CO233 - Computational Techniques

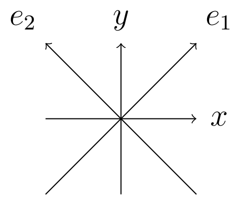
15th January 2020

Vector and Matrix Norms

An orthonormal basis of $\mathbb{R}^n$ are unit vectors that are pairwise mutually perpendicular; such that for $(e_1, \cdots, e_n)$;

- $e_i \cdot e_i = 1$
- $e_i \cdot e_j = 0$, if $i \neq j$

The standard canonical basis of $\mathbb{R}^3$ are the i, j, k vectors, and similar in $\mathbb{R}^2$. However, we can form another orthonormal basis of $\mathbb{R}^2$ by bisecting the angles as such;



If we take a vector $\boldsymbol{v} \in \mathbb{R}^n$, the Euclidean norm (or the ℓ_2-norm) is defined as such;

$$||\boldsymbol{v}||_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$$

A norm, a mapping $|| \; || : \mathbb{R}^n \to \mathbb{R}^+$, must satisfy these 3 axioms;

(i) $||\boldsymbol{v}|| > 0$ given that $\boldsymbol{v} \neq \boldsymbol{0}$

(ii) $||\lambda \boldsymbol{v}|| = |\lambda| \; ||\boldsymbol{v}||$

(iii) $||\boldsymbol{v} + \boldsymbol{w}|| \leq ||\boldsymbol{v}|| + ||\boldsymbol{w}||$ (triangular inequality)

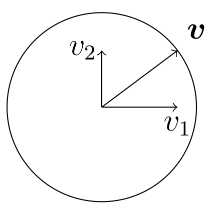
Some other (ℓ_p) norms are defined as follows;

$$\ell_1\text{-norm } ||\boldsymbol{v}||_1 = \sum_{i=1}^{n} |v_i|$$
$$\ell_\infty\text{-norm } ||\boldsymbol{v}||_\infty = \max\{|v_i| : 1 \leq i \leq n\}$$
$$\ell_p\text{-norm } ||\boldsymbol{v}||_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}}$$

In each dimension, we have the following;

- $n = 1$
  - $||\boldsymbol{v}||_1 = |v| = |v|$
  - $||\boldsymbol{v}||_2 = \sqrt{v^2} = |v|$
  - $||\boldsymbol{v}||_\infty = \max\{|v|\} = |v|$
- $n = 2$

  We can represent this geometrically as such;



In our case $||\boldsymbol{v}||_\infty = \max\{|v_1|, |v_2|\} = |v_1|$, but the point is that it's either of the "sides" of the triangle. Obviously, $||\boldsymbol{v}||_2 \geq ||\boldsymbol{v}||_\infty$, as it's the hypotenuse of the triangle, and similarly, $||\boldsymbol{v}||_1 \geq ||\boldsymbol{v}||_2$, due to the triangle inequality. Therefore we have $||\boldsymbol{v}||_1 \geq ||\boldsymbol{v}||_2 \geq ||\boldsymbol{v}||_\infty$.

Even if the orthonormal base changes, the Euclidean norm stays the same, whereas the other norms can change. As such, we can say the ℓ_2-norm is invariant under an **orthogonal transformation** (a basis change from an orthonormal bases to another orthonormal bases).

- $n = ?$ (general)

  The goal is to prove $||\boldsymbol{v}||_\infty \leq ||\boldsymbol{v}||_2 \leq ||\boldsymbol{v}||_1$. If we first take the squares of all of them, such that we have the following;

$$||\boldsymbol{v}||_1^2 = \left(\sum_{i=1}^{n} |v_i|\right)^2$$
$$||\boldsymbol{v}||_2^2 = \sum_{i=1}^{n} |v_i|^2$$
$$||\boldsymbol{v}||_\infty^2 = (\max\{|v_i| : 1 \leq i \leq n\})^2$$

  Since the ℓ_∞-norm corresponds to a single v_i, it's obvious that the following inequality holds (since the ℓ_2-norm squared has all the other terms squared, as well as the ℓ_∞-norm squared);

$$||\boldsymbol{v}||_2^2 = \sum_{i=1}^{n} |v_i|^2 \geq (\max\{|v_i| : 1 \leq i \leq n\})^2 = ||\boldsymbol{v}||_\infty^2 \Rightarrow ||\boldsymbol{v}||_2 \geq ||\boldsymbol{v}||_\infty$$

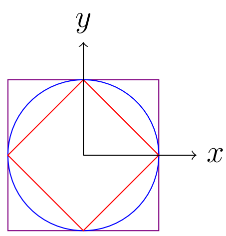
  To prove the other inequality, we see that the square of the sum of absolutes is greater than the sum of the squares, as the square of the sum contains the cross terms (which will be positive).

$$||\boldsymbol{v}||_2^2 = \sum_{i=1}^{n} |v_i|^2 \leq \left(\sum_{i=1}^{n} |v_i|\right)^2 = ||\boldsymbol{v}||_\infty^1 \Rightarrow ||\boldsymbol{v}||_2 \leq ||\boldsymbol{v}||_1$$

  As such, we can conclude that $||\boldsymbol{v}||_\infty \leq ||\boldsymbol{v}||_2 \leq ||\boldsymbol{v}||_1$ in any dimension. ■

Tutorial Question

Find the locus of vectors such that $||\boldsymbol{v}||_p \leq 1$, for $p = 1, 2, \infty$ in $n = 2$;



Imagine they're all shaded from the border to the origin.

ℓ_p-norm

Our goal is to show that as $p \to \infty$, we get the definition of the ℓ_∞-norm previously stated. Take a vector $\boldsymbol{v} \in \mathbb{R}^n$, where both $\boldsymbol{v}$ and n are fixed.

$$||\boldsymbol{v}||_p^p = \sum_{i=1}^{n} |v_i|^p$$

Obviously, this is greater than or equal to $||\boldsymbol{v}||_\infty^p$, as it would only be a single $|v_i|^p$. Similarly, it must be less than or equal to $n|v_i|^p$, as v_i is the maximum of all the components.

$$||\boldsymbol{v}||_\infty^p \leq ||\boldsymbol{v}||_p^p = \sum_{i=1}^{n} |v_i|^p \leq n||\boldsymbol{v}||_\infty^p$$

Taking everything to the power of $\frac{1}{p}$, we obtain the following result (note that $p > 0$ hence the signs don't change);

$$||\boldsymbol{v}||_\infty \leq ||\boldsymbol{v}||_p =\leq n^{\frac{1}{p}}||\boldsymbol{v}||_\infty$$

As $p \to \infty$, since $n \geq 2$ ($n = 1$ is shown to collapse to the same component), we have $n^{\frac{1}{p}} \to 1$, which sandwiches the middle term.

Some Proposition (ℓ_∞-norm vs ℓ_2-norm)

The proposition is as follows; for a vector $\boldsymbol{v} \in \mathbb{R}^n$, $||\boldsymbol{v}||_2 \leq \sqrt{n}||\boldsymbol{v}||_\infty$. To show this, we know that each of $|v_i|$ is less than or equal to $||\boldsymbol{v}||_\infty$, by definition of the maximum. The same can be said for $|v_i|^2$, vs $||\boldsymbol{v}||_\infty^2$. Taking square roots, we have the following;

$$||\boldsymbol{v}||_2^2 = \sum_{i=1}^{n} |v_i|^2 \leq n||\boldsymbol{v}||_\infty^2 \Rightarrow ||\boldsymbol{v}||_2 \leq \sqrt{n}||\boldsymbol{v}||_\infty$$

To show this holds similarly for $||\boldsymbol{v}||_1 \leq \sqrt{n}||\boldsymbol{v}||_2$, we employ the Cauchy-Schwarz inequality, which states $|\boldsymbol{x} \cdot \boldsymbol{y}| \leq ||\boldsymbol{x}||_2||\boldsymbol{y}||_2$. The Cauchy-Schwarz inequality uses the fact that $\boldsymbol{x} \cdot \boldsymbol{y} = ||\boldsymbol{x}||_2||\boldsymbol{y}||_2 \cos\theta$. To do this, we need to define a sign function $\text{sgn} : \mathbb{R} \to \{1, -1\}$ as follows;

$$\text{sgn } x = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0 \end{cases}$$

We also need to craft a vector $\boldsymbol{w}$, as follows;

$$\begin{aligned}
w_i &= \frac{\text{sgn } v_i}{\sqrt{n}} && 1 \leq i \leq n \\
\boldsymbol{v} \cdot \boldsymbol{w} &= \sum_{i=1}^{n} v_i w_i \\
&= \sum_{i=1}^{n} \frac{v_i \cdot \text{sgn } v_i}{\sqrt{n}} && \text{product of same sign becomes positive} \\
&= \sum_{i=1}^{n} \frac{|v_i|}{\sqrt{n}} \\
&= \sqrt{n} \sum_{i=1}^{n} |v_i| \\
&= \sqrt{n}||\boldsymbol{v}||_1 \\
||\boldsymbol{w}||_2 &= \sum_{i=1}^{n} \frac{\pm 1}{\sqrt{n}}^2 \\
&= \sum_{i=1}^{n} \frac{1}{n} \\
&= 1
\end{aligned}$$

By Cauchy-Schwarz, we get;

$$\sqrt{n}||\boldsymbol{v}||_1 = |\boldsymbol{v} \cdot \boldsymbol{w}| \leq ||\boldsymbol{v}||_2||\boldsymbol{w}||_2 = ||\boldsymbol{v}||_2$$

16th January 2020

Note that this recording has **no audio**, and therefore will just be the board transcribed. I honestly have no idea what he was doing in this lecture, it seems to just jump from topic to topic.

Equivalence of Norms?

Take any two norms on $\mathbb{R}^n$; $|| \; ||_a$, and $|| \; ||_b$.

$$\exists r, s \in \mathbb{R}^+ \; \forall \boldsymbol{v} \in \mathbb{R}^n \; [r||\boldsymbol{v}||_b \leq ||\boldsymbol{v}||_a \leq s||\boldsymbol{v}||_b]$$

This means that norms in finite dimensional vector spaces are equivalent (no idea why, look it up).

Convergence of Vector Sequences

$(\boldsymbol{r}_n)$ is a sequence of vectors, and $(a_{i,j})$ is the i, j^{th} entry of $\boldsymbol{A}$. For a vector $\boldsymbol{v}^{(m)} \in \mathbb{R}^n$, where $m = 0, 1, 2, \ldots$

$$\boldsymbol{v}^{(m)} = \begin{bmatrix} v_1^{(m)} \\ v_2^{(m)} \\ \vdots \\ v_n^{(m)} \end{bmatrix}$$

For a vector sequence $\boldsymbol{v}^{(m)}$ to converge to some vector $\boldsymbol{v} \in \mathbb{R}^n$, the following must hold;

$$\boldsymbol{v}^{(m)} \to \boldsymbol{v} \in \mathbb{R}^n \Leftrightarrow \lim_{m \to \infty} ||\boldsymbol{v}^{(m)} - \boldsymbol{v}|| \to 0$$

This is componentwise convergence, such that $\forall i \in [1, n] \; [v_i^{(m)} \to v_i]$.

Matrix Norms

Vectors are a type of matrix. For a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the following properties of its norms must hold, where $|| \; || : \mathbb{R}^{m \times n} \to \mathbb{R}_{\geq 0}$;

(i) $||\boldsymbol{A}|| > 0$ given that $\boldsymbol{A} \neq \boldsymbol{0}$

(ii) $||\lambda \boldsymbol{A}|| = |\lambda| ||\boldsymbol{A}||$

(iii) $||\boldsymbol{A} + \boldsymbol{B}|| \leq ||\boldsymbol{A}|| + ||\boldsymbol{B}||$

(iv) $||\boldsymbol{B}\boldsymbol{A}|| \leq ||\boldsymbol{B}|| ||\boldsymbol{A}||$

$$\begin{array}{ccccc} \boldsymbol{v} \in \mathbb{R}^n & \longrightarrow & \boxed{\boldsymbol{A}} & \longrightarrow & \boldsymbol{A}\boldsymbol{v} \in \mathbb{R}^m \\ || \; ||_a & & & & || \; ||_b \end{array}$$

$$||\boldsymbol{A}\boldsymbol{v}||_b \leq ||\boldsymbol{A}|| ||\boldsymbol{v}||_a$$

For the following example, take $(a_{i,j}) = \boldsymbol{A} \in \mathbb{R}^{m \times n}$;

$$a_j = \begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{m,j} \end{bmatrix} \qquad \text{the } j^{\text{th}} \text{ column of } \boldsymbol{A}$$

$$a^i = \begin{bmatrix} a_{i,1} & a_{i,2} & \cdots & a_{i,n} \end{bmatrix} \qquad \text{the } i^{\text{th}} \text{ row of } \boldsymbol{A}$$

We have the following norms on matrices;

$$\begin{aligned}
||\boldsymbol{A}||_1 &= \max\{||a_j||_1 : 1 \leq j \leq n\} \\
||\boldsymbol{A}||_\infty &= \max\{||(a^i)^\top||_1 : 1 \leq i \leq m\} \\
||\boldsymbol{A}||_F &= \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i,j}|^2} && \text{Frobenius norm} \\
||\boldsymbol{A}||_2 &= \text{largest singular value of } \boldsymbol{A} \\
\text{let } \boldsymbol{A} &= \begin{bmatrix} 2 & 3 & 1 & 4 \\ 1 & 3 & -1 & 5 \\ \sqrt{2} & 0 & -2 & 2 \end{bmatrix} \\
||\boldsymbol{A}||_1 &= \max\{3+\sqrt{2}, 6, 4, 11\} \\
&= 11 \\
||\boldsymbol{A}||_\infty &= \max\{10, 10, 4+\sqrt{2}\} \\
&= 10 \\
||\boldsymbol{A}||_F &= \sqrt{4+9+1+16+1+9+1+25+2+0+4+4} \\
&= 2\sqrt{19}
\end{aligned}$$

Let there be two vector norms, $||\ ||_a$ on $\mathbb{R}^n$ and $||\ ||_b$ on $\mathbb{R}$. If $||\ ||$ (matrix norm) satisfies

$$\forall \boldsymbol{A} \in \mathbb{R}^{m\times n}, x \in \mathbb{R}^n\ [||\boldsymbol{Ax}||_b \leq ||\boldsymbol{A}||||\boldsymbol{x}||_a]$$

then $||\ ||$ is **consistent** with $||\ ||_a$ and $||\ ||_b$. Additionally if $a = b$, then $||\ ||$ is **compatible** with $||\ ||_a$. This gives us the following propositions;

- $||\ ||_1$ (matrix norm) is compatible with $||\ ||_1$ (vector norm)
- $||\ ||_2$ (matrix norm) is compatible with $||\ ||_2$ (vector norm)
- $||\ ||_\infty$ (matrix norm) is compatible with $||\ ||_\infty$ (vector norm)
- $||\ ||_F$ (matrix norm) is compatible with $||\ ||_2$ (vector norm) $\Rightarrow ||\boldsymbol{Ax}||_2 \leq ||\boldsymbol{A}||_F||\boldsymbol{x}||_2$

Given a vector norm $||\ ||$ on $\mathbb{R}^n$ then the matrix norm $||\ ||$ subordinate to vector norm $||\ ||$ is defined by

$$||\boldsymbol{A}|| = \max\{||\boldsymbol{Ax}|| : ||\boldsymbol{x}|| \leq 1\} = \max\{||\boldsymbol{Ax}|| : ||\boldsymbol{x}|| = 1\} = \max\{||\boldsymbol{A}\tfrac{\boldsymbol{x}}{||\boldsymbol{x}||}|| : \boldsymbol{x} \neq \boldsymbol{0}\}$$

Using this, we can prove property (iii) (see above). We claim that $||\boldsymbol{A}+\boldsymbol{B}|| \leq ||\boldsymbol{A}|| + ||\boldsymbol{B}||$ for matrix norm $||\ ||$ subordinate to vector norm $||\ ||$.

$$\begin{aligned}
||\boldsymbol{A}+\boldsymbol{B}|| &= \max\{||(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\} \\
&= \max\{||\boldsymbol{Ax}+\boldsymbol{Bx}|| : ||\boldsymbol{x}|| \leq 1\} \\
&\leq \max\{||\boldsymbol{Ax}|| + ||\boldsymbol{Bx}|| : ||\boldsymbol{x}|| \leq 1\} && \text{triangle inequality for } ||\ || \\
&\leq \max\{||\boldsymbol{Ax}|| : ||\boldsymbol{x}|| \leq 1\} + \max\{||\boldsymbol{Bx}|| : ||\boldsymbol{x}|| \leq 1\} && \text{maximise independently} \\
&= ||\boldsymbol{A}|| + ||\boldsymbol{B}|| && \blacksquare
\end{aligned}$$

We are also able to prove property (iv), which we claim to be $||\boldsymbol{BA}|| \leq ||\boldsymbol{B}||||\boldsymbol{A}||$.

$$\begin{aligned}
||\boldsymbol{BA}|| &= \max\{||\boldsymbol{BAx}|| : ||\boldsymbol{x}|| \leq 1\} \\
||\boldsymbol{BAx}|| &= ||\boldsymbol{B}(\boldsymbol{Ax})|| \\
||\boldsymbol{Ax}|| &\leq ||\boldsymbol{A}||||\boldsymbol{x}|| && \text{matrix norm subordinate to vector norm}
\end{aligned}$$

To show the line above, we consider the two cases, $\boldsymbol{x} \neq \mathbf{0}$ and $\boldsymbol{x} = \mathbf{0}$. If $\boldsymbol{x} = \mathbf{0}$, no work needs to be done, as it is trivial. I have no idea why this works, but he wrote it.

$$\begin{aligned} \text{show } \left|\left|\boldsymbol{A}\frac{\boldsymbol{x}}{||\boldsymbol{x}||}\right|\right| &\leq ||\boldsymbol{A}|| \\ \text{but } \left|\left|\frac{\boldsymbol{x}}{||\boldsymbol{x}||}\right|\right| &= 1 && \Rightarrow \\ ||\boldsymbol{A}\boldsymbol{x}|| &\leq ||\boldsymbol{A}||||\boldsymbol{x}|| \end{aligned}$$

Continuing on, we have

$$\begin{aligned} ||\boldsymbol{B}\boldsymbol{A}\boldsymbol{x}|| &= ||\boldsymbol{B}(\boldsymbol{A}\boldsymbol{x})|| \\ &\leq ||\boldsymbol{B}||||\boldsymbol{A}\boldsymbol{x}|| \\ &\leq ||\boldsymbol{B}||||\boldsymbol{A}||||\boldsymbol{x}|| \\ ||\boldsymbol{B}\boldsymbol{A}|| &= \max\{||\boldsymbol{B}\boldsymbol{A}\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\} \\ &\leq \max\{||\boldsymbol{B}||||\boldsymbol{A}||||\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\} \\ &= ||\boldsymbol{B}||||\boldsymbol{A}|| \max\{||\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\} && \text{obviously 1, as bounded on top} \\ &= ||\boldsymbol{B}||||\boldsymbol{A}|| && \blacksquare \end{aligned}$$

Complex Vectors

$$\mathbb{C}^n = \left\{ \boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} : v_i \in \mathbb{C} \right\}$$

$$\begin{aligned} z \in \mathbb{C} &= a + ib && a \in \mathbb{R}, b \in \mathbb{R}, i = \sqrt{-1} \\ z^* &= a - ib \\ |z| &= \sqrt{a^2 + b^2} \end{aligned}$$

Take a linear map $f : \mathbb{C}^n \to \mathbb{C}^m$, the same properties hold;

$$f(a\boldsymbol{v} + b\boldsymbol{w}) = af(\boldsymbol{v}) + bf(\boldsymbol{w}) \qquad a, b \in \mathbb{C}$$

We also want to define something similar to the dot product in $\mathbb{R}^n$;

$$\begin{aligned} \boldsymbol{v} \cdot \boldsymbol{w} &= ||\boldsymbol{v}||_2||\boldsymbol{w}||_2 \cos\theta_{\boldsymbol{v},\boldsymbol{w}} \\ \boldsymbol{v} \cdot \boldsymbol{v} &= \sqrt{||\boldsymbol{v}||_2^2} \\ &= ||\boldsymbol{v}||_2 \\ <\boldsymbol{v}, \boldsymbol{w}> &= \sum_{i=1}^{n} v_i^* w_i \\ <\boldsymbol{v}, \boldsymbol{v}> &= \sum_{i=1}^{n} v_i^* v_i \\ &= \sum_{i=1}^{n} |v_i|^2 \end{aligned}$$

The standard basis in $\mathbb{R}^n$ is defined as $(\boldsymbol{e_1}, \ldots, \boldsymbol{e_n})$ where

$$\boldsymbol{e_j} = \begin{array}{c} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \end{array} \begin{array}{c} 1^{\text{st}} \\ 2^{\text{nd}} \\ \vdots \\ j^{\text{th}} \\ \vdots \\ n^{\text{th}} \end{array}$$

For any vector $\boldsymbol{v} \in \mathbb{C}^n$, it can be written in the standard basis as such;

$$\begin{aligned}\boldsymbol{v} &= \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \\ &= v_1\boldsymbol{e_1} + \cdots + v_n\boldsymbol{e_n}\end{aligned}$$

Basis Change, Again

Let the linear map $f : \mathbb{R}^n \to \mathbb{R}^m$.

$$\boldsymbol{B} = (\boldsymbol{b_1}, \dots, \boldsymbol{b_n}) \qquad \text{an ordered basis of } \mathbb{R}^n$$
$$\boldsymbol{D} = (\boldsymbol{d_1}, \dots, \boldsymbol{d_m}) \qquad \text{an ordered basis of } \mathbb{R}^m$$

Find the matrix $\boldsymbol{A}$ ($\boldsymbol{A} := f_{\boldsymbol{DB}}$)representing f with respect to (?) $\boldsymbol{B}$ and $\boldsymbol{D}$.

$$\begin{aligned}\boldsymbol{v} &= \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} && \text{coordinates of a point } \boldsymbol{p} \in \mathbb{R}^n \\ \boldsymbol{p} &= \sum_{j=1}^{n} v_j\boldsymbol{b_j} \\ &\boldsymbol{Av} && \text{should be coordinate of } f(\boldsymbol{p}) \in \mathbb{R}^m \\ f(\boldsymbol{p}) &= \sum_{i=1}^{m} (\boldsymbol{Av})_i\boldsymbol{d_i} \\ f(\boldsymbol{b_j}) \in \mathbb{R}^m &= \sum_{i=1}^{m} a_{i,j}\boldsymbol{d_i} && 1 \leq j \leq n\end{aligned}$$

Take $\boldsymbol{B} \in \mathbb{R}^{m\times n}$, where $\boldsymbol{B} = [\boldsymbol{b_1}, \boldsymbol{b_2}, \dots, \boldsymbol{b_n}]$, and $e_1, \dots, \boldsymbol{e_m}$ being the standard basis (see previous).

$$\begin{aligned}\boldsymbol{Be_j} &= \boldsymbol{B}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j^{\text{th}} \\ &= b_{1,j}\boldsymbol{e_1} + \dots + b_{m,j}\boldsymbol{e_m}\end{aligned}$$

Suppose $m = n$ and also $f = \text{id}$ (identity), but $\boldsymbol{B}$ and $\boldsymbol{D}$ are different. The matrix $(\text{id})_{\boldsymbol{DB}}$ represents a change of basis from $\boldsymbol{B}$ to $\boldsymbol{D}$. The point $\boldsymbol{p} \in \mathbb{R}^n$ has coordinates $\boldsymbol{x} \in \mathbb{R}^n$ with respect to $\boldsymbol{B}$, and $\boldsymbol{y} \in \mathbb{R}^n$ with respect to $\boldsymbol{D}$.

$$\boldsymbol{Dy} = \boldsymbol{p} = \boldsymbol{Bx} = x_1\boldsymbol{b_1} + \cdots + x_n\boldsymbol{b_n}$$

From this we gather $\boldsymbol{Dy} = \boldsymbol{Bx}$, therefore $\boldsymbol{y} = \boldsymbol{D}^{-1}\boldsymbol{Bx} = (\text{id})_{\boldsymbol{DB}}\boldsymbol{x}$, which means that

$$(\text{id})_{\boldsymbol{DB}} = \boldsymbol{D}^{-1}\boldsymbol{B}$$

Some stuff on functions between bases?

$$\begin{array}{ccccc} \mathbb{R}^n & \xrightarrow{f} & \mathbb{R}^m & \xrightarrow{g} & \mathbb{R}^k \\ \boldsymbol{B} & & \boldsymbol{C} & & \boldsymbol{D} \end{array}$$

$$\mathbb{R}^n_{\boldsymbol{B}} \xrightarrow{f_{\boldsymbol{CB}}} \mathbb{R}^m_{\boldsymbol{C}} \xrightarrow{g_{\boldsymbol{DC}}} \mathbb{R}^k_{\boldsymbol{D}}$$

$$g_{\boldsymbol{DC}}f_{\boldsymbol{CB}} = (g \circ f)_{\boldsymbol{DB}}$$

This then goes into change of basis, but see last year's **CO145**.

22nd January 2020

Tutorial Question

$f : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map, and $\boldsymbol{E} = (\boldsymbol{e_1}, \boldsymbol{e_2})$ is an ordered basis.

$$\begin{aligned} f(\boldsymbol{e_1}) &= 5\boldsymbol{e_1} - 6\boldsymbol{e_2} \\ f(\boldsymbol{e_2}) &= 3\boldsymbol{e_1} + \boldsymbol{e_2} \end{aligned}$$

We only care about what the linear map does to the ordered basis, as anything else can be done by linearity.

(i) Find $f_{\boldsymbol{EE}}$, the matrix representation of f in $\boldsymbol{E}$ - note that this has the same input space as the output space, but it can be different.

The first column can be done by reading the entry for for $f(\boldsymbol{e_1})$, and similarly for the second column as follows;

$$f_{\boldsymbol{EE}} = \begin{bmatrix} 5 & 3 \\ -6 & 1 \end{bmatrix}$$

(ii) If we have another ordered basis $\boldsymbol{D} = (\boldsymbol{d_1}, \boldsymbol{d_2})$, where $\boldsymbol{d_1} = \boldsymbol{e_1} - \boldsymbol{e_2}$ and $\boldsymbol{d_2} = \boldsymbol{e_1} + \boldsymbol{e_2}$, find $f_{\boldsymbol{DD}}$.

$$\begin{tikzcd} \mathbb{R}^2_{\boldsymbol{E}} \arrow[r, "f_{\boldsymbol{EE}}"] & \mathbb{R}^2_{\boldsymbol{E}} \arrow[d, "\boldsymbol{I_{DE}} = (\boldsymbol{I_{ED}})^{-1}"] \\ \mathbb{R}^2_{\boldsymbol{D}} \arrow[u, "\boldsymbol{I_{ED}}"] \arrow[r, "f_{\boldsymbol{DD}}"] & \mathbb{R}^2_{\boldsymbol{D}} \end{tikzcd}$$

$\boldsymbol{I_{ED}}$ can easily be obtained by reading the entries for $\boldsymbol{d_1}$ for the first column, and similarly for $\boldsymbol{d_2}$ in the second column;

$$\begin{aligned} \boldsymbol{I_{ED}} &= \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ \boldsymbol{I_{DE}} &= (\boldsymbol{I_{ED}})^{-1} \\ &= \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \\ f_{\boldsymbol{DD}} &= \boldsymbol{I_{DE}} f_{\boldsymbol{EE}} \boldsymbol{I_{ED}} \\ &= \frac{1}{2}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 5 & 3 \\ -6 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} 1 & 13 \\ -5 & 3 \end{bmatrix} \end{aligned}$$

Eigenvalues + Generalised Eigenvectors

Working with a matrix $\boldsymbol{A} \in \mathbb{C}^{m \times m}$. For an eigenvector $\boldsymbol{v} \in \mathbb{C}^m \setminus \{\boldsymbol{0}\}$, and an eigenvalue $\lambda \in \mathbb{C}$, $\boldsymbol{Av} = \lambda\boldsymbol{v} \Rightarrow (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0} \Rightarrow |\boldsymbol{A} - \lambda\boldsymbol{I}| = 0 \Rightarrow P_{\boldsymbol{A}}(\lambda) = 0$ (characteristic polynomial). This complex polynomial will be of degree m, and it will have precisely m roots (including multiplicity). Suppose $P_{\boldsymbol{A}}(\lambda) = 0$, then $(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$ has a solution where $\boldsymbol{v} \neq \boldsymbol{0}$.

Assume we have $\lambda_1, \dots, \lambda_t$ distinct eigenvalues, meaning that $P_{\boldsymbol{A}}(\lambda_i) = 0$ for $1 \leq i \leq t$. This means we can write the characteristic polynomial as;

$$P_{\boldsymbol{A}}(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \dots (\lambda - \lambda_t)^{m_t}$$

Where m_i is the **algebraic multiplicity** of λ_i. $m_i \in \mathbb{N}$, and also $1 \leq m_i \leq m$, as it must not exceed the dimension of the matrix. On the other hand, the **geometric multiplicity** of λ_i is ℓ_i, which is the **nullity** of $(\boldsymbol{A} - \lambda_i \boldsymbol{I})$. The nullity is the dimension of the kernel / null-space. $1 \leq \ell_i \leq m_i$, as we already have at least one non-zero solution from $(\boldsymbol{A} - \lambda_i \boldsymbol{I})\boldsymbol{v} = \boldsymbol{0}$.

In the nice case, we have $\ell_i = m_i$, for $1 \leq i \leq t$, which means the matrix is diagonalisable. We have m_i linearly independent vectors $(\boldsymbol{v_{i,1}}, \boldsymbol{v_{i,2}, \dots, v_{i,m_i}})$ which satisfy

$$\boldsymbol{A}\boldsymbol{v_{i,j}} = \lambda_i \boldsymbol{v_{i,j}} \text{ for } 1 \leq j \leq m_i$$

If we take these eigenvectors as an ordered basis;

$$\boldsymbol{B} = [\boldsymbol{v_{1,1}}, \boldsymbol{v_{1,2}}, \dots, \boldsymbol{v_{1,m_1}}, \boldsymbol{v_{2,1}}, \boldsymbol{v_{2,2}}, \dots, \boldsymbol{v_{2,m_2}}, \dots \boldsymbol{v_{t,1}}, \boldsymbol{v_{t,2}}, \dots, \boldsymbol{v_{t,m_t}}]$$

We also want to note that $\sum_{i=1}^{t} m_i = m$, as that is the degree of the characteristic polynomial. Multiplying the basis by the original matrix, we get;

$$\boldsymbol{AB} = [\lambda_1\boldsymbol{v_{1,1}}, \lambda_1\boldsymbol{v_{1,2}}, \dots, \lambda_1\boldsymbol{v_{1,m_1}}, \lambda_2\boldsymbol{v_{2,1}}, \lambda_2\boldsymbol{v_{2,2}}, \dots, \lambda_2\boldsymbol{v_{2,m_2}}, \dots \lambda_t\boldsymbol{v_{t,1}}, \lambda_t\boldsymbol{v_{t,2}}, \dots, \lambda_t\boldsymbol{v_{t,m_t}}]$$

Since all the columns of $\boldsymbol{B}$ are linearly independently, by our definition, the inverse $\boldsymbol{B}^{-1}$ exists. Therefore, we can write

$$\boldsymbol{B}^{-1}\boldsymbol{AB} = \begin{bmatrix} \lambda_1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & \lambda_1 & & & & & & \\ & & & \lambda_2 & & & & & \\ & & & & \ddots & & & & \\ & & & & & \lambda_2 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \lambda_t & \\ & & & & & & & & \ddots & \\ & & & & & & & & & \lambda_t \end{bmatrix} \text{(everything else is 0)}$$

Which has m_1 instances of λ_1, followed by m_2 instances of λ_2, and so on, until m_t instances of λ_t.

Example for $\ell_i = m_i$

$$\begin{aligned} \boldsymbol{A} &= \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix} \\ |\boldsymbol{A} - \lambda\boldsymbol{I}| &= (\lambda - 3)^2(\lambda - 5) \\ \lambda_1 &= 3 && \text{two linearly independently eigenvectors} \\ \boldsymbol{v_{1,1}} &= \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \\ \boldsymbol{v_{1,2}} &= \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \\ \lambda_2 &= 5 \\ \boldsymbol{v_{2,1}} &= \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
\boldsymbol{B} &= \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix} \\
\boldsymbol{B}^{-1} &= \frac{1}{2}\begin{bmatrix} -2 & 2 & -2 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \\
\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B} &= \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}
\end{aligned}
$$

Trivial Example for $\ell_i < m_i$

$$
\begin{aligned}
\boldsymbol{A} &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\
|\boldsymbol{A} - \lambda \boldsymbol{I}| &= \lambda^2 \\
\lambda_1 &= 0 \\
\boldsymbol{v_1} &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} && \text{only solution, hence } \ell_1 = 1 < 2 = m_1 \\
(\boldsymbol{A} - 0\boldsymbol{I})\begin{bmatrix} 0 \\ 1 \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} 1 \\ 0 \end{bmatrix}
\end{aligned}
$$

although this vector is not mapped to zero, it is mapped to something that **will** be mapped to zero

$$
\begin{aligned}
(\boldsymbol{A} - 0\boldsymbol{I})^2\begin{bmatrix} 0 \\ 1 \end{bmatrix} &= (\boldsymbol{A} - 0\boldsymbol{I})(\boldsymbol{A} - 0\boldsymbol{I})\begin{bmatrix} 0 \\ 1 \end{bmatrix} \\
&= (\boldsymbol{A} - 0\boldsymbol{I})\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \boldsymbol{0}
\end{aligned}
$$

We say $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ is a generalised eigenvector for $\lambda_1 = 0$. A vector which is not mapped by $(\boldsymbol{A} - \lambda\boldsymbol{I})$ to $\boldsymbol{0}$, but is $\boldsymbol{0}$ when iterated once more.

Less Trivial Example

$$
\begin{aligned}
\boldsymbol{A} &= \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \\
|\boldsymbol{A} - \lambda \boldsymbol{I}| &= (1 - \lambda)^3 \\
\lambda_1 &= 1 && m_1 = 3 \\
(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{x} &= \boldsymbol{0} && \Leftrightarrow \\
\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\boldsymbol{x} &= \boldsymbol{0}
\end{aligned}
$$

this has rank 1, and therefore by rank-nullity theorem (rank + nullity = 3), has 2 linearly independent solutions, therefore $\ell_1 = 2 < 3 = m_1$

$$
\boldsymbol{v_1} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}
$$

$$\boldsymbol{v_2} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

to find the generalised eigenvector $\boldsymbol{v_3}$, we want to find some vector that is mapped by $(\boldsymbol{A} - 1\boldsymbol{I})$ to the eigenspace, which is some linear combination of $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$

$$\begin{aligned}
(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{v_3} &= \alpha_1 \boldsymbol{v_1} + \alpha_2 \boldsymbol{v_2} && \Leftrightarrow \\
\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} &= \begin{bmatrix} 0 \\ \alpha_1 \\ -\alpha_1 \end{bmatrix} + \begin{bmatrix} \alpha_2 \\ 0 \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} x_2 + x_3 \\ 0 \\ 0 \end{bmatrix} && \Leftrightarrow \\
\alpha_1 &= 0 \\
x_2 + x_3 &= \alpha_2 && \text{let } x_2 = 0 \Rightarrow x_3 = \alpha_2 = 1 \\
\boldsymbol{v_3} &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \\
\boldsymbol{B} &= [\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}] \\
&= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} \\
\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} && \text{the extra 1 is from } \boldsymbol{v_2}\text{?}
\end{aligned}$$

This is in Jordan Normal Form. For some λ_i eigenvalue, and $\ell_i \leq m_i$, the sum of the sizes of the blocks is m_i, and the number of blocks is ℓ_i.

$$\boldsymbol{J} = \begin{bmatrix} \boldsymbol{J_1} & & \\ & \ddots & \\ & & \boldsymbol{J_p} \end{bmatrix}$$

$$\boldsymbol{J_i} = \begin{bmatrix} \lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}$$

If λ_i is an eigenvalue of $\boldsymbol{A}$ with algebraic multiplicity m_i, then the nullity of $(\boldsymbol{A} - \lambda_i \boldsymbol{I})^{m_i} = m_i$.

Definition: $\boldsymbol{v} \in \mathbb{R}^m$ is a generalised eigenvector for λ_i if $(\boldsymbol{A} - \lambda_i \boldsymbol{I})^{m_i} \boldsymbol{v} = \boldsymbol{0}$. The maximum iterations is m_i, but can be less.