## 11. Powers of matrices

Consider the sequence

$$f_0 = 0, 1, 1, 2, 3, 5, 8, 13, \cdots, f_n, \cdots.$$

This sequence satisfies the recurrence

$$f_n = f_{n-2} + f_{n-1}.$$

It is called the **Fibonacci sequence**. As a motivating question, what is the $n$th term? That is, can we find a closed form expression for $f_n$?

Here is a seemingly unrelated problem. Consider the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

What is $A^{100}$? Even computing small powers of $A$ looks like a pain. A much easier problem is to compute powers of

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

It is easy to see that

$$D^n = \begin{pmatrix} 1 & 0 \\ 0 & 2^n \end{pmatrix}.$$

The idea is to reduce computing powers of $A$ to powers of a diagonal matrix, which is easy.

To see how to do this, let us go back to the problem of computing the $n$th term $f_n$ of the Fibonnaci sequence. To compute the $n$th term, we need the previous two terms. This suggests we should create a vector

$$v_n = \begin{pmatrix} f_{n-1} \\ f_n \end{pmatrix}.$$

We then have

$$v_{n+1} = \begin{pmatrix} f_n \\ f_{n+1} \end{pmatrix} = \begin{pmatrix} f_n \\ f_{n-1} + f_n \end{pmatrix}.$$

The key point is that the last vector is just $Av_n$, where

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

In other words $v_n = A^{n-1}v_1$, where

$$v_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Now if we have a diagonal matrix and we apply it to a vector, what happens? If we apply the the diagonal matrix

$$D = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

to $v_1$, we get

$$\begin{pmatrix} 1 \\ \frac{1}{2} \end{pmatrix}.$$

In general we have

$$D^n v_1 = \begin{pmatrix} 1 \\ \frac{1}{2^n} \end{pmatrix}.$$

The key point is that if $n$ is large, then $1/2^n$ is negligible in comparison with 1, so that $D^n v_1$ is very close to

$$e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Note that $De_1 = e_1$. On the other hand

$$De_2 = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} = e_2/2.$$

In fact if $D$ is a diagonal matrix, with entries $\lambda_1, \lambda_2, \dots, \lambda_n$ on the main diagonal, then we have $De_i = \lambda_i e_i$. This motivates:

**Definition 11.1.** *Let $A \in M_{n,n}(F)$. We say that $v \neq 0$ is an **eigenvector** with **eigenvalue** $\lambda$ if $Av = \lambda v$.*

So, a diagonal matrix $D$, with diagonal entries $\lambda_1, \lambda_2, \dots, \lambda_n$, has eigenvectors $e_1, e_2, \dots, e_n$, with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$. Note that the eigenvectors are a basis for $F^n$.

If $P$ is an invertible matrix then the inverse is unique. We denote the inverse by $P^{-1}$.

**Definition 11.2.** *Let $A$ and $B$ be two square $n \times n$ matrices. We say that $A$ and $B$ are **similar**, denoted $A \sim B$, if there is an invertible square $n \times n$ matrix $P$ such that $A = PBP^{-1}$.*

*We say that $A$ is **diagonalisable** if $A$ is similar to a diagonal matrix $D$.*

**Lemma 11.3.** *Suppose that $A$ and $B$ are two $n \times n$ square matrices and that $P$ is an invertible matrix such that*

$$A = PBP^{-1}.$$

*Then*

$$A^n = PB^nP^{-1}.$$

*Proof.* We prove this by induction on $n$. It is true for $n = 1$ by assumption. Suppose that

$$A^n = PB^nP^{-1},$$

for some $n > 0$. Then

$$\begin{aligned} A^{n+1} &= A \cdot A^n \\ &= (PBP^{-1})(PB^nP^{-1}) \\ &= P(P^{-1}P)(BB^n)P^{-1}) \\ &= PB^{n+1}P^{-1}, \end{aligned}$$

as required. Thus the result holds by induction on $n$. □

In other words, if $A$ is diagonalisable, then we can compute its powers very quickly.

**Lemma 11.4.** *Suppose that $A$ is an $n \times n$ matrix and that $D$ is a diagonal matrix with entries $\lambda_1, \lambda_2, \dots, \lambda_n$. Suppose that $P$ is an invertible matrix and that $A = PDP^{-1}$. Let $v_i = Pe_i$.*

*Then the vectors $v_1, v_2, \dots, v_n$ are eigenvectors with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ and they form a basis for $F^n$.*

*Proof.*

$$\begin{aligned} Av_i &= (PDP^{-1})(Pe_i) \\ &= (PD)(P^{-1}P)e_i \\ &= P(De_i) \\ &= P(\lambda_i e_i) \\ &= \lambda_i(Pe_i) = \lambda_i v_i. \end{aligned}$$

Therefore $v_i$ is an eigenvector with eigenvalue $\lambda_i$.

Finally we want to prove that $v_1, v_2, \dots, v_n$ are a basis for $F^n$. Suppose that there are scalars $r_1, r_2, \dots, r_n$ such that

$$0 = r_1v_1 + r_2v_2 + \dots + r_nv_n.$$

Multiply both sides by $P$ to get

$$\begin{aligned} 0 &= P \cdot 0 \\ &= P(r_1v_1 + r_2v_2 + \dots + r_nv_n) \\ &= P(r_1v_1) + P(r_2v_2) + \dots + P(r_nv_n) \\ &= r_1Pv_1 + r_2Pv_2 + \dots + r_nPv_n \\ &= r_1e_1 + r_2e_2 + \dots + r_ne_n. \end{aligned}$$

Here we used the fact that $Pv_i = P(P^{-1}e_i) = e_i$. As $e_1, e_2, \dots, e_n$ are a basis they are independent. Therefore $r_1, r_2, \dots, r_n = 0$. But then $v_1, v_2, \dots, v_n$ are independent. Any independent set of vectors can be extended to a basis. Since any two bases have the same size, it follows that $v_1, v_2, \dots, v_n$ are a basis to begin with. $\square$

What is the matrix $P$? When applied to $e_i$ we get $v_i$. In fact this means the columns of $P$ are the vectors $v_1, v_2, \dots, v_n$.

**Theorem 11.5.** *Let $A \in M_{n,n}(F)$.*

*Then $A$ is diagonalisable if and only if we can find a basis $v_1, v_2, \dots, v_n$ of eigenvectors for $F^n$. In this case,*

$$A = PDP^{-1},$$

*where $P$ is the matrix whose columns are the eigenvectors $v_1, v_2, \dots, v_n$ and $D$ is the diagonal matrix whose diagonal entries are the corresponding eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$.*

*Proof.* We have already seen one direction. By (11.4), if $A = PDP^{-1}$ where $D$ is a diagonal matrix with entries $\lambda_1, \lambda_2, \dots, \lambda_n$ and $P$ is invertible then the vectors $v_1, v_2, \dots, v_n$ are a basis of eigenvectors with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$.

So suppose that $v_1, v_2, \dots, v_n$ are a basis of eigenvectors. Let $P$ be the matrix whose columns are the vectors $v_1, v_2, \dots, v_n$. Since the vectors $v_1, v_2, \dots, v_n$ are independent, the kernel of $P$ is the trivial subspace $\{0\}$. But then $P$ is an invertible matrix. Let $D = P^{-1}AP$. Then

$$\begin{aligned} De_i &= (P^{-1}AP)e_i \\ &= P^{-1}Av_i \\ &= P^{-1}\lambda_i v_i \\ &= \lambda_i P^{-1} v_i \\ &= \lambda_i e_i. \end{aligned}$$

So $D$ is the matrix whose $i$th row is the vector $\lambda_i e_i$. But then $D$ is a diagonal matrix with entries $\lambda_1, \lambda_2, \dots, \lambda_n$ on the main diagonal. We have

$$D = P^{-1}AP.$$

Multiplying both sides by $P$ on the left, we get

$$PD = AP.$$

Finally multiplying both sides on the right by $P^{-1}$ we get

$$A = PDP^{-1}.$$

$\square$

Here is one good reason why $A$ might have a basis of eigenvectors:

**Theorem 11.6.** *Let $A \in M_{n,n}(F)$ and let $v_1, v_2, \dots, v_k$ be eigenvectors of $A$ with distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_k$.*

*Then $v_1, v_2, \dots, v_k$ are independent. In particular if $k = n$ then $v_1, v_2, \dots, v_n$ are a basis of eigenvetors for $F^n$ and $A$ is diagonalisable.*

*Proof.* Suppose not. Suppose that $v_1, v_2, \dots, v_k$ are dependent. We will derive a contradiction. By assumption there are scalars $r_1, r_2, \dots, r_k$, not all zero, such that

$$0 = r_1 v_1 + r_2 v_2 + \dots + r_k v_k.$$

We suppose that $k$ is minimal with this property. In particular we may assume that $r_i \neq 0$ for all $i$. Clearly $k > 1$. We apply $A$ to both sides of the equation above. We get

$$\begin{aligned} 0 &= A \cdot 0 \\ &= A(r_1 v_1 + r_2 v_2 + \dots + r_k v_k) \\ &= r_1 A v_1 + r_2 A v_2 + \dots + r_k A v_k \\ &= r_1 \lambda_1 v_1 + r_2 \lambda_2 v_2 + \dots + r_k \lambda_k v_k. \end{aligned}$$

Take the first equation and multiply by $\lambda_k$. We get

$$\begin{aligned} 0 &= r_1 \lambda_k v_1 + r_2 \lambda_k v_2 + \dots + r_k \lambda_k v_k \\ 0 &= r_1 \lambda_1 v_1 + r_2 \lambda_2 v_2 + \dots + r_k \lambda_k v_k. \end{aligned}$$

We subtract the second equation from the first equation:

$$0 = r_1(\lambda_k - \lambda_1) v_1 + r_2(\lambda_k - \lambda_2) v_2 + \dots + r_{k-1}(\lambda_k - \lambda_{k-1}) v_{k-1}.$$

Now $s_i = r_i(\lambda_k - \lambda_i) \neq 0$, since the eigenvalues are distinct. But then we found a linear dependence involving fewer eigenvectors. This contradicts our choice of $k$. The only possibility is that the eigenvectors are independent to start with. $\square$

So now let us turn to the problem of determining the eigenvectors and eigenvalues of a matrix $A$.

**Definition 11.7.** *Let $A \in M_{n,n}(F)$ and let $\lambda \in F$. Let*

$$E_\lambda(A) = \{\, v \in V \mid Av = \lambda v \,\} \subset F^n.$$

*$E_\lambda(A)$ is called an* ***eigenspace*** *of $A$.*

**Lemma 11.8.** *Let $A \in M_{n,n}(F)$ and let $\lambda \in F$.*

*Then $V = E_\lambda(A)$ is a subspace of $F^n$.*

*Proof.* There are two ways to proceed. In the first way we show that $V$ is non-empty and closed under addition and scalar multiplication.

$$A \cdot 0 = 0 = \lambda 0.$$

So $0 \in V$. Suppose that $v$ and $w \in V$. Then $Av = \lambda v$ and $Aw = \lambda w$. Then

$$A(v + w) = Av + Aw = \lambda v + \lambda w = \lambda(v + w).$$

Hence $v + w \in V$ and so $V$ is closed under addition. Suppose that $v \in V$ and $r$ is a scalar. Then

$$A(rv) = r(Av) = r\lambda v = \lambda(rv).$$

Hence $rv \in V$ and so $V$ is closed under scalar multiplication. Therefore $V$ is a subspace of $F^n$.

Here is another way to proceed:

**Claim 11.9.**

$$E_\lambda(A) = \mathrm{Ker}(A - \lambda I_n).$$

*Proof of* (11.9). Suppose that $v \in E_\lambda(A)$. Then $Av = \lambda v$. But then

$$(A - \lambda I_n)v = Av - \lambda I_n v = Av - \lambda v = 0.$$

Therefore $v \in \mathrm{Ker}(A - \lambda)$ and so $E_\lambda(A) \subset \mathrm{Ker}(A - \lambda I)$. The revere inclusion is just as easy to prove and this establishes the claim. □

Since the kernel is always a subspace, (11.9) implies that $E_\lambda(A)$ is a subspace. □

So what is a quick way to determine if a square matrix has a non-trivial kernel? This is the same as saying the matrix is not invertible. Now for $2 \times 2$ matrices we have seen a quick way to determine if the matrix is invertible. If

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then $B$ is invertible if and only if $ad - bc \neq 0$. For us

$$B = A - \lambda I_2 = \begin{pmatrix} -\lambda & 1 \\ 1 & 1 - \lambda \end{pmatrix}.$$

This is not invertible if and only if

$$-\lambda(1 - \lambda) - 1 = 0.$$

This is a quadratic polynomial in $\lambda$, which is known as the **characteristic polynomial**. Expanding, we get

$$\lambda^2 - \lambda - 1 = 0.$$

Using the quadratic formula gives

$$\lambda = \frac{1 \pm \sqrt{5}}{2}.$$

Note that the Golden ratio turns up as one of the roots. If we plug in $\lambda_1 = (1+\sqrt{5})/2$ then we get

$$B = \begin{pmatrix} -\frac{1+\sqrt{5}}{2} & 1 \\ 1 & \frac{1-\sqrt{5}}{2} \end{pmatrix}.$$

If we multiply the first row by $-(1-\sqrt{5})/2$ and add it to the second row we get

$$\begin{pmatrix} -\frac{1+\sqrt{5}}{2} & 1 \\ 0 & 0 \end{pmatrix},$$

so that this is indeed a matrix of rank one. The kernel is spanned by

$$v_1 = (1, \frac{1+\sqrt{5}}{2}).$$

This is an eigenvector with eigenvalue $\lambda_1$. Similarly

$$v_2 = (1, \frac{1-\sqrt{5}}{2}).$$

is an eigenvector with eigenvalue

$$\lambda_2 = \frac{1-\sqrt{5}}{2}.$$

Thus $A = PDP^{-1}$, where

$$D = \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}$$

and

$$P = \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix}.$$

It follows that

$$P^{-1} = -\frac{1}{\sqrt{5}} \begin{pmatrix} \frac{1-\sqrt{5}}{2} & -1 \\ -\frac{1+\sqrt{5}}{2} & 1 \end{pmatrix}$$

One can check the equality $A = PDP^{-1}$. Now

$$
\begin{aligned}
A^n v_1 &= PD^nP^{-1}v_1 \\
&= \frac{1}{-\sqrt{5}} \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^n \end{pmatrix} \begin{pmatrix} \frac{1-\sqrt{5}}{2} & -1 \\ -\frac{1+\sqrt{5}}{2} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^n \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n \\ -\left(\frac{1-\sqrt{5}}{2}\right)^n \end{pmatrix} \\
&= \frac{1}{\sqrt{5}} \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n \\ ? \end{pmatrix}.
\end{aligned}
$$

It follows that

$$f_n = \frac{1}{\sqrt{5}} \left( \left( \frac{1+\sqrt{5}}{2} \right)^n - \left( \frac{1-\sqrt{5}}{2} \right)^n \right).$$

Now

$$-1 < \frac{1-\sqrt{5}}{2} < 0 \qquad \text{whilst} \qquad \frac{1+\sqrt{5}}{2} > 1.$$

If $n$ is large this means

$$\left( \frac{1-\sqrt{5}}{2} \right)^n \approx 0.$$

and the other term is the one that matters. But $f_n$ is an integer. It follows that $f_n$ is the closest integer to

$$\left( \frac{1}{\sqrt{5}} \right) \left( \frac{1+\sqrt{5}}{2} \right)^n.$$

It is interesting to check this for some values of $n$. Put in $n = 5$ and we get

$$4.956,$$

which is very close to the real answer, namely 5. Put in $n = 6$ and we get

$$8.025,$$

which is even closer to the real answer, namely 8. Put in $n = 100$ (well into maxima, or your favourite computer algebra system) we get

$$3.542248 \times 10^{20}.$$

Actually this is nowhere near the real answer. On the other hand maxima (or YFCAS) has a function to compute $f_{100}$ directly (and more importantly correctly). The problem is as follows. To compute $f_{100}$ accurately using matrices, which involves real numbers, we need twenty significant figures of accuracy. Maxima, let's say, routinely uses ten significant figures of accuracy, so on the first ten digits are correct. On the other hand, the routine which maxima uses to compute the Fibonacci numbers, does the stupid thing and just keeps computing each term in the sequence until it gets to a hundred. The advantage of this is that the computer knows exactly how much accuracy it needs as it computes; if it has an integer like 1450 it needs four significant figures but if it has a number like 123456 it needs six, and so on.