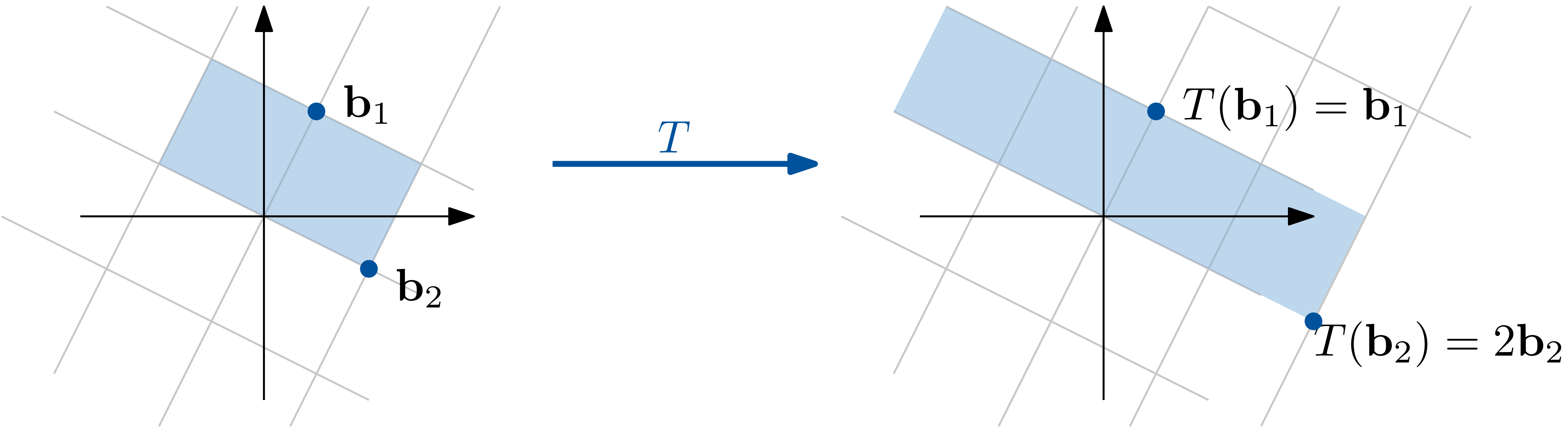
Remember from last week (week 9 p19):
Given a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$, the "right" basis to work in is $\mathcal{B} = \{\mathbf{b}_1, \ldots, \mathbf{b}_n\}$ where $T(\mathbf{b}_i) = \lambda_i \mathbf{b}_i$ for some scalars λ_i. Then the matrix for T relative to $\mathcal{B}$ is a diagonal matrix:

$$[T]_\mathcal{B} = \begin{bmatrix} | & & | \\ [T(\mathbf{b}_1)]_\mathcal{B} & \ldots & [T(\mathbf{b}_n)]_\mathcal{B} \\ | & & | \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}.$$

Computers are much faster and more accurate when they work with diagonal matrices, because many entries are 0.

Also, it's much easier to understand the linear transformation T from a diagonal matrix, e.g. if $T(\mathbf{b}_1) = \mathbf{b}_1$ and $T(\mathbf{b}_2) = 2\mathbf{b}_2$, so $[T]_{\mathcal{B}} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$, then T is an expansion by a factor of 2 in the $\mathbf{b}_2$ direction.



So it is important to study the equation $T(\mathbf{x}) = \lambda\mathbf{x}$.

§5.1-5.2: Eigenvectors and Eigenvalues

Definition: Let A be a square matrix.
An *eigenvector* of A is a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some scalar λ.
Then we call $\mathbf{x}$ an *eigenvector corresponding to λ* (or a λ-eigenvector).
An *eigenvalue* of A is a scalar λ such that $A\mathbf{x} = \lambda\mathbf{x}$ for some nonzero vector $\mathbf{x}$.

If $\mathbf{x}$ is an eigenvector of A, then $\mathbf{x}$ and its image $A\mathbf{x}$ are in the same (or opposite, if $\lambda < 0$) direction. Multiplication by A stretches $\mathbf{x}$ by a factor of λ.
If $\mathbf{x}$ is not an eigenvector, then $\mathbf{x}$ and $A\mathbf{x}$ are not geometrically related in any obvious way.

Warning: eigenvalues and eigenvectors exist for square matrices only. If A is not a square matrix, then $\mathbf{x}$ and $A\mathbf{x}$ are in different vector spaces (they are column vectors with a different number of rows), so it doesn't make sense to ask whether $A\mathbf{x}$ is a multiple of $\mathbf{x}$.

$$A\mathbf{x} = \lambda\mathbf{x}$$

eigenvector: cannot be $\mathbf{0}$.
$A\mathbf{0} = \lambda\mathbf{0}$ is always true, so it holds no information about A.

eigenvalue: can be 0.
$A\mathbf{x} = 0\mathbf{x}$ for a nonzero vector $\mathbf{x}$ does hold information about A - it tells you that A is not invertible.
In fact, A is invertible if and only if 0 is not an eigenvalue.

Important computations:

i given an eigenvalue, how to find the corresponding eigenvectors (p5-8, §5.1);

ii how to find the eigenvalues (p9-11, §5.2);

iii how to determine if there is a basis $\mathcal{B} = \{\mathbf{b}_1, \dots, \mathbf{b}_n\}$ of $\mathbb{R}^n$ where each $\mathbf{b}_i$ is an eigenvector (p13-28, §5.3).

Warm up:

Example: Let $A = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix}$. Determine whether $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ are eigenvectors of A.

Answer:

$\begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 12 \\ -3 \end{bmatrix}$ is not a multiple of $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ (because its entries are not equal), so $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is not an eigenvector of A.

$\begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \begin{bmatrix} -12 \\ 6 \end{bmatrix} = 6 \begin{bmatrix} -2 \\ 1 \end{bmatrix}$, so $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ is an eigenvector of A corresponding to the eigenvalue 6.

i: Given the eigenvalues, find the corresponding eigenvectors:

i.e. we know λ, and we want to solve $A\mathbf{x} = \lambda\mathbf{x}$.
This equation is equivalent to $A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$,
which is equivalent to $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

So the eigenvectors of A corresponding to the eigenvalue λ are the nonzero solutions to $(A - \lambda I)\mathbf{x} = \mathbf{0}$, which we can find by row-reducing $A - \lambda I$.

i: Given the eigenvalues, find the corresponding eigenvectors:

i.e. we know λ, and we want to solve $A\mathbf{x} = \lambda\mathbf{x}$.
This equation is equivalent to $A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$,
which is equivalent to $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

So the eigenvectors of A corresponding to the eigenvalue λ are the nonzero solutions to $(A - \lambda I)\mathbf{x} = \mathbf{0}$, which we can find by row-reducing $A - \lambda I$.

Example: Let $A = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix}$. Find the eigenvectors of A corresponding to the eigenvalue 2.

Answer: $A - 2I_2 = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 8-2 & 4 \\ -3 & 0-2 \end{bmatrix} = \begin{bmatrix} 6 & 4 \\ -3 & -2 \end{bmatrix}$.
$\left[\begin{array}{cc|c} 6 & 4 & 0 \\ -3 & -2 & 0 \end{array}\right] \xrightarrow{\text{row reduction}} \left[\begin{array}{cc|c} 1 & 2/3 & 0 \\ 0 & 0 & 0 \end{array}\right]$ so the eigenvectors are $\begin{bmatrix} -2/3 \\ 1 \end{bmatrix} s$ where s is any nonzero value. A nicer-looking answer: $\begin{bmatrix} -2 \\ 3 \end{bmatrix} s$ where s is any nonzero value.

Because it is sometimes convenient to talk about the eigenvectors and $\mathbf{0}$ together:

Definition: The *eigenspace* of A corresponding to the eigenvalue λ (or the λ-eigenspace of A, sometimes written $E_\lambda(A)$) is the solution set to $(A - \lambda I)\mathbf{x} = \mathbf{0}$.

Because λ-eigenspace of A is the null space of $A - \lambda I$, eigenspaces are subspaces. In the previous example, the eigenspace is a line, but there can also be two-dimensional eigenspaces:

Example: Let $B = \begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix}$. Find a basis for the eigenspace corresponding to the eigenvalue -3.

Answer:

$$B - (-3)I_3 = \begin{bmatrix} -3+3 & 0 & 0 \\ -1 & -2+3 & 1 \\ -1 & 1 & -2+3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \xrightarrow{\text{row-reduction}} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

So solutions are $x_2 \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, for all values of x_2, x_3. So a basis is $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$.

Checking our answer: $\begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -3 \\ -3 \\ 0 \end{bmatrix}$, $\begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -3 \\ 0 \\ -3 \end{bmatrix}$.

Also, be careful how you write your answer, depending on what the question asks for:

The eigenvectors: $s \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$, where s, t are not both zero.

The eigenspace: $s \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ where s, t can take any value.

A basis for the eigenspace: $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$

Tip: if you found that $(B - \lambda I)\mathbf{x} = \mathbf{0}$ has no nonzero solutions, then you've made an arithmetic error. Please do not write that the eigenvector is $\mathbf{0}$!

ii: Given a matrix, find its eigenvalues:

λ is an eigenvalue of A if $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has non-trivial solutions.
By the Invertible Matrix Theorem, this happens precisely when $A - \lambda I$ is not invertible.
So we must have $\det(A - \lambda I) = 0$.

$\det(A - \lambda I)$ is the characteristic polynomial of A (sometimes written χ_A). If A is $n \times n$, then this is a polynomial of degree n. So A has at most n different eigenvalues.
$\det(A - \lambda I) = 0$ is the characteristic equation of A.
We find the eigenvalues by solving the characteristic equation.

ii: Given a matrix, find its eigenvalues:

λ is an eigenvalue of A if $(A - \lambda I)\mathbf{x} = \mathbf{0}$ has non-trivial solutions.
By the Invertible Matrix Theorem, this happens precisely when $A - \lambda I$ is not invertible.
So we must have $\det(A - \lambda I) = 0$.

$\det(A - \lambda I)$ is the characteristic polynomial of A (sometimes written χ_A). If A is $n \times n$, then this is a polynomial of degree n. So A has at most n different eigenvalues.
$\det(A - \lambda I) = 0$ is the characteristic equation of A.
We find the eigenvalues by solving the characteristic equation.

Example: Find the eigenvalues of $A = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix}$.

Answer: $\det(A - \lambda I) = \begin{vmatrix} 8-\lambda & 4 \\ -3 & 0-\lambda \end{vmatrix} = (8-\lambda)(0-\lambda) - (-3)4 = \lambda^2 - 8\lambda + 12.$

So the eigenvalues are the solutions to $\lambda^2 - 8\lambda + 12 = 0$.
Factor: $(\lambda - 2)(\lambda - 6) = 0 \implies \lambda = 2$ and 6.

We find the eigenvalues by solving the characteristic equation $\det(A - \lambda I) = 0$.

Example: Find the eigenvalues of $B = \begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix}$.

Answer:

(expand along top row)

$$\det(B - \lambda I) = \begin{vmatrix} -3-\lambda & 0 & 0 \\ -1 & -2-\lambda & 1 \\ -1 & 1 & -2-\lambda \end{vmatrix} = (-3-\lambda)\begin{vmatrix} -2-\lambda & 1 \\ 1 & -2-\lambda \end{vmatrix}$$

$$= (-3-\lambda)[(-2-\lambda)(-2-\lambda) - 1]$$

Tip: if you already have a factor, don't expand it

$$= (-3-\lambda)[\lambda^2 + 4\lambda + 3]$$

$$= (-3-\lambda)(\lambda+3)(\lambda+1).$$

So the eigenvalues are the solutions to $(-3-\lambda)(\lambda+3)(\lambda+1) = 0$,

which are $-3, \quad -3, \quad -1.$

Tips:

- Because of the variable λ, it is easier to find $\det(A - \lambda I)$ by expanding across rows or down columns than by using row operations.
- If you already have a factor, do not expand it (e.g. previous page)
- Do not "cancel" λ in the characteristic equation: remember that $\lambda = 0$ can be an eigenvalue (see below).
- The eigenvalues of A are usually not related to the eigenvalues of rref(A).

Tips:

- Because of the variable λ, it is easier to find $\det(A - \lambda I)$ by expanding across rows or down columns than by using row operations.
- If you already have a factor, do not expand it (e.g. previous page)
- Do not "cancel" λ in the characteristic equation: remember that $\lambda = 0$ can be an eigenvalue (see below).
- The eigenvalues of A are usually not related to the eigenvalues of rref(A).

Example: Find the eigenvalues of $C = \begin{bmatrix} 3 & 6 & -2 \\ 0 & 0 & 2 \\ 0 & 0 & 6 \end{bmatrix}$.

Answer: $C - \lambda I = \begin{bmatrix} 3-\lambda & 6 & -2 \\ 0 & -\lambda & 2 \\ 0 & 0 & 6-\lambda \end{bmatrix}$ is upper triangular, so its determinant is the product of its diagonal entries: $\det(C - \lambda I) = (3-\lambda)(-\lambda)(6-\lambda)$, whose solutions are 3, 0, 6.

By a similar argument (for upper or lower triangular matrices):

Fact: The eigenvalues of a triangular matrix are the diagonal entries.

Summary: To find the eigenvalues and eigenvectors of a square matrix A:

Step 1 Solve the characteristic equation $\det(A - \lambda I) = 0$ to find the eigenvalues;

Step 2 For each eigenvalue λ, solve $(A - \lambda I)\mathbf{x} = \mathbf{0}$ to find the eigenvectors.

Summary: To find the eigenvalues and eigenvectors of a square matrix A:
Step 1 Solve the characteristic equation $\det(A - \lambda I) = 0$ to find the eigenvalues;
Step 2 For each eigenvalue λ, solve $(A - \lambda I)\mathbf{x} = \mathbf{0}$ to find the eigenvectors.

Thinking about eigenvectors conceptually:

Suppose $\mathbf{v}$ is an eigenvector of A corresponding to the eigenvalue λ.
Then
$$A^2(\mathbf{v}) = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda A\mathbf{v} = \lambda(\lambda\mathbf{v}) = \lambda^2\mathbf{v}.$$
So any eigenvector of A is also an eigenvector of A^2, corresponding to the square of the previous eigenvalue.

Summary: To find the eigenvalues and eigenvectors of a square matrix A:
Step 1 Solve the characteristic equation $\det(A - \lambda I) = 0$ to find the eigenvalues;
Step 2 For each eigenvalue λ, solve $(A - \lambda I)\mathbf{x} = \mathbf{0}$ to find the eigenvectors.

Thinking about eigenvectors conceptually:

Suppose $\mathbf{v}$ is an eigenvector of A corresponding to the eigenvalue λ.
Then

$$A^2(\mathbf{v}) = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda A\mathbf{v} = \lambda(\lambda\mathbf{v}) = \lambda^2\mathbf{v}.$$

So any eigenvector of A is also an eigenvector of A^2, corresponding to the square of the previous eigenvalue.

We can also define eigenvalues and eigenvectors for a linear transformation $T : V \to V$ on an abstract vector space V: a nonzero vector $\mathbf{v}$ in V is an eigenvector of T with corresponding eigenvalue λ if $T(\mathbf{v}) = \lambda\mathbf{v}$.

Example: Consider $T : \mathbb{P}_3 \to \mathbb{P}_3$ given by $T(\mathbf{p}) = x\frac{d}{dx}\mathbf{p}$. Then $\mathbf{p}(x) = x^2$ is an eigenvector of T corresponding to the eigenvalue 2, because $T(x^2) = x\frac{d}{dx}x^2 = x2x = 2x^2$.

§5.3: Diagonalisation

Remember that our motivation for finding eigenvectors is to find a basis relative to which a linear transformation is represented by a diagonal matrix.

Definition: (week 9 p16) Two square matrices A and B are *similar* if there is an invertible matrix P such that $A = PBP^{-1}$.

From the change-of-coordinates formula (week 9 p13)

$$[T]_{\mathcal{E}} = \underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}}[T]_{\mathcal{B}}\underset{\mathcal{B}\leftarrow\mathcal{E}}{\mathcal{P}} = \underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}}[T]_{\mathcal{B}}\underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}}^{-1},$$

similar matrices represent the same linear transformation relative to different bases.

§5.3: Diagonalisation

Remember that our motivation for finding eigenvectors is to find a basis relative to which a linear transformation is represented by a diagonal matrix.

Definition: (week 9 p16) Two square matrices A and B are *similar* if there is an invertible matrix P such that $A = PBP^{-1}$.

From the change-of-coordinates formula (week 9 p13)

$$[T]_{\mathcal{E}} = \underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}}[T]_{\mathcal{B}}\underset{\mathcal{B}\leftarrow\mathcal{E}}{\mathcal{P}} = \underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}}[T]_{\mathcal{B}}\underset{\mathcal{E}\leftarrow\mathcal{B}}{\mathcal{P}}^{-1},$$

similar matrices represent the same linear transformation relative to different bases.

Definition: A square matrix A is *diagonalisable* if it is similar to a diagonal matrix, i.e. if there is an invertible matrix P and a diagonal matrix D such that $A = PDP^{-1}$.

Theorem 5: Diagonalisation Theorem: An $n \times n$ matrix A is diagonalisable (i.e. $A = PDP^{-1}$) if and only if A has n linearly independent eigenvectors.

Proof: we prove a stronger theorem: An $n \times n$ matrix A satisfies $AP = PD$ for a $n \times k$ matrix P and a diagonal $k \times k$ matrix D if and only if the ith column of P is an eigenvector of A with eigenvalue d_{ii}, or is the zero vector. This comes from equating column by column the right hand sides of the following equations:

$$AP = A \begin{bmatrix} | & & | \\ \mathbf{p}_1 & \dots & \mathbf{p}_k \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ A\mathbf{p}_1 & \dots & A\mathbf{p}_k \\ | & & | \end{bmatrix}$$

$$PD = \begin{bmatrix} | & & | \\ \mathbf{p}_1 & \dots & \mathbf{p}_k \\ | & & | \end{bmatrix} \begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & & \\ \vdots & & \ddots & \\ 0 & & & d_{kk} \end{bmatrix} = \begin{bmatrix} | & & | \\ d_{11}\mathbf{p}_1 & \dots & d_{kk}\mathbf{p}_k \\ | & & | \end{bmatrix}$$

Theorem 5: Diagonalisation Theorem: An $n \times n$ matrix A is diagonalisable (i.e. $A = PDP^{-1}$) if and only if A has n linearly independent eigenvectors.

Proof: we prove a stronger theorem: An $n \times n$ matrix A satisfies $AP = PD$ for a $n \times k$ matrix P and a diagonal $k \times k$ matrix D if and only if the ith column of P is an eigenvector of A with eigenvalue d_{ii}, or is the zero vector. This comes from equating column by column the right hand sides of the following equations:

$$AP = A\begin{bmatrix} | & & | \\ \mathbf{p}_1 & \dots & \mathbf{p}_k \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ A\mathbf{p}_1 & \dots & A\mathbf{p}_k \\ | & & | \end{bmatrix}$$

$$PD = \begin{bmatrix} | & & | \\ \mathbf{p}_1 & \dots & \mathbf{p}_k \\ | & & | \end{bmatrix}\begin{bmatrix} d_{11} & 0 & \dots & 0 \\ 0 & d_{22} & & \\ \vdots & & \ddots & \\ 0 & & & d_{kk} \end{bmatrix} = \begin{bmatrix} | & & | \\ d_{11}\mathbf{p}_1 & \dots & d_{kk}\mathbf{p}_k \\ | & & | \end{bmatrix}$$

To deduce The Diagonalisation Theorem, note that $A = PDP^{-1}$ if and only if $AP = PD$ and P is invertible, i.e. (using Invertible Matrix Theorem) if and only if $AP = PD$ and the n columns of P are linearly independent.

Example: Diagonalise $A = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix}$.

Answer:

Step 1 Solve the characteristic equation $\det(A - \lambda I) = 0$ to find the eigenvalues.

From p9, $\det(A - \lambda I) = \lambda^2 - 8\lambda + 12$, eigenvalues are 2 and 6.

Step 2 For each eigenvalue λ, solve $(A - \lambda I)\mathbf{x} = \mathbf{0}$ to find a basis for the λ-eigenspace.

From p6, $\left\{ \begin{bmatrix} -2 \\ 3 \end{bmatrix} \right\}$ is a basis for the 2-eigenspace,

You can check that $\left\{ \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right\}$ is a basis for the 6-eigenspace.

Notice that these two eigenvectors are linearly independent (this is automatic, p20).

If Step 2 gives fewer than n vectors, A is not diagonalisable (p25). Otherwise, continue:

Step 3 Put the eigenvectors from Step 2 as the columns of P. $P = \begin{bmatrix} -2 & -2 \\ 3 & 1 \end{bmatrix}$.

Step 4 Put the corresponding eigenvalues as the diagonal entries of D. $D = \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix}$.

Checking our answer: $PDP^{-1} = \begin{bmatrix} -2 & -2 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 1 & 2 \\ -3 & -2 \end{bmatrix} = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix} = A.$

The matrices P and D are not unique:

- In Step 2, we can choose a different basis for the eigenspaces: e.g. using $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ instead of $\begin{bmatrix} -2 \\ 3 \end{bmatrix}$ as a basis for the 2-eigenspace, we can take $P = \begin{bmatrix} 2 & -2 \\ -3 & 1 \end{bmatrix}$, and then $PDP^{-1} = \begin{bmatrix} 2 & -2 \\ -3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 0 & 6 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 1 & 2 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix} = A.$
- In Step 3, we can choose a different order for the columns of P, as long as we put the entries of D in the corresponding order: e.g. $P = \begin{bmatrix} -2 & -2 \\ 1 & 3 \end{bmatrix}$, $D = \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix}$ then $PDP^{-1} = \begin{bmatrix} -2 & -2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} \frac{1}{4} \begin{bmatrix} 3 & 2 \\ -1 & -2 \end{bmatrix} = \begin{bmatrix} 8 & 4 \\ -3 & 0 \end{bmatrix} = A.$

Example: Diagonalise $B = \begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix}$.

Answer:

Step 1 Solve the characteristic equation $\det(B - \lambda I) = 0$ to find the eigenvalues.

From p10, $\det(B - \lambda I) = (-3 - \lambda)(\lambda + 3)(\lambda + 1)$, so the eigenvalues are -3 and -1.

Step 2 For each eigenvalue λ, solve $(B - \lambda I)\mathbf{x} = \mathbf{0}$ to find a basis for the λ-eigenspace.

From p7, $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for the -3-eigenspace; you can check that $\left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a basis for the -1-eigenspace. You can check that these three eigenvectors are linearly independent (this is automatic, see p20).

If Step 2 gives fewer than n vectors, A is not diagonalisable (p25). Otherwise, continue:

Step 3 Put the eigenvectors from Step 2 as the columns of P.

Step 4 Put the corresponding eigenvalues as the diagonal entries of D.

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}, D = \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Remember that $B = PDP^{-1}$ if and only if $BP = PD$ and P is invertible. This allows us to check our answer without inverting P:

$$BP = \begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -3 & -3 & 0 \\ -3 & 0 & -1 \\ 0 & -3 & -1 \end{bmatrix},$$

$$PD = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -3 & -3 & 0 \\ -3 & 0 & -1 \\ 0 & -3 & -1 \end{bmatrix} = BP, \text{ and}$$

$$\det P = \begin{vmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} = 1 \begin{vmatrix} 0 & 1 \\ 1 & 1 \end{vmatrix} = -1 \neq 0.$$

We can use the matrices P and D to quickly calculate powers of B (see also week 9 p18):

$$\begin{aligned} B^3 &= (PDP^{-1})^3 \\ &= (PDP^{-1})(PDP^{-1})(PDP^{-1}) \\ &= PD^3P^{-1} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & -1 \end{bmatrix}^3 \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} -27 & 0 & 0 \\ 0 & -27 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -27 & 0 & 0 \\ -13 & -14 & 13 \\ -13 & 13 & -14 \end{bmatrix}. \end{aligned}$$

(This sometimes works for fractional and negative powers too, see Homework 5 Q3.)

At the end of Step 2, after finding a basis for each eigenspace, it is unnecessary to explicitly check that the eigenvectors in the different bases, together, are linearly independent:

Theorem 7c: Linear Independence of Eigenvectors: If $\mathcal{B}_1, \dots, \mathcal{B}_p$ are linearly independent sets of eigenvectors of a matrix A, corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_p$, then the total collection of vectors in the sets $\mathcal{B}_1, \dots, \mathcal{B}_p$ is linearly independent. (Proof idea: see practice problem 3 in §5.1 of textbook.)

Example: In the previous example, $\mathcal{B}_1 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a linearly independent set in the -3-eigenspace, $\mathcal{B}_2 = \left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ is a linearly independent set in the -1-eigenspace, so the theorem says that $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \right\}$ is linearly independent.

An important special case of Theorem 7c is when each $\mathcal{B}_i$ contains a single vector:
Theorem 2: Linear Independence of Eigenvectors: If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are eigenvectors of a matrix A corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_p$, then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly independent.

To give you an idea of why this is true, we prove it in the simple case $p = 2$:
Suppose $A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda_2 \mathbf{v}_2$, and $\mathbf{v}_1, \mathbf{v}_2 \neq \mathbf{0}$ and $\lambda_1 \neq \lambda_2$. We want to show that $c_1 = c_2 = 0$ is the only solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}. \qquad (*)$$

An important special case of Theorem 7c is when each $\mathcal{B}_i$ contains a single vector:

Theorem 2: Linear Independence of Eigenvectors: If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are eigenvectors of a matrix A corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_p$, then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly independent.

To give you an idea of why this is true, we prove it in the simple case $p = 2$:
Suppose $A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda_2 \mathbf{v}_2$, and $\mathbf{v}_1, \mathbf{v}_2 \neq \mathbf{0}$ and $\lambda_1 \neq \lambda_2$. We want to show that $c_1 = c_2 = 0$ is the only solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}. \qquad (*)$$

Multiply both sides by A:

$$c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 = \mathbf{0}.$$

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 = \mathbf{0}. \qquad (**)$$

An important special case of Theorem 7c is when each $\mathcal{B}_i$ contains a single vector:
Theorem 2: Linear Independence of Eigenvectors: If $\mathbf{v}_1, \dots, \mathbf{v}_p$ are eigenvectors of a matrix A corresponding to distinct eigenvalues $\lambda_1, \dots, \lambda_p$, then the set $\{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ is linearly independent.

To give you an idea of why this is true, we prove it in the simple case $p = 2$:
Suppose $A\mathbf{v}_1 = \lambda_1 \mathbf{v}_1$ and $A\mathbf{v}_2 = \lambda_2 \mathbf{v}_2$, and $\mathbf{v}_1, \mathbf{v}_2 \neq \mathbf{0}$ and $\lambda_1 \neq \lambda_2$. We want to show that $c_1 = c_2 = 0$ is the only solution to

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 = \mathbf{0}. \qquad (*)$$

Multiply both sides by A:

$$c_1 A\mathbf{v}_1 + c_2 A\mathbf{v}_2 = \mathbf{0}.$$

$$c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 = \mathbf{0}. \qquad (**)$$

Multiply equation $(*)$ by λ_1, and subtract from equation $(**)$:

$$\mathbf{0} = c_1(\lambda_1 - \lambda_1)\mathbf{v}_1 + c_2(\lambda_2 - \lambda_1)\mathbf{v}_2 = \mathbf{0}.$$

Because $\lambda_2 \neq \lambda_1$ and $\mathbf{v}_2 \neq \mathbf{0}$, this implies $c_2 = 0$; substituting into $(*)$ shows $c_1 = 0$.

One proof for $p > 2$ is to do this (multiply by A, multiply by λ_i, subtract) $p - 1$ times. P288 in the textbook phrases this differently, as a proof by contradiction.

iii.ii: Determine if a matrix is diagonalisable

From the Diagonalisation Theorem, we know that A is diagonalisable if and only if A has n linearly independent eigenvectors. Can we determine if A has enough eigenvectors without finding all those eigenvectors?

To do so, we need an extra idea:
Definition: The (algebraic) multiplicity of an eigenvalue λ_k is its multiplicity as a root of the characteristic equation, i.e. it is the number of times the linear factor $(\lambda - \lambda_k)$ occurs in $\det(A - \lambda I)$.

Example: Consider $B = \begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix}$. From p10, the characteristic polynomial of B is $\det(B - \lambda I) = (-3 - \lambda)(\lambda + 3)(\lambda + 1) = -(\lambda + 3)(\lambda + 3)(\lambda + 1)$. So -3 has multiplicity 2, and -1 has multiplicity 1.

Theorem 7b: Diagonalisability Criteria: An $n \times n$ matrix A is diagonalisable if and only if both the following conditions are true:

i. the characteristic polynomial $\det(A - \lambda I)$ factors completely into linear factors (i.e. it has n solutions counting with multiplicity);

ii. for each eigenvalue λ_k, the dimension of the λ_k-eigenspace is equal to the multiplicity of λ_k.

Theorem 7b: Diagonalisability Criteria: An $n \times n$ matrix A is diagonalisable if and only if both the following conditions are true:

i the characteristic polynomial $\det(A - \lambda I)$ factors completely into linear factors (i.e. it has n solutions counting with multiplicity);

ii for each eigenvalue λ_k, the dimension of the λ_k-eigenspace is equal to the multiplicity of λ_k.

Example: (failure of i) Consider $\begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix}$, the standard matrix for rotation through $\frac{\pi}{6}$. Its characteristic polynomial is $\begin{vmatrix} \sqrt{3}/2 - \lambda & -1/2 \\ 1/2 & \sqrt{3}/2 - \lambda \end{vmatrix} = (\frac{\sqrt{3}}{2} - \lambda)^2 + \frac{1}{4}$.

This polynomial cannot be written in the form $(\lambda - a)(\lambda - b)$ because it has no solutions, as its value is always $\geq \frac{1}{4}$. So this rotation matrix is not diagonalisable. (This makes sense because, after a rotation through $\frac{\pi}{6}$, no vector is in the same or opposite direction.)

The failure of i can be “fixed” by allowing eigenvalues to be complex numbers, so we concentrate on condition ii.

Theorem 7b: Diagonalisability Criteria: An $n \times n$ matrix A is diagonalisable if and only if both the following conditions are true:

i the characteristic polynomial $\det(A - \lambda I)$ factors completely into linear factors (i.e. it has n solutions counting with multiplicity);

ii for each eigenvalue λ_k, the dimension of the λ_k-eigenspace is equal to the multiplicity of λ_k.

Example: (failure of ii) Consider $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. It is upper triangular, so its eigenvalues are its diagonal entries (with the same multiplicities), i.e. 0 with multiplicity 2. The eigenspace of eigenvalue 0 is the set of solutions to $\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} - 0I_2\right)\mathbf{x} = \mathbf{0}$, which is Span $\left\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right\}$. So the eigenspace has dimension $1 < 2$, and therefore $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ is not diagonalisable.

Fact: (theorem 7a in textbook): the dimension of the λ_k-eigenspace is at most the multiplicity of λ_k.

Theorem 7b: Diagonalisability Criteria: An $n \times n$ matrix A is diagonalisable if and only if both the following conditions are true:

i the characteristic polynomial $\det(A - \lambda I)$ factors completely into linear factors;

ii for each eigenvalue λ_k, the dimension of the λ_k-eigenspace is equal to the multiplicity of λ_k.

Proof: Obvious if you believe that the dimension of each eigenspace is at most the multiplicity, or:

"if" part: if both conditions hold, then A has n linearly independent eigenvectors (using Linear Independence of Eigenvectors Theorem on p20), so A is diagonalisable.

"only if" part: (sketch) the main idea is that the two conditions are true for a diagonal matrix, and similar matrices have the same characteristic polynomial:

$$\begin{aligned}\det(PBP^{-1} - \lambda I) &= \det(PBP^{-1} - \lambda PP^{-1}) \\ &= \det(P(B - \lambda I)P^{-1}) \\ &= \det P \det(B - \lambda I) \det(P^{-1}) \\ &= \det P \det(B - \lambda I) \frac{1}{\det P} = \det(B - \lambda I).\end{aligned}$$

Example: Determine if $B = \begin{bmatrix} -3 & 0 & 0 \\ -1 & -2 & 1 \\ -1 & 1 & -2 \end{bmatrix}$ is diagonalisable.

Answer:

Step 1 Solve $\det(B - \lambda I) = 0$ to find the eigenvalues and multiplicities.

From p10, $\det(B - \lambda I) = (-3 - \lambda)(\lambda + 3)(\lambda + 1)$, so the eigenvalues are -3 (with multiplicity 2) and -1 (with multiplicity 1).

Step 2 For each eigenvalue λ of multiplicity more than 1, find the dimension of the λ-eigenspace (e.g. by row-reducing $(B - \lambda I)$ to echelon form):

The dimensions of all eigenspaces are equal to their multiplicities $\rightarrow$ diagonalisable

The dimension of one eigenspace is less than its multiplicity $\rightarrow$ not diagonalisable

$\lambda = -1$ has multiplicity 1, so we don't need to study it (see p28 for the reason).

$\lambda = -3$ has multiplicity 2, so we need to examine it more closely:

$$B - (-3)I_3 = \begin{bmatrix} -3+3 & 0 & 0 \\ -1 & -2+3 & 1 \\ -1 & 1 & -2+3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \xrightarrow{\text{row-reduction}} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

This has two free variables (x_2, x_3), so the dimension of the -3-eigenspace is two, which is equal to its multiplicity. So B is diagonalisable.

Example: Let $K = \begin{bmatrix} 6 & -4 & 4 & 9 \\ -9 & 9 & 8 & -17 \\ 0 & 0 & 5 & 0 \\ -5 & 4 & -4 & -8 \end{bmatrix}$. Given that $\det(K - \lambda I) = (\lambda - 1)^2(\lambda - 5)^2$, determine if K is diagonalisable.

Answer:

Step 1 Solve $\det(K - \lambda I) = 0$ to find the eigenvalues and multiplicities.

The eigenvalues are 1 (with multiplicity 2) and 5 (with multiplicity 2).

Step 2 For each eigenvalue λ of multiplicity more than 1, find the dimension of the λ-eigenspace (e.g. by row-reducing $(K - \lambda I)$ to echelon form):

The dimensions of all eigenspaces are equal to their multiplicities $\rightarrow$ diagonalisable

The dimension of one eigenspace is less than its multiplicity $\rightarrow$ not diagonalisable

$$\lambda = 1:\ K - 1I_4 = \begin{bmatrix} 5 & -4 & 4 & 9 \\ -9 & 8 & 8 & -17 \\ 0 & 0 & 4 & 0 \\ -5 & 4 & -4 & -9 \end{bmatrix} \rightarrow \begin{bmatrix} 5 & -4 & 4 & 9 \\ 0 & 4 & * & * \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \begin{array}{l} \\ 5R_2 - 9R_1 \\ \\ R_4 - R_1 \end{array}$$

x_4 is the only one free variable, so the dimension of the 1-eigenspace is one, which is less than its multiplicity. So K is not diagonalisable. (We don't need to also check $\lambda = 5$.)

In Step 2, why don't we need to look at eigenvalues with multiplicity 1?

In Step 2, why don't we need to look at eigenvalues with multiplicity 1?
Answer: because the dimension of an eigenspace is always at least 1. So if an eigenvalue has multiplicity 1, then the dimension of its eigenspace must be exactly 1.

So things are very simple when all eigenvalues have multiplicity 1:
Theorem 6: Distinct eigenvalues implies diagonalisable: If an $n \times n$ matrix has n distinct eigenvalues, then it is diagonalisable.

Example: Is $C = \begin{bmatrix} 3 & 6 & -2 \\ 0 & 0 & 2 \\ 0 & 0 & 6 \end{bmatrix}$ diagonalisable?

Answer: C is upper triangular, so its eigenvalues are its diagonal entries (with the same multiplicities), i.e 3, 0 and 6. Since C is 3×3 and it has 3 different eigenvalues, C is diagonalisable.

Warning: an $n \times n$ matrix with fewer than n eigenvalues can still be diagonalisable!

Non-examinable: what to do when A is not diagonalisable:

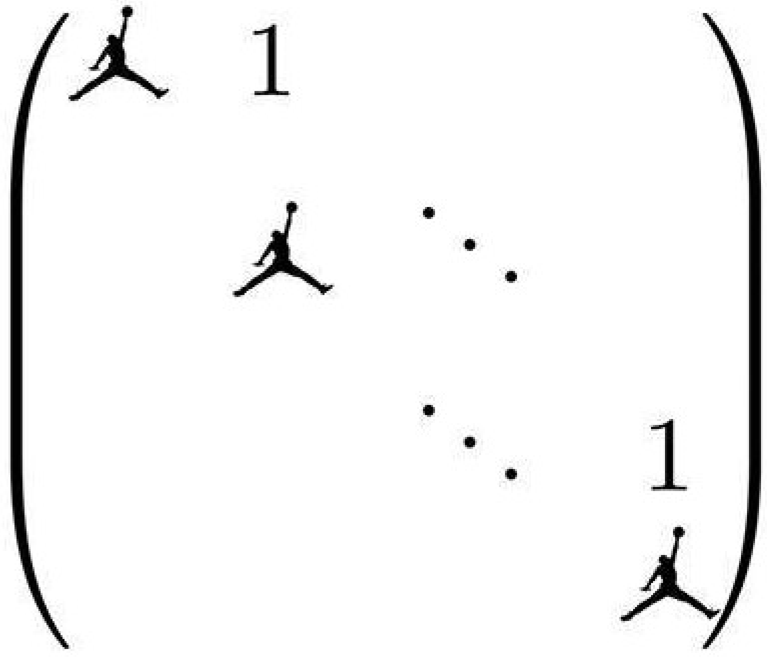
We can still write A as PJP^{-1}, where J is "easy to understand and to compute with". Such a J is called a Jordan form.

For example, all non-diagonalisable 2×2 matrices are similar to $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, where λ is the only eigenvalue (allowing complex eigenvalues).

Non-examinable: what to do when A is not diagonalisable:

We can still write A as PJP^{-1}, where J is “easy to understand and to compute with”. Such a J is called a Jordan form.

For example, all non-diagonalisable 2×2 matrices are similar to $\begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}$, where λ is the only eigenvalue (allowing complex eigenvalues).



A Jordan block of size-n with eigenvalue ,

(A Jordan form may contain more than one Jordan block, e.g. $\begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}$ contains two 2×2 Jordan blocks.)

Non-examinable: rectangular matrices (see §7.4 of textbook):

Any $m \times n$ matrix A can be decomposed as $A = QDP^{-1}$ where:

- P is an invertible $n \times n$ matrix with columns $\mathbf{p}_i$;
- Q is an invertible $m \times m$ matrix with columns $\mathbf{q}_i$;
- D is a "diagonal" $m \times n$ matrix with diagonal entries d_{ii}:

e.g. $\begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & 0 & 0 \end{bmatrix}, \begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \\ 0 & 0 \end{bmatrix}$. So the maximal number of nonzero entries is the smaller of m and n.

Instead of $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$, this decomposition satisfies $A\mathbf{p}_i = d_{ii}\mathbf{q}_i$ for all $i \leq m, n$.

Non-examinable: rectangular matrices (see §7.4 of textbook):

Any $m \times n$ matrix A can be decomposed as $A = QDP^{-1}$ where:

P is an invertible $n \times n$ matrix with columns $\mathbf{p}_i$;

Q is an invertible $m \times m$ matrix with columns $\mathbf{q}_i$;

D is a "diagonal" $m \times n$ matrix with diagonal entries d_{ii}:

e.g. $\begin{bmatrix} d_{11} & 0 & 0 & 0 \\ 0 & d_{22} & 0 & 0 \end{bmatrix}$, $\begin{bmatrix} d_{11} & 0 \\ 0 & d_{22} \\ 0 & 0 \end{bmatrix}$. So the maximal number of nonzero entries is the smaller of m and n.

Instead of $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$, this decomposition satisfies $A\mathbf{p}_i = d_{ii}\mathbf{q}_i$ for all $i \leq m, n$.

An important example is the singular value decomposition $A = U\Sigma V^T$. Each diagonal entry of Σ is a singular value of A, which is the squareroot of an eigenvalue of $A^T A$ (a diagonalisable $n \times n$ matrix with non-negative eigenvalues). The singular values contain a lot of information about A, e.g. the largest singular value is the "maximal length scaling factor" of A. (Even for a square matrix, this is in general not true with the eigenvalues of A, so depending on the problem the SVD may be more useful than the diagonalisation of A.)