Q(L) \circ P_1(L) = P_1(L) \circ Q(L)$ and we note that we always have

$$\mathrm{Ker}(M) \subset \mathrm{Ker}(N \circ M) \quad \text{and} \quad \mathrm{Im}(M \circ N) \subset \mathrm{Im}(M).$$

In the first case, because if $\vec{v} \in \mathrm{Ker}(M)$ then $N \circ M(\vec{v}) = N(\vec{0}) = \vec{0}$.
In the second case, because if $\vec{v} \in \mathrm{Im}(M \circ N)$ so that $\vec{v} = M \circ N(\vec{w})$ then $\vec{v} = M(N(\vec{w}))$.
This now gives us the result.

Lemma If $P(x)$ and $Q(x)$ are any two polynomials and $L : V \to V$ is any linear transformation then:

$$P(L) \circ Q(L)$$

is injective (respectively surjective, respectively bijective) if and only if each of $P(L)$ and $Q(L)$ is injective (respectively surjective, respectively bijective)

Proof sketch:

First note: f and g injective (resp. surjective, bijective) implies $f \circ g$ injective (resp. surjective, bijective). These are useful facts about functions

Next note: $f \circ g$ injective (resp. surjective) implies g injective (resp. f surjective). These are useful facts about functions

Next note that because $P(L) \circ Q(L) = Q(L) \circ P(L)$ we can apply these results in both orders to conclude the if and only if statements.

Again an important case is if we factor $P(X) = \prod(x - \lambda_i)$ then $P(L)$ is injective (respectively surjective, respectively bijective) if and only if all of $(L - \lambda_1 \mathrm{Id}_n), \ldots, (L - \lambda_m \mathrm{Id}_n)$ are.

Lemma

Suppose $P(x)$ is any polynomial, and $L : V \to V$ any linear transformation. If $L(\vec{v}) = \lambda\vec{v}$ then $P(L)(\vec{v}) = P(\lambda)\vec{v}$.

Proof:

First we claim that for all ℓ we have $L^\ell(\vec{v}) = \lambda^\ell\vec{v}$, we prove this by induction.

The base case $\ell = 0$ is, we leave it as an exercise

For the inductive case we assume $\ell > 0$ and that $L^{\ell-1}(\vec{v}) = \lambda^{\ell-1}\vec{v}$ then we have:

$$L^\ell(\vec{v}) = L^{\ell-1}(L(\vec{v})) = L^{\ell-1}(\lambda\vec{v}) = \lambda L^{\ell-1}(\vec{v}) = \lambda(\lambda^{\ell-1}\vec{v})$$

and so

$$L^\ell(\vec{v}) = \lambda^\ell\vec{v}$$

by induction.

To complete the proof, now we may calculate that

$$\begin{aligned} P(L)(\vec{v}) &= a_\ell L^\ell(\vec{v}) + \cdots + a_1 L(\vec{v}) + a_0 \mathrm{Id}_V(\vec{v}) \\ &= a_\ell\lambda^\ell\vec{v} + \cdots + a_1\lambda\vec{v} + a_0\vec{v} \\ &= (a_\ell\lambda^\ell + \cdots + a_1\lambda + a_0)\vec{v} \\ &= P(\lambda)\vec{v} \end{aligned}$$

which is the result.

Polynomial Invariants of Linear Transformations

In order to describe what we can do, in terms of finding a nice basis, in the case of a linear transformation

$$L : V \to V,$$

it is useful to first take stalk of what we **can not** change, things that can't change, are called **invariants**.

For $L : V \to W$, we could not change the rank!
In this context, $L : V \to V$, we will look for more invariants (though we still can't change the rank).
The extra invariants we shall use will come primarily from polynomials we associate to L.

Associating Polynomials to a Linear Transformation

The key strategy for giving invariants of linear transformations $L : V \to V$ is to associate to any such transformation a pair of polynomials.

That is, given L we will define polynomials:

- $\text{char}_L(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$
- $\min_L(x) = x^m + b_{m-1}x^{m-1} + \cdots + b_0$

These polynomials will be the key invariants of linear transformations. (An invariant just means a quantity which can be associated to L that doesn't depend on any choices and so somehow gives information about L) That is:

- The coefficients: a_i are invariants (a_0 and $a_n - 1$ especially important and well known, though we won't discuss this).
- The coefficients: b_i are invariants.
- The roots of the polynomials are invariants.
- The multiplicities of the roots are invariants.

The key usefulness of these invariants is that because we **can not** change them, they ultimately tell us the limits of what we **can** do.

Suppose $L : V \to V$ is any linear transformation on a finite dimensional vector space V. Pick any basis $\vec{e}_1, \ldots, \vec{e}_n$ for V and let A be the matrix associated to L in this basis.

Lemma

The polynomial

$$\det(x\mathrm{Id}_n - A)$$

does not depend on the choice of basis for V.

This is A4Q4a

Define the **characteristic polynomial** of L to be

$$\mathrm{char}_L(x) = \det(x\mathrm{Id}_n - A)$$

If the polynomial depended on the basis, then it wouldn't be an invariant, but as it does not, it is.

Suppose $L : V \to V$ is any linear transformation on an n-dimensional vector space V.

Lemma

There exists a non-zero polynomials $P(x)$ such that $P(L) = 0_{V,V}$.

Proof The vector space of linear transformations has dimension n^2 (because n by n matricies do) but $\mathrm{Id}_n, L, L^2, \ldots, L^{n^2}$ is a list of $n^2 + 1$ vectors, hence is linearly dependent. This dependence gives

$$b_{n^2} L^{n^2} + b_{n^2-1} L^{n^2-1} + \cdots + b_1 L + b_0 \mathrm{Id}_n = 0$$

and so

$$P(x) = b_{n^2} x^{n^2} + b_{n^2-1} x^{n^2-1} + \cdots + b_1 x + b_0$$

is such a polynomial.

Lemma

There exists a smallest integer ℓ such that there exists a non-zero polynomial $P(x)$ of degree ℓ such that $P(L) = 0_{V,V}$.

Proof-idea This is the well ordering principal.

Lemma

Among polynomials of this smallest degree ℓ there exists a unique non-zero polynomial $P(x)$ such that $P(L) = 0_{V,V}$ and the lead coefficient is 1.

Proof-idea By dividing by the lead coefficient we can ensure the lead is 1,

If $P(x)$ and $Q(x)$ are two such polynomials, then their difference would have degree at most $\ell - 1$, and still have $P(L) - Q(L) = 0$ hence the difference must be the zero-polynomial so the solution is unique.

The **minimal polynomial**, $\min_L(x)$, of a linear transformation $L : V \to V$ is the lowest degree non-zero monic polynomial P such that

$$P(L) : V \to V,$$

is the zero transformation, that is $P(L) = 0_{V,V}$.

Such a thing exists by the previous few lemmas.

The minimal polynomial is primarily a theoretical gadget to prove things.
For complicated matricies we almost never compute it directly.
We will see how to compute it indirectly from the Jordan canonical form later.

Examples

Consider

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

The characteristic polynomial is the determinant of

$$\det\left(\begin{pmatrix} x-1 & -1 \\ 0 & x-1 \end{pmatrix}\right) = (x-1)^2$$

We notice that

$$A^2 - 2A + \mathrm{Id}_2 = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} - 2\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

So that with $P(x) = x^2 - 2x + 1$ we have $P(A) = 0$.

By noticing that with $Q(x) = x - a$ we have

$$Q(A) = \begin{pmatrix} 1-a & 1 \\ 0 & 1-a \end{pmatrix} \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

We conclude that

$$\min_A(x) = x^2 - 2x + 1$$

Theorem
Let $L : V \to V$ be any linear transformation of a finite dimensional vector space. If $P(x)$ is any polynomial so that $P(L) = 0$ then $\min_L(x)|P(x)$.

Proof: By the polynomial division algorithm (long division) we can write:

$$P(x) = Q(x)\min_L(x) + R(x)$$

where the degree of $R(x)$ is strictly less than the degree of $\min_L(x)$.
By definition $\min_L(x)|P(x)$ if and only if $R(x) = 0$.
Now note that:

$$R(L) = P(L) - Q(L)\min_L(L) = 0 - Q(L)0 = 0$$

hence $R(L)$ is a polynomial of strictly lower degree than $\min_L(x)$ such that $\min_L(L) = 0$.
It follows that $R(x) = 0$, and hence $\min_L(x)|P(x)$.

This gives us a useful characterization of the minimal polynomial, and we shall use it to prove things about it.
Fact: $\min_L(x)$ is **the unique** monic polynomial that satisfies this property.
(This is a short exercise in polynomial algebra, we will not use it for anything.)

Factors of the Characteristic Polynomial

Lemma
λ is a root of $\mathrm{char}_L(x)$ if and only if $L - \lambda \mathrm{Id}_V$ is not injective, that is $\mathrm{Ker}(L - \lambda \mathrm{Id}_V) \neq \{\vec{0}\}$.
Equivalently, $x - \lambda$ divides $\mathrm{char}_L(x)$ if and only if $L - \lambda \mathrm{Id}_V$ is not injective.
This is on the assignment.

Lemma
$L - \lambda \mathrm{Id}_V$ is not injective if and only if there exists $\vec{0} \neq \vec{v} \in V$ with $L(\vec{v}) = \lambda \vec{v}$
This is on the assignment.

Lemma
λ is a root of $\mathrm{char}_L(x)$ if and only if there exists $\vec{v} \in V$ with $L(\vec{v}) = \lambda \vec{v}$
This is on the assignment.

You have possibly seen the above before in the context of eigenvalues for matricies.

Factors of the Minimal Polynomial

Lemma

Suppose $P(x)$ divides $\min_L(x)$ then $P(L)$ is not injective.

(This applies in particular when $P(x) = x - \lambda$.)

Proof:

Write $\min_L(x) = P(x)Q(x)$

We note that $Q(L) \neq 0$, because it has strictly lower degree than $\min_L(x)$, and thus there is a vector $\vec{v} \in V$ with $\vec{w} = Q(L)(\vec{v}) \neq \vec{0}$.

We claim $P(L)(\vec{w}) = \vec{0}$. Indeed we know that

$$P(L)(\vec{w}) = P(L)(Q(L)(\vec{v})) = P(L) \circ Q(L)(\vec{v}) = \min_L(L)(\vec{v}) = \vec{0},$$

and so $P(L)$ is not injective.

Lemma

If $P(L)$ is not injective, then $P(x)$ and $\min_L(x)$ have a common factor.
(In particular if $L - \lambda \mathrm{Id}_V$ is not injective then $x - \lambda$ divides $\min_L(x)$.)

Proof:

Assume for the purpose of contradiction that $P(x)$ and $\min_L(x)$ have no common factors then we may write

$$1 = S_1(x)P(x) + S_2(x)\min_L(x).$$

Now because $P(L)$ is not injective, there exists $\vec{0} \neq \vec{v} \in V$ with $P(L)(\vec{v}) = \vec{0}$.
Now we compute:

$$\begin{aligned}
\vec{v} &= \mathrm{Id}_V(\vec{v}) \\
&= (S_1(L)P(L) + S_2(L)\min_L(L))(\vec{v}) && \text{choice of } S_1(x) \text{ and } S_2(x) \\
&= (S_1(L)P(L))(\vec{v}) && \min_L(L)) = 0 \\
&= S_1(L)(P(L)(\vec{v})) && \text{definition of composition} \\
&= S_1(L)(\vec{0}) && P(L)(\vec{v}) = \vec{0} \\
&= \vec{0}
\end{aligned}$$

But this is a contradiction.
In the linear case the only possible common factor is the whole polynomial.

Factors of the Minimal Polynomial vs the Characteristic Polynomial

Theorem
For any linear transformation $L : V \to V$ the roots of $\mathrm{char}_L(x)$ and $\mathrm{min}_L(x)$ are the same. That is, they are precisely the values λ such that $L - \lambda \mathrm{Id}_V$ is not injective.

Proof:
We have just shown that the roots of $\mathrm{min}_L(x)$ are precisely those λ for which $L - \lambda \mathrm{Id}_n$ are not injective.
But we have already seen that these are precisely the roots of $\mathrm{char}_L(x)$.

What sort of relationship do you think exists between $\mathrm{min}_L(x)$ and $\mathrm{char}_L(x)$?

Examples

Consider

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

Both have characteristic polynomial x^2. The first has minimal polynomial x^2, the second has minimal polynomial x.

Theorem (Very special case of a later result)
If the minimal and characteristic polynomial of a matrix are both x^2 then there exists a basis so that the matrix for L is

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

If the minimal polynomial of a matrix is x, then the matrix is the zero matrix.

Idea: Let $\vec{v}$ be any vector such that $L(\vec{v}) \neq \vec{0}$.

Then $\vec{e_1} = L(\vec{v}), \vec{e_2} = \vec{v}$ is a basis with respect to which the matrix is $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$

Theorem(we will not prove, and you don't need)
For 2 by 2 and 3 by 3 matricies the characteristic and minimal polynomials determine if two matricies are similar. So A and B are similar if and only if both the characteristic/minimal polynomials are identical.

Examples

Consider

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

Both have characteristic polynomial x^4 and minimal polynomial x^2, but they are not similar.

Why?

Natural Questions About Characteristic and Minimal Polynomials

- Given some description of a linear transformation $L : \mathbb{R}^n \to \mathbb{R}^n$, find its minimal and characteristic polynomials.
  For the characteristic polynomial you must know the definition, for the minimal polynomial we will come back to this.
- Given some description of a linear transformation $L : V \to V$, find its minimal and characteristic polynomials.
  Translate this to a question about matricies!
- What information do the coefficients contain? are there formulas for these? what sorts of properties do they have?
  These questions are open ended, we will mostly ignore these.
- What information do the roots contains? are there formulas for these? what sorts of properties do they have?
  These questions are open ended, we will come back to this.