

4.1 Eigenvectors, Eigenvalues, and Spectrum

EX Recall our new standard example of $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by reflecting over a line. If $S = \{\vec{e}_1, \vec{e}_2\}$, then usually $[T]_S^S$ is hard to compute, but if $B = \{\vec{b}_1, \vec{b}_2\}$ is given by \vec{b}_1 parallel to the line and \vec{b}_2 perpendicular to it then

$$[T]_B^B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We know $[T]_S^S$ and $[T]_B^B$ are similar, we can use the homework to see

$$\text{Tr}([T]_S^S) = \text{Tr}([T]_B^B) = 0$$

and

$$\det([T]_S^S) = \det([T]_B^B) = (1) \cdot (-1) = -1.$$

Also, since

$$([T]_B^B)^d = \begin{cases} [T]_B^B & \text{if } d \text{ odd} \\ I_2 & \text{if } d \text{ even} \end{cases}$$

we see

$$\underbrace{T \circ T \circ \dots \circ T}_d = \begin{cases} T & \text{if } d \text{ odd} \\ I & \text{if } d \text{ even.} \end{cases}$$

All of this is to say that finding the basis B on which T acted simply, allows us do computations for T much more easily. □

- In this section, given a lin trans. $T: V \rightarrow V$ we will look for vectors \vec{v} on which T acts as simply as possible; namely, $T(\vec{v}) = \lambda \vec{v}$ for some scalar λ .

Def Let $T: V \rightarrow V$ be a linear transformation. A non-zero vector $\vec{v} \in V$ is called an eigenvector of T if

$$T(\vec{v}) = \lambda \vec{v}$$

for some scalar λ . We call λ an eigenvalue of T . The set of all eigenvalues of T is called the spectrum of T , and is denoted $\sigma(T)$.

- Observe that if \vec{v} is an eigenvector of T with eigenvalue λ iff $\vec{v} \neq \vec{0}$ and $(T - \lambda I)(\vec{v}) = \vec{0}$.

That is, iff $\vec{v} \in \text{Ker}(T - \lambda I) \setminus \{\vec{0}\}$.

Def If λ is an eigenvalue of T , we call $\text{Ker}(T - \lambda I)$ the eigen space of λ for T .

Ex Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be reflection over the line $y = 2x$. Then

$$T\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right) = 1 \cdot \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } T\left(\begin{pmatrix} -2 \\ 1 \end{pmatrix}\right) = -1 \cdot \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

So $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 1 \end{pmatrix}$ are eigenvectors of T with eigenvalues 1 and -1, respectively. It turns out these are

all of the eigenvalues of T , so $\sigma(T) = \{-1, 1\}$. Moreover, one can show

$$\ker(T - 1 \cdot I) = \text{span} \left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$$

$$\text{and } \ker(T - (-1) \cdot I) = \text{span} \left\{ \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\}.$$

Finding Eigenvectors and Eigenvalues

• Suppose you knew λ is an eigenvalue of a lin trans. T . Then computing $\ker(T - \lambda I)$ yields all the eigenvectors of T with eigenvalue λ . If you do this for all λ in the spectrum of T , you'll find all the eigenvectors of T . But how do you find the eigenvalues of T in the first place?

• Consider the special case $A: \mathbb{F}^n \rightarrow \mathbb{F}^n$ with $A \in M_{n \times n}$.

Prop For $A \in M_{n \times n}$, λ is an eigenvalue of A if and only if $\det(A - \lambda I_n) = 0$.

Proof We have

λ is an eigenvalue of $A \iff \ker(A - \lambda I_n)$ contains a non-zero vector

$\iff (A - \lambda I_n)\vec{x} = \vec{0}$ does not have a unique solution

com puts $\iff A - \lambda I_n$ is not invertible

Section 3.3 $\iff \det(A - \lambda I_n) = 0$.

Def Let z be a variable. Then for $A \in M_{n \times n}$

$$\text{char}_A(z) := \det(A - zI_n)$$

is a degree n polynomial in z called the characteristic polynomial of A

• The previous proposition says λ is an eigenvalue of A iff it is a root of its characteristic polynomial.

10/31

Ex (1) Let $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Then

$$\text{char}_A(z) = \det(A - zI) = \begin{vmatrix} 1-z & 0 \\ 0 & -1-z \end{vmatrix} = (1-z)(-1-z) = (z-1)(z+1)$$

The roots of $\text{char}_A(z)$ are $z = \pm 1$. Thus $\sigma(A) = \{-1, 1\}$.

(2) Let $A = \begin{pmatrix} a_{1,1} & & * \\ & a_{2,2} & \\ 0 & & \ddots \\ & & & a_{n,n} \end{pmatrix}$ be upper triangular.

$$\text{char}_A(\lambda) = \det(A - \lambda I) = \det \begin{pmatrix} a_{1,1} - \lambda & & * \\ & a_{2,2} - \lambda & \\ 0 & & \ddots \\ & & & a_{n,n} - \lambda \end{pmatrix} = (a_{1,1} - \lambda)(a_{2,2} - \lambda) \cdots (a_{n,n} - \lambda)$$

upper triangular

which has roots $z = a_{1,1}, a_{2,2}, \dots, a_{n,n}$. Thus $\sigma(A) = \{a_{1,1}, a_{2,2}, \dots, a_{n,n}\}$.

Let's record the last example as a proposition. Note a similar argument works for lower triangular.

Prop If $A = (a_{ij}) \in M_{\mathbb{F}}(n \times n)$ is triangular, then $\sigma(A) = \{a_{11}, a_{22}, \dots, a_{nn}\}$.

Lemma Suppose $A, B \in M_{\mathbb{F}}(n \times n)$ are similar, then $\text{char}_A(z) = \text{char}_B(z)$.

Proof Since A and B are similar, there exists invertible Q st.

$$A = Q^{-1} B Q.$$

Note that

$$Q^{-1}(B - \lambda I_n)Q = Q^{-1} B Q - \lambda Q^{-1} Q = A - \lambda I_n.$$

Thus $A - \lambda I_n$ and $B - \lambda I_n$ are similar for scalars λ . By Exercise 5 on Homework 8, they have the same determinant. Hence for all scalars λ

$$\text{char}_A(\lambda) = \det(A - \lambda I_n) = \det(B - \lambda I_n) = \text{char}_B(\lambda). \quad \square$$

Let us return to the abstract setting $T: V \rightarrow V$, but assume V is finite-dim'l. By taking a basis $B = \{\vec{v}_1, \dots, \vec{v}_n\}$ for V , we can always return to our concrete setting:

$$\begin{array}{ccc} V & \xrightarrow{T} & V \\ \downarrow \text{[C]}_B & & \downarrow \text{[C]}_B \\ \mathbb{F}^n & \xrightarrow{[T]_B^B} & \mathbb{F}^n \end{array}$$

Suppose λ is an eigenvalue of $[T]_B^B$ with eigenvector $\vec{x} = (x_1, \dots, x_n)^T \in \mathbb{F}^n$. Consider $\vec{v} := x_1 \vec{v}_1 + \dots + x_n \vec{v}_n \in V$.

Then \vec{v} is an eigenvector of T . Indeed,

$$[T(\vec{v})]_B = [T]_B^B [\vec{v}]_B = [T]_B^B \vec{x} = \lambda \vec{x} = \lambda [\vec{v}]_B = [\lambda \vec{v}]_B.$$

Since $T(\vec{v})$ and $\lambda \vec{v}$ have same coordinate vector, it must be that $T(\vec{v}) = \lambda \vec{v}$.

In particular, λ is also an eigenvalue of T .

But what if we choose another basis A ? Will we still get the same eigenvalues and eigenvectors? Yes, because the two matrix reps are similar:

$$[T]_A^A = [I]_B^A [T]_B^B [I]_A^B = ([I]_A^B)^{-1} [T]_B^B [I]_A^B$$

so by the previous lemma

$$\text{char}_{[T]_A^A}(z) = \text{char}_{[T]_B^B}(z).$$

In light of this, we make the following definition:

Def Let V be a finite-dimensional vector space and let $T: V \rightarrow V$ be a linear transformation.

Then the characteristic polynomial of T is

$$\text{char}_T(z) = \det([T]_B^B - z I)$$

where B is any basis for V .

Ex Let $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be the reflection over the line $y=2x$. Then for the two bases
 $S = \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ and $B = \left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right\}$

we have seen that

$$[T]_S^S = \begin{pmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{pmatrix} \quad \text{and} \quad [T]_B^B = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

We've already seen the characteristic polynomial of the latter matrix is $(z-1)(z+1)$, so $\text{char}_T(z) = (z-1)(z+1)$. We should get the same thing for the other matrix:

$$\begin{aligned} \det([T]_S^S - zI_2) &= \det \begin{pmatrix} -3/5 - z & 4/5 \\ 4/5 & 3/5 - z \end{pmatrix} = \left(-\frac{3}{5} - z\right)\left(\frac{3}{5} - z\right) - \frac{16}{25} \\ &= -\frac{1}{25} - z^2 - \frac{16}{25} \\ &= z^2 - 1 = (z-1)(z+1) \end{aligned}$$

Real vs Complex Roots

Fact (Fundamental Theorem of Algebra) Every non-constant polynomial has at least one complex root.

• So for any non-constant $p(z) \in \mathbb{P}_d(\mathbb{C})$, there exists $\lambda \in \mathbb{C}$ s.t. $p(\lambda) = 0$.

Warning This is not true over \mathbb{R} . That is, if $p(x) \in \mathbb{P}_d(\mathbb{R})$ (i.e. has real coefficients) it may be that $p(\lambda) \neq 0 \quad \forall \lambda \in \mathbb{R}$. For example $p(x) = x^2 + 1$ has no real roots, but

$$p(x) = (x-i)(x+i)$$

so it does have complex ones. We can always treat $p(x)$ as a complex polynomial (that is $\mathbb{P}_d(\mathbb{R}) \subseteq \mathbb{P}_d(\mathbb{C})$), so in order to use the Fundamental Theorem of Algebra, we will usually treat all polynomials as complex ones.

Multiplicities of Eigenvalues

• Suppose $p(z) \in \mathbb{P}_d(\mathbb{C})$ has a root $\lambda \in \mathbb{C}$. Recall that this implies

$$p(z) = (z-\lambda)q(z)$$

for some $q(z) \in \mathbb{P}_d(\mathbb{C})$. We say $z-\lambda$ divides $p(z)$. If λ is also a root of $q(z)$, then

$$p(z) = (z-\lambda)^2 r(z)$$

for some $r(z) \in \mathbb{P}_{d-2}(\mathbb{C})$. Thus $(z-\lambda)^2$ divides $p(z)$.

Def Let λ be an eigenvalue of a linear transformation T . The algebraic multiplicity of λ is the largest integer k s.t. $(z-\lambda)^k$ divides $\text{char}_T(z)$. We denote $m_\lambda(T) = k$.
 The geometric multiplicity of λ is $\dim(\text{Ker}(T - \lambda I))$.

• Let $\lambda_1, \dots, \lambda_d$ be the distinct eigenvalues of T . It follows that

$$\text{char}_T(z) = c (z-\lambda_1)^{m_{\lambda_1}(T)} (z-\lambda_2)^{m_{\lambda_2}(T)} \dots (z-\lambda_d)^{m_{\lambda_d}(T)}$$

for some $c \in \mathbb{C}$. Observe that we then have that the degree of $\text{char}_T(z)$ is $m_{\lambda_1}(T) + m_{\lambda_2}(T) + \dots + m_{\lambda_d}(T)$.

On the other hand,

$$\text{char}_T(z) = \det([T]_B^B - zI)$$

where B is a basis for V and so the degree is also $\dim(V)$. Thus we obtain the following:

Prop Let $\lambda_1, \dots, \lambda_d$ be the distinct eigenvalues of $T: V \rightarrow V$. Then $\sum_{i=1}^d m_{\lambda_i}(T) = \dim(V)$.

• Given $A \in M_{n \times n}$, if we say $\lambda_1, \lambda_2, \dots, \lambda_n$ are its eigenvalues "counting multiplicities", we mean each distinct eigenvalue λ is repeated $m_\lambda(A)$ times in this list.

Prop Let λ be an eigenvalue of T . Then $\dim(\ker(T - \lambda I)) \leq m_\lambda(T)$. 11/5

Proof Homework 9. □

• The inequality in this proposition can be strict, as the next example shows:

Ex Define $T: P_2 \rightarrow P_2$ by $T(p(x)) = p'(x)$. Recall that for $S = \{1, x, x^2\}$

$$[T]_S^S = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

< 0

$$\text{char}_T(z) = \det([T]_S^S - zI_3) = \det \begin{pmatrix} -z & 1 & 0 \\ 0 & -z & 2 \\ 0 & 0 & -z \end{pmatrix} = -z^3$$

← upper triangular

Thus $\sigma(T) = \{0\}$ and $m_0(T) = 3$. On the other hand,

$$\ker(T - 0 \cdot I) = \ker(T) = \{\text{constant polynomials}\} = \text{span}\{1\}.$$

Thus $\dim(\ker(T - 0 \cdot I)) = 1 < 3 = m_0(T)$. □

The trace, the determinant, and the eigenvalues

• We conclude this section by relating the trace and determinant of a matrix to its eigenvalues.

Thm Let $A \in M_{n \times n}$ have eigenvalues $\lambda_1, \dots, \lambda_n$ counting multiplicities. Then

① $\text{Tr}(A) = \lambda_1 + \dots + \lambda_n$

② $\det(A) = \lambda_1 \cdots \lambda_n$

Proof Homework 9.

4.2 Diagonalization

Def Let V be a finite-dimensional vector space. A linear transformation $T: V \rightarrow V$ is diagonalizable if there exists a basis B for V s.t. $[T]_B^B$ is diagonal.

• We've seen that any reflection $T: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is diagonal.

Thm $T: V \rightarrow V$ is diagonalizable if and only if there exists a basis for V consisting of eigenvectors.

Proof (\Rightarrow) Suppose T is diagonalizable with basis $B = \{\vec{v}_1, \dots, \vec{v}_n\}$ for V and

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

Then for each $i=1, \dots, n$

$$[T(\vec{v}_i)]_B = [T]_B^B [\vec{v}_i]_B = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} \cdot \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow \text{ith position} = \begin{pmatrix} 0 \\ \vdots \\ \lambda_i \\ \vdots \\ 0 \end{pmatrix} = [\lambda_i \vec{v}_i]_B$$

Since coordinate vectors are unique, we must have $T(\vec{v}_i) = \lambda_i \vec{v}_i$. So B consists of eigenvectors.

(\Leftarrow) Let $B = \{\vec{v}_1, \dots, \vec{v}_n\}$ be a basis for V of eigenvectors for T with eigenvalues $\lambda_1, \dots, \lambda_n$ respectively. Then the j th column of $[T]_B^B$ is:

$$[T(\vec{v}_j)]_B = [\lambda_j \vec{v}_j]_B = \begin{pmatrix} 0 \\ \vdots \\ \lambda_j \\ \vdots \\ 0 \end{pmatrix} \leftarrow \text{jth position}$$

Hence $[T]_B^B$ is diagonal and T is diagonalizable. \square

Ex $T: \mathbb{P}_2 \rightarrow \mathbb{P}_2$ defined by $T(p(x)) = p'(x)$ is not diagonalizable. Recall that $\sigma(T) = \{0\}$. So every eigenvector of T belongs to $\ker(T - 0 \cdot I) \setminus \{0\}$. So any basis of eigenvectors would need to be contained in the subspace $\ker(T - 0 \cdot I)$. But we already saw that this subspace is 1-dimensional, so it cannot contain a basis for \mathbb{P}_2 . \square

Prop $A \in M_{n \times n}$ is diagonalizable if and only if it is similar to a diagonal matrix.

Proof (\Rightarrow) Suppose A is diagonalizable for a basis B of \mathbb{F}^n . Then for the standard basis S we have $A = [A]_S^S$, which is similar to $[A]_B^B$.

(\Leftarrow) Suppose $A = QDQ^{-1}$ for $D, Q \in M_{n \times n}$ with $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ and Q invertible. Since Q is invertible and square, its columns form a basis $B = \{\vec{v}_1, \dots, \vec{v}_n\}$ for \mathbb{F}^n . Observe that

$$A\vec{v}_j = A(Q\vec{e}_j) = (QDQ^{-1})(Q\vec{e}_j) = Q(D\vec{e}_j) = Q(\lambda_j \vec{e}_j) = \lambda_j Q\vec{e}_j = \lambda_j \vec{v}_j$$

Thus \vec{v}_j is an eigenvector of A (with eigenvalue λ_j). Since this holds for each $j=1, \dots, n$ and $\vec{v}_1, \dots, \vec{v}_n$ forms a basis for \mathbb{F}^n , the previous theorem implies A is diagonalizable. \square

Rem Note that $Q[\vec{v}_j]_B = Q\vec{e}_j = \vec{v}_j = [\vec{v}_j]_S$. This implies $Q = [I]_B^S$. Also $D = Q^{-1}AQ = [I]_S^B [A]_S^S [I]_B^S = [A]_B^B$.

Def For $A \in M_{n \times n}$ diagonalizable, we call a factorization $A = QDQ^{-1}$ a diagonalization of A when $D \in M_{n \times n}$ is diagonal.

The previous theorem says that the diagonalization of A is given by $Q = [I]_{\mathcal{B}}$ where \mathcal{B} is a basis of eigenvectors of A and $D = [A]_{\mathcal{B}}^{\mathcal{B}} = \text{diag}(\lambda_1, \dots, \lambda_n)$ where $\lambda_1, \dots, \lambda_n$ are the corresponding eigenvalues.

Ex Let $A = \begin{pmatrix} -2 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{pmatrix}$. Since A is triangular, $\sigma(A) = \{-2, 1, 3\}$ and as Quiz 9 you showed it had eigenvalues:

$$\begin{aligned} \lambda_1 &= -2 & \vec{v}_1 &= \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \\ \lambda_2 &= 1 & \vec{v}_2 &= \begin{pmatrix} 1/3 \\ 1 \\ 0 \end{pmatrix} \\ \lambda_3 &= 3 & \vec{v}_3 &= \begin{pmatrix} 2/5 \\ 1 \\ 1 \end{pmatrix} \end{aligned}$$

So if we define

$$D := \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} \quad \text{and} \quad Q := \begin{pmatrix} 1 & 1/3 & 2/5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

note Q is invertible:

$$\det(Q) = \underbrace{1 \cdot 1 \cdot 1}_{\text{diagonal entries}} = 1 \neq 0.$$

then we should have

$$A = QDQ^{-1}$$

This is equivalent to $AQ = QD$ (which does not require us to compute Q^{-1}), so we'll check that:

$$AQ = \begin{pmatrix} -2 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1/3 & 2/5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} -2 & 1/3 & 6/5 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix}$$

$$QD = \begin{pmatrix} 1 & 1/3 & 2/5 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix} = \begin{pmatrix} -2 & 1/3 & 6/5 \\ 0 & 1 & 3 \\ 0 & 0 & 3 \end{pmatrix} = \checkmark$$

Functional Calculus

Let $A \in M_{n \times n}$ have diagonalization $A = QDQ^{-1}$ with $D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Observe that

$$A^2 = (QDQ^{-1})(QDQ^{-1}) = QD^2Q^{-1} = Q \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} Q^{-1} = Q \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_n^2 \end{pmatrix} Q^{-1}$$

Iterating this argument we have

$$A^k = QD^kQ^{-1} = Q \begin{pmatrix} \lambda_1^k & & \\ & \ddots & \\ & & \lambda_n^k \end{pmatrix} Q^{-1}$$

Now, let $p(t) = a_d t^d + a_{d-1} t^{d-1} + \dots + a_1 t + a_0$ be a degree d polynomial. Then using the above we have:

$$\begin{aligned} p(A) &= a_d A^d + \dots + a_1 A + a_0 I_n = a_d QD^dQ^{-1} + \dots + a_1 QDQ^{-1} + a_0 QIQ^{-1} \\ &= Q [a_d D^d + \dots + a_1 D + a_0 I] Q^{-1} = Q p(D) Q^{-1} \\ &= Q \left[a_d \begin{pmatrix} \lambda_1^d & & \\ & \ddots & \\ & & \lambda_n^d \end{pmatrix} + \dots + a_1 \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix} + a_0 \begin{pmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{pmatrix} \right] Q^{-1} \\ &= Q \begin{pmatrix} p(\lambda_1) & & \\ & \ddots & \\ & & p(\lambda_n) \end{pmatrix} Q^{-1} \end{aligned}$$

That is, $p(A) = Q p(D) Q^{-1}$ and $p(D)$ is just $p(t)$ applied to the diagonal entries (eigenvalues) of A .

• Note that $\sigma(p(A)) = p(\sigma(A))$.

• We can extend this from polynomials to power series whose interval of convergence contains $\sigma(A)$;
if $f(t) = \sum_{k=0}^{\infty} a_k (t-c)^k$ has radius of convergence R and $\sigma(A) \subseteq (c-R, c+R)$, then $f(A)$ exists and
$$f(A) = Q \begin{pmatrix} f(\lambda_1) & & \\ & \ddots & \\ & & f(\lambda_n) \end{pmatrix} Q^{-1}.$$

EX Let $A = \begin{pmatrix} -2 & 6 \\ -3 & 7 \end{pmatrix}$. Then $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1}$ (Exercise). Recall $f(t) = \sqrt{t}$ can be written as a power series

$$f(t) = \sum_{k=0}^{\infty} \frac{f^{(k)}(c)}{k!} (t-c)^k$$

for any $c > 0$, with radius of convergence $R=c$, so that the interval of convergence contains $(0, 2c)$.

Choosing $c=3$ (for example) yields $\sigma(A) = \{1, 4\} \subseteq (0, 6)$. Thus

$$\begin{aligned} \sqrt{A} &= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{1} & 0 \\ 0 & \sqrt{4} \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ -1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix}. \end{aligned}$$

Note that

$$(\sqrt{A})^2 = \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ -1 & 3 \end{pmatrix} = \begin{pmatrix} -2 & 6 \\ -3 & 7 \end{pmatrix} = A$$

That is, \sqrt{A} acts just like the square root of a number. □

Distinct eigenvalues

It turns out if all of the eigenvalues of T are distinct, then A is diagonalizable.

Thm Let $T: V \rightarrow V$ be a linear transformation with $\lambda_1, \dots, \lambda_r$ distinct eigenvalues and $\vec{v}_1, \dots, \vec{v}_r \in V$ corresponding eigenvectors. Then $\vec{v}_1, \dots, \vec{v}_r$ are linearly independent.

Proof We will use proof by induction on r . For the base case $r=1$, since $\vec{v}_1 \neq \vec{0}$ by definition of an eigenvector, the system \vec{v}_1 is linearly independent.

For the induction step, suppose we know the statement is true for $r-1$. Suppose

$$\alpha_1 \vec{v}_1 + \dots + \alpha_r \vec{v}_r = \vec{0}.$$

Applying $A - \lambda_r I$ to each side yields:

$$\alpha_1 (\lambda_1 - \lambda_r) \vec{v}_1 + \dots + \alpha_{r-1} (\lambda_{r-1} - \lambda_r) \vec{v}_{r-1} + \vec{0} = \vec{0}$$

Since $\vec{v}_1, \dots, \vec{v}_{r-1}$ are lin. indep. by our induction hypothesis, we must have

$$\alpha_1 (\lambda_1 - \lambda_r) = \dots = \alpha_{r-1} (\lambda_{r-1} - \lambda_r) = 0.$$

Recall that the eigenvalues are distinct, so $\lambda_i - \lambda_r \neq 0$ and so it must be that

$\alpha_i = 0$ for $i=1, \dots, r-1$. So our original linear combination reduces to

$$\vec{0} + \alpha_r \vec{v}_r = \vec{0}.$$

Since $\vec{v}_r \neq \vec{0}$ (by virtue of being an eigenvector) we must have $\alpha_r = 0$. Thus $\vec{v}_1, \dots, \vec{v}_r$ are linearly independent. Therefore induction yields the theorem for all $r \in \mathbb{N}$. □

Cor If $T: V \rightarrow V$ has $\dim(V)$ distinct eigenvalues, then it is diagonalizable. 11/7

Proof Let $n = \dim(V)$, and let $\vec{v}_1, \dots, \vec{v}_n$ be eigenvectors of T with distinct eigenvalues. The theorem implies they are linearly independent. Since there are $n = \dim(V)$ of them, they form a basis, which means T is diagonalizable. \square

Criteria for Diagonalizability

Thm Let V be a finite-dimensional vector space with $\dim(V) = n$. Let $T: V \rightarrow V$ have eigenvalues $\lambda_1, \dots, \lambda_n$ counting multiplicities. Then T is diagonalizable if and only if

$$\dim(\ker(T - \lambda I)) = m_\lambda(T)$$

for all $\lambda \in \sigma(T)$.

Proof (\Rightarrow) Let B be a basis s.t. $[T]_B^B$ is diagonal. By reordering B , we can assume

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

Thus $\dim(\ker(T - \lambda I))$ is the number of $\lambda_i = \lambda$, which equals $m_\lambda(T)$.

(\Leftarrow) Suppose $\dim(\ker(T - \lambda I)) = m_\lambda(T) \forall \lambda \in \sigma(T)$. Let $\lambda_1, \dots, \lambda_p$ be the distinct eigenvalues of T .

For each $k=1, \dots, p$, let $V_k = \ker(T - \lambda_k I)$ and let B_k be a basis for V_k . We claim

$$B := \bigcup_{k=1}^p B_k$$

is a basis for V , which will imply T is diagonalizable since B consists of eigenvectors. Note that by a Proposition from Section 4.1

$$\sum_{k=1}^p \dim(V_k) = \sum_{k=1}^p m_{\lambda_k}(T) = \dim(V).$$

So B contains $\dim(V)$ vectors and therefore it suffices to show they are lin. indep.

Let $B = \{\vec{u}_1, \dots, \vec{u}_n\}$ and set $I_k = \{i : \vec{u}_i \in V_k\}$, $k=1, \dots, p$. Suppose

$$\alpha_1 \vec{u}_1 + \dots + \alpha_n \vec{u}_n = \vec{0}.$$

Write $\vec{v}_k := \sum_{i \in I_k} \alpha_i \vec{u}_i \in V_k = \ker(T - \lambda_k I)$. We claim $\vec{v}_1 = \dots = \vec{v}_p = \vec{0}$. Indeed, $\vec{v}_1 + \dots + \vec{v}_p = \vec{0}$

and so if any are non-zero then we have a non-trivial linear comb. summing to zero. At the same time, the non-zero vectors are eigenvectors with distinct eigenvalues and hence are linearly independent, a contradiction. Thus

$$\sum_{i \in I_k} \alpha_i \vec{u}_i = \vec{v}_k = \vec{0}$$

and since B_k is a basis for V_k , we must have $\alpha_i = 0 \forall i \in I_k$, and each $k=1, \dots, p$. Thus B is a lin. indep. system and therefore a basis. \square

Ex Recall $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ has $\dim(\ker(A - 0 \cdot I)) = 1 < 3 = m_0(A)$. So A is not diagonalizable. Consequently, for any invertible matrix Q , $Q A Q^{-1}$ is not diagonal. \square

• Note that A in the previous example is at least upper triangular. It turns out this is the best we can hope for in general.

Thm Let V be finite-dimensional with $\dim(V) = n$. Let $T: V \rightarrow V$ have eigenvalues $\lambda_1, \dots, \lambda_n$. Then there exists a basis B for V s.t.

$$[T]_B^B = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \quad \leftarrow \text{arbitrary}$$

Proof We proceed by induction on n .

Base Case: $n=1$. Let $\vec{v} \in V$ be s.t. $V = \text{span}\{\vec{v}\}$. Then $T: V \rightarrow V$ implies \vec{v} is an eigenvector of T , say with eigenvalue λ . Then for $B = \{\vec{v}\}$, we have $[T]_B^B = (\lambda)$.

Induction Step: Suppose the theorem holds for dimension $n-1$. Let $\vec{v}_1 \in V$ be an eigenvector of T with eigenvalue λ_1 . Complete to a basis $B = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\}$. By Exercise 5 on Homework 9,

$$[T]_B^B = \begin{pmatrix} \lambda_1 & X \\ 0 & Y \end{pmatrix}$$

for some $X \in M_{1 \times (n-1)}$, $Y \in M_{(n-1) \times (n-1)}$. Let $C = \{\vec{v}_2, \dots, \vec{v}_n\}$, $W = \text{span} C$, and define $S: W \rightarrow W$ by $[S]_C^C = Y$. By Exercise 4 on Homework 9,

$$\text{char}_T(z) = (\lambda_1 - z) \cdot \text{char}_Y(z) = (\lambda_1 - z) \text{char}_S(z).$$

It follows that $\lambda_2, \dots, \lambda_n$ are eigenvalues of S . Since $\dim(W) = n-1$, our induction hypothesis implies there is a basis C' for W s.t.

$$[S]_{C'}^{C'} = \begin{pmatrix} \lambda_2 & X \\ 0 & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

Define $B' := \{\vec{v}_1\} \cup C'$. Then

$$\begin{aligned} [T]_{B'}^{B'} &= [I]_B^B [T]_B^B [I]_{B'}^B \\ &= \begin{pmatrix} 1 & 0 \\ 0 & [I]_{C'}^{C'} \end{pmatrix} \begin{pmatrix} \lambda_1 & X \\ 0 & [S]_C^C \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & [I]_{C'}^{C'} \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & * \\ 0 & [I]_{C'}^{C'} [S]_C^C [I]_{C'}^{C'} \end{pmatrix} = \begin{pmatrix} \lambda_1 & * \\ 0 & [S]_{C'}^{C'} \end{pmatrix} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix} \end{aligned}$$

Induction then completes the proof. □

Cor Every $A \in M_{n \times n}(\mathbb{C})$ is similar to an upper triangular matrix.

• Since \det and Tr are invariant under similarity, this gives a much simpler proof of the formulas

$$\begin{aligned} \text{Tr}(A) &= \lambda_1 + \dots + \lambda_n \\ \det(A) &= \lambda_1 \cdots \lambda_n. \end{aligned}$$

Fact For every $A \in M_{n \times n}(\mathbb{C})$ there exists a unitary matrix $U \in M_{n \times n}(\mathbb{C})$ ($U^* = U^{-1}$) such that $U A U^{-1} = U A U^*$ is upper triangular.

Cor If $A \in M_{n \times n}(\mathbb{C})$ satisfies $A = A^*$, then A is diagonalizable and all of its eigenvalues are real.

Proof By the fact, $\exists U \in \mathbb{R}^{n \times n}$ unitary s.t. $T := UAU^*$ is upper triangular. But then

$$T^* = (U^*)^* A^* U^* = UAU^* = T$$

So $T = T^*$. This implies T is also lower triangular and hence diagonal. Moreover, we have

$$\overline{(T)_{ii}} = (T^*)_{ii} = (T)_{ii}$$

So the diagonal entries of T are real, but these are precisely the eigenvalues of A . \square

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

$$A \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, A^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, A^3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = A \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \end{pmatrix}, A^4 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = A \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 3 \end{pmatrix}, \dots, A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix}$$

where F_n is the n th Fibonacci number. Recall from Homework 9 that $\sigma(A) = \left\{ \frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2} \right\}$.

Denote

$$\tau := \frac{1-\sqrt{5}}{2}$$

$$\varphi := \frac{1+\sqrt{5}}{2} \leftarrow \text{Golden ratio}$$

You also showed in Homework 9 that

$$\ker(A - \tau I) = \text{span} \left\{ \begin{pmatrix} \tau \\ 1 \end{pmatrix} \right\}$$

$$\ker(A - \varphi I) = \text{span} \left\{ \begin{pmatrix} \varphi \\ 1 \end{pmatrix} \right\}$$

Thus A has the following diagonalization:

$$A = \begin{pmatrix} \varphi & \tau \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \varphi & 0 \\ 0 & \tau \end{pmatrix} \frac{1}{\varphi - \tau} \begin{pmatrix} 1 & -\tau \\ -1 & \varphi \end{pmatrix}$$

Therefore

$$\begin{aligned} A^n &= \begin{pmatrix} \varphi & \tau \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \varphi^n & 0 \\ 0 & \tau^n \end{pmatrix} \frac{1}{\varphi - \tau} \begin{pmatrix} 1 & -\tau \\ -1 & \varphi \end{pmatrix} \\ &= \begin{pmatrix} \varphi^{n+1} & \tau^{n+1} \\ \varphi^n & \tau^n \end{pmatrix} \frac{1}{\varphi - \tau} \begin{pmatrix} 1 & -\tau \\ -1 & \varphi \end{pmatrix} \\ &= \frac{1}{\varphi - \tau} \begin{pmatrix} \varphi^{n+1} - \tau^{n+1} & \varphi \tau^{n+1} - \tau \varphi^{n+1} \\ \varphi^n - \tau^n & \varphi \tau^n - \tau \varphi^n \end{pmatrix} \end{aligned}$$

Thus

$$\begin{pmatrix} F_{n+1} \\ F_n \end{pmatrix} = A^n \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{\varphi - \tau} \begin{pmatrix} \varphi^{n+1} - \tau^{n+1} \\ \varphi^n - \tau^n \end{pmatrix}.$$

This gives us a formula for the Fibonacci numbers:

$$F_n = \frac{\varphi^n - \tau^n}{\varphi - \tau} = \frac{\varphi^n - (-\varphi)^{-n}}{\sqrt{5}} \quad (\tau = -\frac{1}{\varphi})$$

Consider $x, y > 0$ arbitrary. Then

$$\begin{pmatrix} a \\ b \end{pmatrix} := A^n \begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\varphi - \tau} \begin{pmatrix} (\varphi^{n+1} - \tau^{n+1})x + \varphi\tau(\tau^n - \varphi^n)y \\ (\varphi^n - \tau^n)x + \varphi\tau(\tau^{n-1} - \varphi^{n-1})y \end{pmatrix}$$

So using $\varphi \approx 1.618$ and $\tau \approx -0.618$ (so that $\varphi > 1$ and $|\tau| < 1$)

$$\frac{b}{a} = \frac{(\varphi^{n+1} - \tau^{n+1})x + \varphi\tau(\tau^n - \varphi^n)y}{(\varphi^n - \tau^n)x + \varphi\tau(\tau^{n-1} - \varphi^{n-1})y} \approx \frac{\varphi^{n+1}x - \varphi^{n+1}\tau y}{\varphi^n x - \varphi^n \tau y} = \varphi$$

\square