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;

$$w_i = \frac{\text{sgn } v_i}{\sqrt{n}} \qquad\qquad 1 \leq i \leq n$$

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

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 $m = 0, 1, 2, \dots$

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

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

$$\boldsymbol{v} \in \mathbb{R}^n \longrightarrow \boxed{\boldsymbol{A}} \longrightarrow \boldsymbol{A}\boldsymbol{v} \in \mathbb{R}^m$$
$$||\ ||_a \qquad\qquad\qquad ||\ ||_b$$

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

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

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

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

We have the following norms on matrices;

$$||\boldsymbol{A}||_1 = \max\{||a_j||_1 : 1 \leq j \leq n\}$$

$$||\boldsymbol{A}||_\infty = \max\{||(a^i)^\top||_1 : 1 \leq i \leq m\}$$

$$||\boldsymbol{A}||_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{i,j}|^2} \qquad\qquad \text{Frobenius norm}$$

$$||\boldsymbol{A}||_2 = \text{largest singular value of } \boldsymbol{A}$$

$$\text{let } \boldsymbol{A} = \begin{bmatrix} 2 & 3 & 1 & 4 \\ 1 & 3 & -1 & 5 \\ \sqrt{2} & 0 & -2 & 2 \end{bmatrix}$$

$$||\boldsymbol{A}||_1 = \max\{3+\sqrt{2}, 6, 4, 11\}$$
$$= 11$$

$$||\boldsymbol{A}||_\infty = \max\{10, 10, 4+\sqrt{2}\}$$
$$= 10$$

$$||\boldsymbol{A}||_F = \sqrt{4+9+1+16+1+9+1+25+2+0+4+4}$$
$$= 2\sqrt{19}$$

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

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

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) $\Rightarrow ||\boldsymbol{A}\boldsymbol{x}||_2 \leq ||\boldsymbol{A}||_F||\boldsymbol{x}||_2$

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

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

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 $|| \ ||$.

$$||\boldsymbol{A}+\boldsymbol{B}|| = \max\{||(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\}$$
$$= \max\{||\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\}$$
$$\leq \max\{||\boldsymbol{A}\boldsymbol{x}|| + ||\boldsymbol{B}\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\} \qquad \text{triangle inequality for } || \ ||$$
$$\leq \max\{||\boldsymbol{A}\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\} + \max\{||\boldsymbol{B}\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\} \qquad \text{maximise independently}$$
$$= ||\boldsymbol{A}|| + ||\boldsymbol{B}|| \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \blacksquare$$

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

$$||\boldsymbol{B}\boldsymbol{A}|| = \max\{||\boldsymbol{B}\boldsymbol{A}\boldsymbol{x}|| : ||\boldsymbol{x}|| \leq 1\}$$
$$||\boldsymbol{B}\boldsymbol{A}\boldsymbol{x}|| = ||\boldsymbol{B}(\boldsymbol{A}\boldsymbol{x})||$$
$$||\boldsymbol{A}\boldsymbol{x}|| \leq ||\boldsymbol{A}||||\boldsymbol{x}|| \qquad\qquad \text{matrix norm subordinate to vector norm}$$

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

$$\text{show } \left|\left|\boldsymbol{A}\frac{\boldsymbol{x}}{||\boldsymbol{x}||}\right|\right| \leq ||\boldsymbol{A}||$$

$$\text{but } \left|\left|\frac{\boldsymbol{x}}{||\boldsymbol{x}||}\right|\right| = 1 \qquad\qquad\qquad \Rightarrow$$

$$||\boldsymbol{A}\boldsymbol{x}|| \leq ||\boldsymbol{A}||\,||\boldsymbol{x}||$$

Continuing on, we have

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

Complex Vectors

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

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

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

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

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

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

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

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

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

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

$$= v_1 \boldsymbol{e_1} + \cdots + v_n \boldsymbol{e_n}$$

Basis Change, Again

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

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

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

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} \qquad\qquad \text{coordinates of a point } \boldsymbol{p} \in \mathbb{R}^n$$

$$\boldsymbol{p} = \sum_{j=1}^{n} v_j \boldsymbol{b_j}$$

$$\boldsymbol{A}\boldsymbol{v} \qquad\qquad \text{should be coordinate of } f(\boldsymbol{p}) \in \mathbb{R}^m$$

$$f(\boldsymbol{p}) = \sum_{i=1}^{m} (\boldsymbol{A}\boldsymbol{v})_i \boldsymbol{d_i}$$

$$f(\boldsymbol{b_j}) \in \mathbb{R}^m = \sum_{i=1}^{m} a_{i,j} \boldsymbol{d_i} \qquad\qquad 1 \le j \le n$$

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

$$\boldsymbol{B}\boldsymbol{e_j} = \boldsymbol{B} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \leftarrow j^{\text{th}}$$

$$= b_{1,j} \boldsymbol{e_1} + \ldots + b_{m,j} \boldsymbol{e_m}$$

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 $\boldsymbol{D}$.

$$\boldsymbol{D}\boldsymbol{y} = \boldsymbol{p} = \boldsymbol{B}\boldsymbol{x} = x_1 \boldsymbol{b_1} + \cdots + x_n \boldsymbol{b_n}$$

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

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

Some stuff on functions between bases?

$$\mathbb{R}^n \xrightarrow{\ f\ } \mathbb{R}^m \xrightarrow{\ g\ } \mathbb{R}^k$$
$$\boldsymbol{B} \qquad\quad \boldsymbol{C} \qquad\quad \boldsymbol{D}$$

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

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

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.

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

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 $f_{\boldsymbol{DD}}$.

$$
\begin{array}{ccc}
\mathbb{R}^2_{\boldsymbol{E}} & \xrightarrow{f_{\boldsymbol{EE}}} & \mathbb{R}^2_{\boldsymbol{E}} \\
\big\uparrow {\boldsymbol{I}_{\boldsymbol{ED}}} & & \big\downarrow {\boldsymbol{I}_{\boldsymbol{DE}} = (\boldsymbol{I}_{\boldsymbol{ED}})^{-1}} \\
\mathbb{R}^2_{\boldsymbol{D}} & \xrightarrow{f_{\boldsymbol{DD}}} & \mathbb{R}^2_{\boldsymbol{D}}
\end{array}
$$

$\boldsymbol{I}_{\boldsymbol{ED}}$ 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;

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

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 $\lambda \in \mathbb{C}$, $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v} \Rightarrow (\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{v} = \boldsymbol{0} \Rightarrow |\boldsymbol{A} - \lambda\boldsymbol{I}| = 0 \Rightarrow P_{\boldsymbol{A}}(\lambda) = 0$ (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 $\lambda_1, \ldots, \lambda_t$ 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;

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

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 $(\boldsymbol{v_{i,1}}, \boldsymbol{v_{i,2}}, \dots, \boldsymbol{v_{i,m_i}})$ 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;

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

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

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

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 $\ell_i = m_i$

$$\boldsymbol{A} = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix}$$

$$|\boldsymbol{A} - \lambda\boldsymbol{I}| = (\lambda - 3)^2(\lambda - 5)$$

$$\lambda_1 = 3 \qquad\qquad\qquad\qquad \text{two linearly independently eigenvectors}$$

$$\boldsymbol{v_{1,1}} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

$$\boldsymbol{v_{1,2}} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

$$\lambda_2 = 5$$

$$\boldsymbol{v_{2,1}} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & 2 \\ 0 & 1 & 1 \end{bmatrix}$$

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

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

Trivial Example for $\ell_i < m_i$

$$\boldsymbol{A} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$|\boldsymbol{A} - \lambda\boldsymbol{I}| = \lambda^2$$

$$\lambda_1 = 0$$

$$\boldsymbol{v_1} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad\qquad \text{only solution, hence } \ell_1 = 1 < 2 = m_1$$

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

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

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

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

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

$$= \boldsymbol{0}$$

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

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$|\boldsymbol{A} - \lambda\boldsymbol{I}| = (1 - \lambda)^3$$

$$\lambda_1 = 1 \qquad\qquad m_1 = 3$$

$$(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{x} = \boldsymbol{0} \qquad\qquad \Leftrightarrow$$

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{0}$$

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 $\boldsymbol{v_2}$

$$(\boldsymbol{A} - \boldsymbol{I})\boldsymbol{v_3} = \alpha_1 \boldsymbol{v_1} + \alpha_2 \boldsymbol{v_2} \qquad\qquad \Leftrightarrow$$

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ \alpha_1 \\ -\alpha_1 \end{bmatrix} + \begin{bmatrix} \alpha_2 \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} x_2 + x_3 \\ 0 \\ 0 \end{bmatrix} \qquad\qquad \Leftrightarrow$$

$$\alpha_1 = 0$$

$$x_2 + x_3 = \alpha_2 \qquad\qquad\qquad \text{let } x_2 = 0 \Rightarrow x_3 = \alpha_2 = 1$$

$$\boldsymbol{v_3} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{B} = [\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}]$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$$\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \qquad\qquad \text{the extra 1 is from } \boldsymbol{v_2}?$$

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

23rd January 2020

In general;

- $\forall j \; [\ell_j = m_j] \Rightarrow \boldsymbol{A}$ is diagonalisable
- $\exists j \; [\ell_j < m_j] \Rightarrow \boldsymbol{A}$ is diagonalisable or **almost** diagonalisable

The following is a very poor explanation of what the form looks like, because I don't know how to draw blocks in LaTeX. See Panopto at timestamp 13:17 for a proper drawing.

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{B}_{\lambda_1} & & \\ & \ddots & \\ & & \boldsymbol{B}_{\lambda_t} \end{bmatrix}$$

$$\boldsymbol{B}_j = \begin{bmatrix} \boldsymbol{J}_j & & \\ & \ddots & \\ & & \boldsymbol{J}_j \end{bmatrix} \qquad\qquad \text{all } \boldsymbol{J}_j \text{ can be of different sizes, } \boldsymbol{B}_j \in \mathbb{R}^{m_j \times m_j}$$

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

Spectral Decomposition of Symmetric Matrices

In almost all areas of computer science, symmetric matrices are the dominant paradigm. All the eigenvalues of real symmetric matrices are real, and are always diagonalisable. A symmetric matrix is always a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, which also satisfies $\boldsymbol{A}^\top = \boldsymbol{A}$, meaning it is symmetric along the diagonal (the elements above the diagonal determine the elements below, and vice versa). We are to prove the following properties;

(i) All eigenvalues of $\boldsymbol{A}$ are real.

First assume that $\boldsymbol{v} \in \mathbb{C}^n$, as we don't yet know that it is real, therefore we can have the case $\lambda \in \mathbb{C}$.

$$\begin{aligned} \boldsymbol{Av} &= \lambda \boldsymbol{v} & \Rightarrow \\ \boldsymbol{v}^{*\top} \boldsymbol{Av} &= \lambda \boldsymbol{v}^{*\top} \boldsymbol{v} \\ &= \lambda ||\boldsymbol{v}||^2 & (1) \\ \boldsymbol{v}^{*\top} \boldsymbol{Av} &= \left(\boldsymbol{v}^{*\top} \boldsymbol{A} \right) \boldsymbol{v} \\ &= \left((\boldsymbol{Av})^\top \right)^* \boldsymbol{v} \end{aligned}$$

briefly proving the above;

$$\begin{aligned} \left((\boldsymbol{Av})^\top \right)^* &= \left(\boldsymbol{v}^\top \boldsymbol{A}^\top \right)^* & \text{taking the transpose inside the bracket} \\ &= \left(\boldsymbol{v}^\top \right)^* \boldsymbol{A}^* & \text{taking the conjugate inside and } \boldsymbol{A}^\top = \boldsymbol{A} \\ &= \left(\boldsymbol{v}^\top \right)^* \boldsymbol{A} & \boldsymbol{A} \text{ is real valued matrix} \end{aligned}$$

continuing on;

$$\begin{aligned} &= \left((\lambda \boldsymbol{v})^\top \right)^* \boldsymbol{v} & \boldsymbol{v} \text{ is an eigenvector} \\ &= \lambda^* \left(\boldsymbol{v}^\top \right)^* \boldsymbol{v} & \text{transposing a scalar does nothing} \\ &= \lambda^* ||\boldsymbol{v}||^2 & (2) \end{aligned}$$

By comparing results, (1) = (2), hence $\lambda^* = \lambda$ (dividing through by $||\boldsymbol{v}||^2$ is allowed, since we know it is non-zero by properties of vector norms, and $\boldsymbol{v} \neq \boldsymbol{0}$ by properties of eigenvectors). As the complex conjugate of λ is the same as λ, we can conclude that $\lambda \in \mathbb{R}$.

(ii) Eigenvectors corresponding to different eigenvalues of $\boldsymbol{A}$ are perpendicular.

Note that with the above result, we know that $\boldsymbol{v} \in \mathbb{R}^n$, as there is no possible way to obtain complex values with a real valued matrix and a real valued eigenvalue.

$$\begin{aligned} \boldsymbol{Av_1} &= \lambda_1 \boldsymbol{v_1} \\ \boldsymbol{Av_2} &= \lambda_2 \boldsymbol{v_2} \\ \lambda_1 &\neq \lambda_2 \\ \boldsymbol{v_2}^\top \boldsymbol{Av_1} &= \lambda_1 \boldsymbol{v_2}^\top \boldsymbol{v_1} \\ &= \lambda_1 (\boldsymbol{v_1} \cdot \boldsymbol{v_2}) & (1) \\ \boldsymbol{v_2}^\top \boldsymbol{Av_1} &= \left(\boldsymbol{v_2}^\top \boldsymbol{A} \right) \boldsymbol{v_1} \\ &= (\boldsymbol{Av_2})^\top \boldsymbol{v_1} & \text{adding a transpose and } \boldsymbol{A}^\top = \boldsymbol{A} \\ &= \lambda_2 \boldsymbol{v_2}^\top \boldsymbol{v_1} & \boldsymbol{v_2} \text{ is an eigenvector} \\ &= \lambda_2 (\boldsymbol{v_1} \cdot \boldsymbol{v_2}) & (2) \end{aligned}$$

Similarly, comparing results (1) and (2), we get

$$\lambda_1(v_1 \cdot v_2) = \lambda_2(v_1 \cdot v_2) \qquad \Leftrightarrow$$
$$\lambda_1(v_1 \cdot v_2) - \lambda_2(v_1 \cdot v_2) = 0$$
$$(\lambda_1 - \lambda_2)(v_1 \cdot v_2) = 0 \qquad \Leftrightarrow$$
$$\lambda_1 = \lambda_2$$
$$\text{or } v_1 \cdot v_2 = 0$$

The former is not possible due to the condition that they are different eigenvalues, therefore $v_1 \cdot v_2 = 0$. As neither are the zero vector, by definition of eigenvectors, $\cos\theta = 0$, therefore they are perpendicular.

(iii) (Tutorial 3) If λ_j is an eigenvalue of A then $\ell_j = m_j$, therefore A is diagonalisable.

Come back to this in two weeks?

To conclude, we get all the eigenvectors and we normalise them (divide by the ℓ_2-norm), and we can assume they are pairwise orthogonal. There is an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbb{A}$; $(v_1, \ldots, v_n)$, where $v_j \in \mathbb{R}^n$. This has the following properties;

- $Av_j = \lambda_j v_j$
- $v_i \perp v_j$, when $i \neq j$ perpendicular
- $\|v_i\| = 1$ for $i = 1, \ldots, n$

Examples

For an example in $\mathbb{R}^{2 \times 2}$;

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$
$$|A - \lambda I| = (1 - \lambda)^2 - 4$$
$$\lambda_1 = 3$$
$$\lambda_2 = -1$$
$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{normalise straight away}$$
$$v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$
$$V = [v_1, v_2] \qquad \text{orthonormal basis of } \mathbb{R}^2$$
$$= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$
$$AV = [Av_1, Av_2]$$
$$= [3v_1, -v_2]$$

For a symmetric matrix, if we have a matrix V representing the orthonormal basis of the eigenvectors, $V^{-1} = V^\top$

$$V^{-1}AV = V^\top AV$$
$$= V^\top [3v_1, -v_2]$$
$$= \begin{bmatrix} v_1^\top \\ v_2^\top \end{bmatrix} [3v_1, -v_2]$$
$$= \begin{bmatrix} 3 & 0 \\ 0 & -1 \end{bmatrix} \qquad \text{dot product with unit length } (v_1 \perp v_2)$$

For an example in $\mathbb{R}^{3\times 3}$;

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 3 & 1 \\ 3 & 1 & 1 \end{bmatrix}$$

now he just shows some intuition on getting an eigenvalue?

$$\begin{aligned}
|\boldsymbol{A} - \lambda\boldsymbol{I}| &= \begin{vmatrix} 1-\lambda & 1 & 3 \\ 1 & 3-\lambda & 1 \\ 3 & 1 & 1-\lambda \end{vmatrix} \\
&= \begin{vmatrix} 5-\lambda & 1 & 3 \\ 5-\lambda & 3-\lambda & 1 \\ 5-\lambda & 1 & 1-\lambda \end{vmatrix} \qquad\qquad \boldsymbol{c_1}' = \boldsymbol{c_1} + \boldsymbol{c_2} + \boldsymbol{c_3} \\
&= 0 \\
\lambda_1 &= 5 \\
\mathrm{trace}(\boldsymbol{A}) &= \lambda_1 + \lambda_2 + \lambda_3 \\
&= 5 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow \\
\lambda_2 + \lambda_3 &= 0 \\
|\boldsymbol{A}| &= \lambda_1\lambda_2\lambda_3 \\
&= -20 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \Rightarrow \\
\lambda_2\lambda_3 &= -4 \\
\lambda_2 &= 2 \\
\lambda_3 &= -2 \\
\boldsymbol{v_1} &= \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\
\boldsymbol{v_2} &= \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \\
\boldsymbol{v_3} &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \\
\boldsymbol{V} &= [\boldsymbol{v_1}, \boldsymbol{v_2}, \boldsymbol{v_3}] \\
\boldsymbol{V}^{-1} &= \boldsymbol{V}^{\top} \\
\boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V} &= \boldsymbol{V}^{\top}\boldsymbol{A}\boldsymbol{V} \\
&= \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \qquad\qquad \text{spectral decomposition of a symmetric matrix}
\end{aligned}$$

Proof of $\boldsymbol{V}^{-1} = \boldsymbol{V}^{\top}$

We define $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ as an **orthogonal matrix** if all columns $\boldsymbol{b_j}$ for $1 \leq j \leq n$ form an orthonormal basis of $\mathbb{R}^n$, such that the two properties hold;

- $\boldsymbol{b_j} \cdot \boldsymbol{b_i} = 0$ when $i \neq j$
- $\boldsymbol{b_j} \cdot \boldsymbol{b_j} = 1$

We want to prove that $\boldsymbol{B}^{-1} = \boldsymbol{B}^\top$:

$$\boldsymbol{B} = [\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}]$$

$$\boldsymbol{B}^\top = \begin{bmatrix} \boldsymbol{b_1}^\top \\ \vdots \\ \boldsymbol{b_n}^\top \end{bmatrix}$$

$$\boldsymbol{B}^\top \boldsymbol{B} = \begin{bmatrix} \boldsymbol{b_1}^\top \\ \vdots \\ \boldsymbol{b_n}^\top \end{bmatrix} [\boldsymbol{b_1}, \ldots, \boldsymbol{b_n}]$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

$$= \boldsymbol{I}_n$$

Preservation of Length

The length of a vector $\boldsymbol{x}$ is preserved under orthogonal transformation;

$$(\boldsymbol{V}\boldsymbol{x}) \cdot (\boldsymbol{V}\boldsymbol{x}) = (\boldsymbol{V}\boldsymbol{x})^\top (\boldsymbol{V}\boldsymbol{x})$$
$$= \boldsymbol{x}^\top \boldsymbol{V}^\top \boldsymbol{V}\boldsymbol{x}$$
$$= \boldsymbol{x}^\top \boldsymbol{V}^{-1} \boldsymbol{V}\boldsymbol{x}$$
$$= \boldsymbol{x}^\top \boldsymbol{x}$$
$$= \boldsymbol{x} \cdot \boldsymbol{x}$$

Geometric Intuition

Take a real symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, such that $\boldsymbol{A}^\top = \boldsymbol{A}$. Assume we have processed it to get the eigenvalues $\lambda_1, \ldots, \lambda_n$, such that $\boldsymbol{V}^\top \boldsymbol{A} \boldsymbol{V} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, and $\boldsymbol{V}$ is an orthonormal matrix formed of the eigenvectors.

Take a vector $\boldsymbol{x}$ from the input space, to $\boldsymbol{A}\boldsymbol{x}$ in the output space, mapped by $\boldsymbol{A}$;

$$
\begin{array}{ccc}
\text{input space} & \boldsymbol{A} = f_{\boldsymbol{EE}} & \text{output space} \\
\boldsymbol{x} & \longmapsto\!\!\!\longrightarrow & \boldsymbol{A}\boldsymbol{x}
\end{array}
$$

$$
\begin{array}{ccc}
\mathbb{R}^n_{\boldsymbol{E}} & \xrightarrow{\ f_{\boldsymbol{EE}} = \boldsymbol{A}\ } & \mathbb{R}^n_{\boldsymbol{E}} \\
{\scriptstyle \boldsymbol{I_{EV}} = \boldsymbol{V}}\big\uparrow & & \big\downarrow{\scriptstyle \boldsymbol{I_{VE}} = (\boldsymbol{I_{ED}})^{-1} = \boldsymbol{V}^{-1}} \\
\mathbb{R}^n_{\boldsymbol{V}} & \xrightarrow{\ f_{\boldsymbol{DD}} = \boldsymbol{V}^{-1}\boldsymbol{A}\boldsymbol{V}\ } & \mathbb{R}^n_{\boldsymbol{V}}
\end{array}
$$

Consider a unit sphere, created by $||\boldsymbol{x}|| = 1$, where $\boldsymbol{x} \in \mathbb{R}^3$. The transformations $\boldsymbol{V}^{-1} = \boldsymbol{V}^\top$ and $\boldsymbol{V}$ are rotations as they preserve length. Take a point $\boldsymbol{x}$ on the surface of this unit sphere, applying $\boldsymbol{V}^{-1}$ to it simply rotates it to another point on the surface of the sphere (going from $\mathbb{R}^3_{\boldsymbol{E}}$ to $\mathbb{R}^3_{\boldsymbol{V}}$) in the input space. Let this point have coordinates $\boldsymbol{u}$ with respect to $\boldsymbol{V}$. Under the diagonal map $\mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, we get the following;

$$u_1 \boldsymbol{v_1} + \cdots + u_n \boldsymbol{v_n} \mapsto \lambda_1 u_1 \boldsymbol{v_1} + \cdots + \lambda_n u_n \boldsymbol{v_n}$$

As $\boldsymbol{u}$ is on the sphere, we can say;

$$||\boldsymbol{u}||_2^2 = u_1^1 + \cdots + u_n^2 = 1$$

Therefore, the mapped point $\boldsymbol{y}$ can be written as $\begin{bmatrix} \lambda_1 u_1 \\ \vdots \\ \lambda_n u_n \end{bmatrix}$, satisfying (in three dimensions);

$$\frac{y_1^2}{\lambda_1^2} + \frac{y_2^2}{\lambda_2^2} + \frac{y_3^2}{\lambda_3^2} = 1$$

This gives a locus of an ellipsoid.

Rank-Nullity Theorem

Take a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$.

$$\operatorname{Im}(\boldsymbol{A}) \subseteq \mathbb{R}^m = \{\boldsymbol{Ax} : \boldsymbol{x} \in \mathbb{R}^n\}$$
$$\dim(\operatorname{Im}(\boldsymbol{A})) = \text{column rank of } \boldsymbol{A}$$
$$\operatorname{Null}(\boldsymbol{A}) \subseteq \mathbb{R}^n = \{\boldsymbol{v} : \boldsymbol{Av} = \boldsymbol{0}\} \qquad\qquad \text{same as kernel}$$
$$\operatorname{rk}(\text{column space}) = \operatorname{rk}(\text{row space})$$
$$\text{row space } \boldsymbol{A} = \text{column space } \boldsymbol{A}^\top$$

For elementary row operations, we have the following property;

(I) they are invertible.

If there is one of the three elementary row operations R that takes $\boldsymbol{A} \rightsquigarrow \boldsymbol{B}$, then $\exists R^{-1}$ that takes $\boldsymbol{B} \rightsquigarrow \boldsymbol{A}$.

If $\boldsymbol{A} \rightsquigarrow \boldsymbol{B}$, then $\boldsymbol{B} = \boldsymbol{M_R A}$. We obtain $\boldsymbol{M_R}$ by applying R to $\boldsymbol{I}_m$, such that $\boldsymbol{I}_m \rightsquigarrow \boldsymbol{M_R}$.

A reduction to row echelon form can be written as follows (with $-$ as a pivot, and $\times$ being any number);

$$\boldsymbol{A} \overset{R_1}{\rightsquigarrow} \cdots \overset{R_t}{\rightsquigarrow} \boldsymbol{C} = \begin{bmatrix} - & \times & \times & \times & \times & \times \\ 0 & - & \times & \times & \times & \times \\ 0 & 0 & 0 & - & \times & \times \\ 0 & 0 & 0 & 0 & - & \times \\ 0 & 0 & 0 & 0 & 0 & - \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

In general it creates a staircase of the pivots, and everything below it is 0. Note that the example above is just a random example, as long as the general structure remains, it's fine.

An ERO doesn't change the row space, as it is operating on the rows. Swapping rows, multiplying by a non-zero number, and adding rows will not change anything. The claim is that while the column space can change, the rank of the column space doesn't change.

$$\boldsymbol{A} = [\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}]$$
$$\boldsymbol{A} \overset{R}{\rightsquigarrow} \boldsymbol{M_R A} = [\boldsymbol{M_R a_1}, \ldots, \boldsymbol{M_R a_n}]$$
$$\boldsymbol{M_{R^{-1}}} = (\boldsymbol{M_R})^{-1} \qquad\qquad \text{therefore } \boldsymbol{M_R} \text{ is invertible}$$

taking some linear combination of the vectors in $\boldsymbol{A}$

$$\lambda_1 \boldsymbol{a_{i_1}} + \cdots + \lambda_1 \boldsymbol{a_{i_t}} = 0 \qquad\qquad\qquad \Leftrightarrow$$
$$\boldsymbol{M_R}(\lambda_1 \boldsymbol{a_{i_1}} + \cdots + \lambda_1 \boldsymbol{a_{i_t}}) = 0$$

This is due to the fact that multiplying a set of linearly independent vectors by an invertible matrix preserves independence. From this it follows that the dimension of the columns of $\boldsymbol{A}$ is equal to the dimension of the columns of $\boldsymbol{M_R A}$, since $\boldsymbol{M_R}$ is invertible. Therefore neither the row rank nor the column rank changes.

Since $\boldsymbol{C}$ is in REF, the row rank is equal to the column rank, which is equal to the number of pivots, let it be r, then $\operatorname{rk}(\boldsymbol{A}) = r$.

29th January 2020

Properties of Symmetric Matrices

Take $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, such that $\boldsymbol{A}^\top = \boldsymbol{A}$.

All eigenvalues of $\boldsymbol{A}$ are positive (non-negative) iff $\forall \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ $[\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0 (\geq 0)]$. $\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}$ is called a quadratic form. For example;

$$\boldsymbol{A} = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= a x_1^2 + 2 b x_1 x_2 + c x_2^2$$

$$= k$$

assume $x_2 \neq 0$

$$\frac{\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}}{x_2^2} = a y^2 + 2 b^y + c \qquad\qquad \text{where } y = \frac{x_1}{x_2}$$

$$> 0 \qquad\qquad \text{focusing on the positive case}$$

If $\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0$, then we have an ellipse, with slanted major / minor axes. This uses a similar result to the result with the sphere's orthonormal basis being formed from the eigenvectors.

Considering this algebraically, if the quadratic in y is always greater than 0, then it must have no roots, hence the discriminant must be negative. Therefore we have the case where $(2b)^2 - 4ac < 0$, so $b^2 - ac < 0$, and $a > 0$.

Using the properties of the trace and the determinant, we can state the following;

$$\lambda_1 + \lambda_2 = a + c$$
$$\lambda_1 \lambda_2 = ac - b^2$$

We want both $ac - b^2$ and $a + b$ to be positive. The former is shown with the result for the discriminant, as if $b^2 - ac$ is negative, then $ac - b^2$ must be positive. The latter then follows as c is positive only if a is positive (due to $ac > b^2 \geq 0$), which we have previously. This shows it holds for the simple case.

The proof for any $n > 1$ is as follows;

$$\boldsymbol{S} = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

$$= \boldsymbol{V}^\top \boldsymbol{A} \boldsymbol{V} \qquad\qquad \Rightarrow$$

$$\boldsymbol{A} = \boldsymbol{V} \boldsymbol{S} \boldsymbol{V}^\top \qquad\qquad \Rightarrow$$

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{V} \boldsymbol{S} \boldsymbol{V}^\top \boldsymbol{x}$$

$$= \boldsymbol{z}^\top \boldsymbol{S} \boldsymbol{z} \qquad\qquad \text{where } \boldsymbol{z} = \boldsymbol{V}^\top \boldsymbol{x} (\text{which preserves } \ell_2\text{-norm})$$

$$= \sum_{i=1}^{n} \lambda_i z_i^2$$

To say $\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x}$ is positive for any non-zero $\boldsymbol{x}$ is to say that $\sum_{i=1}^{n} \lambda_i z_i^2$ is positive for any non-zero $\boldsymbol{v}$.

Then obviously all the eigenvalues (λ_i) must be positive iff $\sum_{i=1}^{n} \lambda_i z_i^2$ is positive.

Singular Value Decomposition (SVD)

This works for a **general** matrix (not necessarily square). Take any matrix $\mathbb{R}^{m \times n}$.

$$\exists V \in \mathbb{R}^{n \times n}, U \in \mathbb{R}^{m \times m}, \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$$

Where V, U are orthogonal matrices, $p = \min(m, n)$, and $r = \mathrm{rk}(A)$. We have a finitely decreasing sequence of σ. This allows us to write the following (assuming $n \leq m$);

$$A = USV^\top, \text{ where } S = \begin{bmatrix} \sigma_1 & & & & & & & \\ & \sigma_2 & & & & & & \\ & & \ddots & & & & & \\ & & & \sigma_r & & & & \\ & & & & 0 & & & \\ & & & & & \ddots & & \\ & & & & & & 0 & \\ \hline 0 & & \cdots & & & & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & & \cdots & & & & 0 \end{bmatrix}$$

This is the singular value decomposition of A, and $\sigma_1, \ldots, \sigma_r$ are the singular values of A. σ_1 is the largest singular value of A, which is the ℓ_2 matrix norm of A ($\|A\|_2$). The geometric intuition of this, in two dimensions, is as follows;
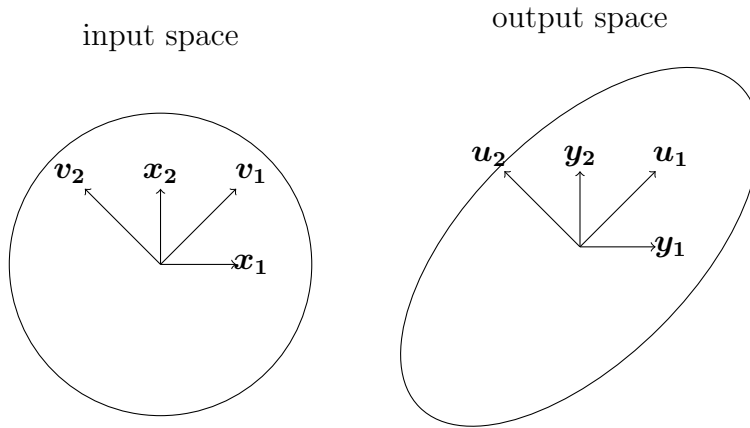
Starting with our coordinates in the basis (x_1, x_2), the matrix $V^\top$ performs a rotation to the new basis (v_1, v_2).

Take some point x on the unit circle with coordinates $\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, we obtain the new coordinates $\begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = V^\top x$.

We still maintain $\|x\| = 1$ and $\|v\| = 1$.

We're now in the basis of (u_1, u_2). We can say that $u_1 = \sigma_1 v_1$, $u_2 = \sigma_2 v_2$ (and so on, for higher dimensions). This implies the following result;

$$\frac{u_1^2}{\sigma_1^2} + \frac{u_2^2}{\sigma_2^2} = v_1^2 + v_2^2 = 1$$

Hence we obtain an ellipse.

Example

$$\boldsymbol{A} \in \mathbb{R}^{3\times 2} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\operatorname{rk}(\boldsymbol{A}) = 1$$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$$

$$\boldsymbol{S} \in \mathbb{R}^{3\times 2} = \operatorname{diag}(\sigma_1, \sigma_2, \dots)$$

$$\boldsymbol{A}^\top = \left(\boldsymbol{V}^\top\right)^\top \boldsymbol{S}^\top \boldsymbol{U}^\top$$

$$= \boldsymbol{V}\boldsymbol{S}^\top \boldsymbol{U}^\top$$

$$\boldsymbol{A}^\top \boldsymbol{A} = \boldsymbol{V}\boldsymbol{S}^\top \boldsymbol{U}^\top \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$$

$$= \boldsymbol{V}\boldsymbol{S}^\top \boldsymbol{S}\boldsymbol{V}^\top \qquad\qquad \boldsymbol{U} \text{ is an orthogonal matrix}$$

brief note on $\boldsymbol{S}^\top \boldsymbol{S} \in \mathbb{R}^{2\times 2}$

$$\boldsymbol{S}^\top \boldsymbol{S} = \operatorname{diag}(\sigma_1, \sigma_2, \dots)\,\operatorname{diag}(\sigma_1, \sigma_2, \dots)$$

$$= \operatorname{diag}(\sigma_1^2, \sigma_2^2, \dots)$$

brief proof for symmetry of $\boldsymbol{A}^\top \boldsymbol{A}$

$$\left(\boldsymbol{A}^\top \boldsymbol{A}\right)^\top = \boldsymbol{A}^\top \left(\boldsymbol{A}^\top\right)^\top$$

$$= \boldsymbol{A}^\top \boldsymbol{A}$$

$\boldsymbol{A}^\top \boldsymbol{A}$ is also positive semi-definite, meaning all its eigenvalues are non-negative (≥ 0) since

$$\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} = (\boldsymbol{A}\boldsymbol{x})^\top (\boldsymbol{A}\boldsymbol{x})$$

$$\geq 0$$

In this form, all we have to do is to find the spectral decomposition for $\boldsymbol{A}^\top \boldsymbol{A}$, since we have a symmetric matrix.

$$\boldsymbol{A}^\top \boldsymbol{A} = \begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix}$$

$$|\boldsymbol{A}^\top \boldsymbol{A} - \lambda \boldsymbol{I}| = (3 - \lambda)^2 - 9$$

$$\lambda_1 = 6 \qquad\qquad\qquad\qquad\qquad \Rightarrow$$

$$\sigma_1 = \sqrt{6}$$

$$\boldsymbol{v_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\lambda_2 = 0 \qquad\qquad\qquad\qquad\qquad \Rightarrow$$

$$\sigma_2 = 0$$

$$\boldsymbol{v_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$\boldsymbol{V} = [\boldsymbol{v_1}, \boldsymbol{v_2}]$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S} \qquad\qquad\qquad\qquad \Rightarrow$$

$$[\boldsymbol{A}\boldsymbol{v_1}, \boldsymbol{A}\boldsymbol{v_2}] = \boldsymbol{U} \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$= [\sqrt{6}\boldsymbol{u_1}, \boldsymbol{0}] \qquad\qquad\qquad\qquad \Rightarrow$$

$$\boldsymbol{u_1} = \frac{1}{\sqrt{6}} \boldsymbol{A} \boldsymbol{v_1}$$

$$= \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix}$$

$$\boldsymbol{u_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{unit vector orthogonal to } \boldsymbol{u_1}$$

$$\boldsymbol{u_3} = \boldsymbol{u_1} \times \boldsymbol{u_2}$$

$$= \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ -2 \end{bmatrix}$$

$$\boldsymbol{A} = [\boldsymbol{u_1}, \boldsymbol{u_2}, \boldsymbol{u_3}] \begin{bmatrix} \sqrt{6} & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} [\boldsymbol{v_1}, \boldsymbol{v_2}]$$

30th January 2020

Note that this lecture has no audio recording for some of the first part. To recap; the goal of SVD is to find the following, such that $\boldsymbol{A} = \boldsymbol{U} \boldsymbol{S} \boldsymbol{V}^\top$

- $\boldsymbol{A} \in \mathbb{R}^{m \times n}$
- $\boldsymbol{U} \in \mathbb{R}^{m \times m}$ $\qquad\qquad\qquad\qquad\qquad \boldsymbol{U}^\top = \boldsymbol{U}$
- $\boldsymbol{V} \in \mathbb{R}^{n \times n}$ $\qquad\qquad\qquad\qquad\qquad \boldsymbol{V}^\top = \boldsymbol{V}$
- $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = \cdots = \sigma_p = 0$ $\qquad p = \min\{m, n\}, \ r = \mathrm{rk}(\boldsymbol{A})$
- $\boldsymbol{S} \in \mathbb{R}^{m \times n} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$

Example

While normally $n \leq m$, we have the following example for $m < n$, where we work with $\boldsymbol{A}\boldsymbol{A}^\top \in \mathbb{R}^{m \times m}$;

$$\boldsymbol{A} = \begin{bmatrix} 3 & 2 & 2 \\ 2 & 3 & -2 \end{bmatrix}$$

$$\boldsymbol{A}^\top = \begin{bmatrix} 3 & 2 \\ 2 & 3 \\ 2 & -2 \end{bmatrix}$$

$$\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{U}\boldsymbol{S}\boldsymbol{S}^\top\boldsymbol{U}^\top$$

$$\boldsymbol{A}\boldsymbol{A}^\top = \begin{bmatrix} 17 & 8 \\ 8 & 17 \end{bmatrix}$$

$$\lambda_1 + \lambda_2 = \text{trace}$$
$$= 34$$
$$\lambda_1 \lambda_2 = \text{determinant}$$
$$= 225$$
$$\lambda_1 = 25$$
$$\sigma_1 = 5$$
$$\lambda_2 = 9$$
$$\sigma_2 = 3$$

$$\boldsymbol{u_2} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$\boldsymbol{u_1} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \text{orthogonal to } \boldsymbol{u_2}$$

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top \qquad\qquad \Rightarrow$$

$$\boldsymbol{A}^\top = \boldsymbol{V}\boldsymbol{S}^\top\boldsymbol{U}^\top \qquad\qquad \Rightarrow$$

$$\boldsymbol{A}^\top\boldsymbol{U} = \boldsymbol{V}\boldsymbol{S}^\top \qquad\qquad \Rightarrow$$

$$[\boldsymbol{A}^\top\boldsymbol{u_1}, \boldsymbol{A}^\top\boldsymbol{u_2}] = [5\boldsymbol{v_1}, 3\boldsymbol{v_2}] \qquad\qquad \Rightarrow$$

$$\boldsymbol{v_1} = \frac{1}{5}\boldsymbol{A}^\top\boldsymbol{u_1}$$

$$= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$\boldsymbol{v_2} = \frac{1}{3}\boldsymbol{A}^\top\boldsymbol{u_2}$$

$$= \frac{1}{3\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 4 \end{bmatrix}$$

$$\boldsymbol{v_3} = \boldsymbol{v_1} \times \boldsymbol{v_2} \qquad\qquad \text{orthogonal to both}$$

$$= \frac{1}{3} \begin{bmatrix} 2 \\ -2 \\ -1 \end{bmatrix}$$

General Strategy

$$\underbrace{\boldsymbol{A}^\top\boldsymbol{A}}_{\substack{\text{symmetric} \\ \text{+ve semi-definite}}} = \boldsymbol{V}\boldsymbol{S}^\top\underbrace{\boldsymbol{U}^\top\boldsymbol{U}}_{\boldsymbol{I}}\boldsymbol{S}\boldsymbol{V}^\top = \boldsymbol{V}\underbrace{\boldsymbol{S}^\top\boldsymbol{S}}_{\in\mathbb{R}^{n\times n}}\boldsymbol{V}^\top$$

Solve for the spectral decomposition of $\boldsymbol{A}^\top\boldsymbol{A}$ to get $\boldsymbol{V}$, and $\lambda_1 = \sigma_1^2, \lambda_2 = \sigma_2^2, \ldots, \lambda_r = \sigma_r^2$. From here we can do the following;

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top \qquad\qquad \Rightarrow$$

$$\boldsymbol{A}\boldsymbol{V} = \boldsymbol{U}\boldsymbol{S} \qquad\qquad \Rightarrow$$

$$\boldsymbol{A}\boldsymbol{v_i} = \sigma_i\boldsymbol{u_i} \qquad\qquad (1 \leq i \leq r), \Rightarrow$$

$$\boldsymbol{u_i} = \frac{\boldsymbol{A}\boldsymbol{v_i}}{\sigma_i}$$

For the remaining vectors, $\boldsymbol{u_{r+1}}, \ldots, \boldsymbol{u_n}$ find vectors to form an orthonormal basis. Use systems of linear equations in dimensions higher than 3, if the dimension is 3, just use the cross product. Solve $\boldsymbol{u_{r+1}}$ to $\boldsymbol{u_n}$ with

$$\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{u_j} = \boldsymbol{0} \text{ for } r + 1 \leq j \leq n$$

such that $(\boldsymbol{u_1}, \ldots, \boldsymbol{u_r}, \boldsymbol{u_{r+1}}, \ldots, \boldsymbol{u_n})$ forms an orthonormal basis.

input space

output space
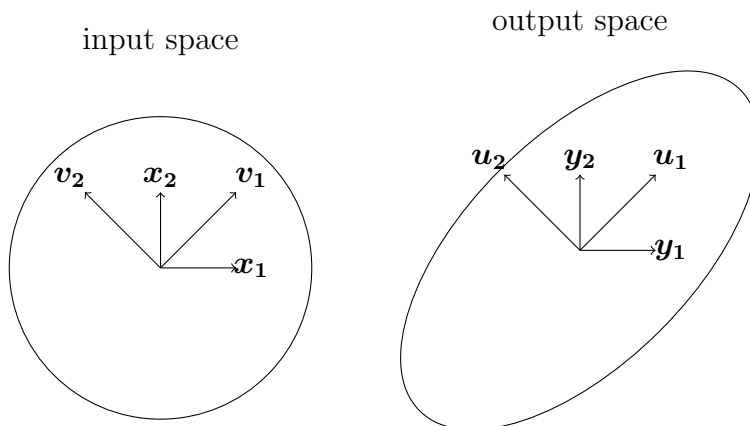
Using the same example as before, because I'm too lazy to draw this again, we see that $\boldsymbol{v_i} \mapsto \sigma_i \boldsymbol{v_i} = \boldsymbol{u_i}$. $\boldsymbol{v_1}$ is called the **principal component**, as it is the direction of the biggest change (since we have an ordering on σ).

A brief note on why $\boldsymbol{SS}^\top$ gives a square matrix;

$$\boldsymbol{S} = \begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ \hline 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$$

$$\boldsymbol{S}^\top = \begin{bmatrix} \sigma_1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & \ddots & \vdots \\ & & \sigma_r & 0 & \cdots & 0 \end{bmatrix}$$

$$\boldsymbol{S}^\top \boldsymbol{S} = \boldsymbol{SS}^\top \qquad\qquad \text{by symmetry}$$

$$= \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_r^2 \end{bmatrix}$$

Cholesky Factorisation / Decomposition of Symmetric Positive Definite Matrices

Take a matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, with $\boldsymbol{A} = \boldsymbol{A}^\top$ and $\boldsymbol{A}$ is positive (semi-)definite. Then the following holds

$$\exists \boldsymbol{L} \in \mathbb{R}^{n \times n} \, [\boldsymbol{A} = \boldsymbol{LL}^\top]$$

where $\boldsymbol{L}$ is a lower triangular matrix, with all diagonal elements being positive (non-zero);

$$\boldsymbol{L} = \begin{bmatrix} l_{1,1} & 0 & \cdots & 0 \\ \times & l_{2,2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \times & \cdots & \times & l_{n,n} \end{bmatrix}$$

This is a general result to the following, in the real numbers; $a > 0 \Rightarrow \exists l > 0 \, [a = l^2]$. This algorithm is done with brute force, by setting $\boldsymbol{A} = (a_{i,j})$, $\boldsymbol{L} = (l_{i,j})$, and $\boldsymbol{A} = \boldsymbol{LL}^\top$. The following is a simple example of this algorithm, with a 3×3 matrix;

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 10 & -1 \\ 1 & -1 & 5 \end{bmatrix}$$

$$= \boldsymbol{LL}^\top$$

$$\boldsymbol{L} = \begin{bmatrix} l_{1,1} & 0 & 0 \\ l_{2,1} & l_{2,2} & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} \end{bmatrix}$$

$$\boldsymbol{L}^\top = \begin{bmatrix} l_{1,1} & l_{2,1} & l_{3,1} \\ 0 & l_{2,2} & l_{3,2} \\ 0 & 0 & l_{3,3} \end{bmatrix}$$

by multiplying out $\boldsymbol{LL}^\top$, we can solve per column, starting with the first

$$l_{1,1}^2 = 1 \qquad\qquad \Rightarrow$$
$$l_{1,1} = 1 \qquad\qquad \text{taking positive square root}$$
$$l_{2,1}l_{1,1} = -1 \qquad\qquad \Rightarrow$$
$$l_{2,1} = -1$$
$$l_{3,1}l_{1,1} = 1 \qquad\qquad \Rightarrow$$
$$l_{3,1} = 1$$

second column

$$l_{2,1}^2 + l_{2,2}^2 = 10 \qquad\qquad \Rightarrow$$
$$l_{2,2} = 3$$
$$l_{3,1}l_{2,1} + l_{3,2}l_{2,2} = -1 \qquad\qquad \Rightarrow$$
$$-1 + 3l_{3,2} = -1 \qquad\qquad \Rightarrow$$
$$l_{3,2} = 0$$

third (final) column

$$l_{3,1}^2 + l_{3,2}^2 + l_{3,3}^2 = 5 \qquad\qquad \Rightarrow$$
$$1 + 0 + l_{3,3}^2 = 5 \qquad\qquad \Rightarrow$$
$$l_{3,3} = 2$$

setting the matrix

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

This can be used to easily solve the following;

$$\boldsymbol{Ax} = \boldsymbol{b} \qquad\qquad \Leftrightarrow$$
$$\boldsymbol{LL}^\top \boldsymbol{x} = \boldsymbol{b} \qquad\qquad \Leftrightarrow$$
$$\boldsymbol{Ly} = \boldsymbol{b} \qquad\qquad \text{let } \boldsymbol{y} = \boldsymbol{L}^\top \boldsymbol{x}$$

This is trivial to solve with forward substitution (y_1 can be solved easily, then substituted into the next equation to get y_2 and so on) as we have a lower triangular matrix

$$b_1 = l_{1,1}y_1 + 0 + \cdots + 0$$
$$b_2 = l_{2,1}y_1 + l_{2,2}y_2 + 0 + \cdots + 0$$
$$b_3 = l_{3,1}y_1 + l_{3,2}y_2 + l_{3,3}y_3 + 0 + \cdots + 0$$
$$\vdots$$
$$b_n = l_{n,1}y_1 + \cdots + l_{n,n}y_n$$

If $\boldsymbol{L}$ is lower triangular, then $\boldsymbol{L}^\top$ must be upper triangular, and thus can be solved with backwards substitution (such that x_n is used to solve x_{n-1}, and so on)

$$\boldsymbol{U} = \boldsymbol{L}^\top$$
$$y_n = 0 + \cdots + 0 + u_{n,n}x_n$$
$$y_{n-1} = 0 + \cdots + 0 + u_{n-1,n-1}x_{n-1} + u_{n-1,n}x_n$$
$$\vdots$$
$$y_1 = u_{1,1}x_1 + \cdots + u_{1,n}x_n$$

Rank-Nullity Theorem, Again

For a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the nullity of $\boldsymbol{A}$ (dimension of the kernel of $\boldsymbol{A}$) is $n - \text{rk}(\boldsymbol{A})$. To recap, a reduction to reduced row echelon form (RREF) is being in row echelon form, with the additional properties that all the pivots must be 1, and all other numbers in a column with a pivot are 0.

Applying the rank-nullity theorem to $\boldsymbol{A}^\top$, where $\boldsymbol{A}^\top \in \mathbb{R}^{n \times m}$;

$$\text{nullity}(\boldsymbol{A}^\top) + \text{rk}(\boldsymbol{A}^\top) = m \qquad\qquad \Rightarrow$$
$$\dim(\ker(\boldsymbol{A}^\top)) + \text{rk}(\boldsymbol{A}) = m \qquad \text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}^\top), \Rightarrow$$
$$\dim(\ker(\boldsymbol{A}^\top)) + \dim(\text{Im}(\boldsymbol{A})) = m \qquad \text{rk}(\boldsymbol{A}) = \dim(\text{Im}(\boldsymbol{A}))$$

It's important to now note that these are both subspaces of $\mathbb{R}^m$

$$\text{Im}(\boldsymbol{A}) \subseteq \mathbb{R}^m \qquad\qquad \text{output space of } \boldsymbol{A}$$
$$\ker(\boldsymbol{A}^\top) \subseteq \mathbb{R}^m$$

Working through an example matrix, we can observe the following;

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

$$\boldsymbol{A}^\top = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix}$$

$$\text{Im}(\boldsymbol{A}) = \left\{ x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} : x_1, x_2 \in \mathbb{R} \right\}$$

geometrically, $\text{Im}(\boldsymbol{A})$ is a plane going through the origin - intuitively, we can represent this with a normal (take cross product)

$$= \left\{ \boldsymbol{v} : \boldsymbol{v} \cdot \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} = 0 \right\}$$

$$\ker(\boldsymbol{A}^\top) = \{ \boldsymbol{w} : \boldsymbol{A}^\top \boldsymbol{w} = \boldsymbol{0} \}$$

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \boldsymbol{0} \qquad\qquad \Rightarrow$$

$$\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \boldsymbol{w} = 0$$

$$\text{and } \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \cdot \boldsymbol{w} = 0$$

therefore $\boldsymbol{w}$ is orthogonal to both those matrices, hence we can take the cross product, which we already know

$$\boldsymbol{w} = k \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \qquad\qquad k \in \mathbb{R}$$

We can then conclude the following;

$$\boldsymbol{w} \in \ker(\boldsymbol{A}^\top) \Leftrightarrow \boldsymbol{w} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = 0 = \boldsymbol{w} \cdot \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \Leftrightarrow \forall \boldsymbol{y} \in \text{Im}(\boldsymbol{A}) \, [\boldsymbol{w} \cdot \boldsymbol{y} = 0]$$

This gives the result that $\boldsymbol{w}$ is in the null space of $\boldsymbol{A}^\top$ iff it is orthogonal to every vector in the image space of $\boldsymbol{A}$.

Our conjecture is the following; $\ker(\boldsymbol{A}^\top) \perp \text{Im}(\boldsymbol{A})$, which is the same as saying

$$\boldsymbol{v} \in \ker(\boldsymbol{A}^\top) \Leftrightarrow \forall \boldsymbol{y} \in \text{Im}(\boldsymbol{A}) \, [\boldsymbol{y} \cdot \boldsymbol{v} = 0]$$

This can be proven generally as follows;

$$
\begin{aligned}
&\boldsymbol{v} \in \ker(\boldsymbol{A}^\top) \\
&\Leftrightarrow \boldsymbol{A}^\top \boldsymbol{v} = \boldsymbol{0} && \text{by definition of the null space / kernel} \\
&\Leftrightarrow \forall i \in [1, n] \, [\boldsymbol{a_i} \cdot \boldsymbol{v} = 0] && (1) \\
&\Leftrightarrow \forall \boldsymbol{x} \in \mathbb{R}^n \left[\left(\sum_{i=1}^{n} x_i \boldsymbol{a_i} \right) \cdot \boldsymbol{v} = 0 \right] && \text{generalise to any linear combination} \\
&\Leftrightarrow \boldsymbol{y} \in \text{Im}(\boldsymbol{A}) \, [\boldsymbol{y} \cdot \boldsymbol{v} = 0] && \blacksquare
\end{aligned}
$$

(1) row i of $\boldsymbol{A}^\top \in \mathbb{R}^{n \times m}$ is the same as column i of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, transposed, and all of the components must become 0

We say that $\mathbb{R}^m$ is the direct sum of $\ker(\boldsymbol{A}^\top)$ and $\text{Im}(\boldsymbol{A})$. We can get the following result, following what was proven earlier;

$$\boldsymbol{v} \in \text{Im}(\boldsymbol{A}) \wedge \boldsymbol{v} \in \ker(\boldsymbol{A}^\top) \Leftrightarrow \boldsymbol{v} \cdot \boldsymbol{v} = 0 \Leftrightarrow \boldsymbol{v} = \boldsymbol{0}$$

We want to prove the result for any vector $\boldsymbol{x} \in \mathbb{R}^m$, there exists a unique vector $\boldsymbol{x_r} \in \text{Im}(\boldsymbol{A})$ and $\boldsymbol{x_n} \in \ker(\boldsymbol{A}^\top)$, such that $\boldsymbol{x} = \boldsymbol{x_r} + \boldsymbol{x_n}$, or formally;

$$\forall \boldsymbol{x} \in \mathbb{R}^m \, \exists! \boldsymbol{x_r} \in \text{Im}(\boldsymbol{A}), \boldsymbol{x_n} \in \ker(\boldsymbol{A}^\top) \, [\boldsymbol{x} = \boldsymbol{x_r} + \boldsymbol{x_n}]$$

By taking the union of a basis of $\text{Im}(\boldsymbol{A})$, let it be $(\boldsymbol{v_1}, \ldots, \boldsymbol{v_r})$, and a basis of $\ker(\boldsymbol{A}^\top)$, let it be $(\boldsymbol{v_{r+1}}, \ldots, \boldsymbol{v_m})$, we get a basis of $\mathbb{R}^m$: $(\boldsymbol{v_1}, \ldots, \boldsymbol{v_r}, \boldsymbol{v_{r+1}}, \ldots, \boldsymbol{v_m})$. This gives us the following result;

$$\boldsymbol{x} \in \mathbb{R}^m = \sum_{i=1}^{m} x_i \boldsymbol{v_i} = \underbrace{\sum_{i=1}^{r} x_i \boldsymbol{v_i}}_{\in \text{Im}(\boldsymbol{A})} + \underbrace{\sum_{i=r+1}^{m} x_i \boldsymbol{v_i}}_{\in \ker(\boldsymbol{A}^\top)} = \boldsymbol{x_r} + \boldsymbol{x_n}$$

In order to prove uniqueness, we assume $\boldsymbol{x} = \boldsymbol{x_r} + \boldsymbol{x_n} = \boldsymbol{x_r'} + \boldsymbol{x_n'}$. The violet part only holds iff $\boldsymbol{x_r} - \boldsymbol{x_r'} = \boldsymbol{x_n'} - \boldsymbol{x_n}$. As they are both in their respective subspaces of $\text{Im}(\boldsymbol{A})$ and $\ker(\boldsymbol{A}^\top)$, they must both be $\boldsymbol{0}$, as shown above. This gives the result $\boldsymbol{x_r} = \boldsymbol{x_r'} \wedge \boldsymbol{x_n} = \boldsymbol{x_n'}$, hence proving uniqueness.

5th February 2020

Positive Definiteness

The conditions for positive definiteness ($\forall \boldsymbol{x} \neq \boldsymbol{0} \, [\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0]$) of $\boldsymbol{A}$ are as follows ;

- all diagonal elements of $\boldsymbol{A}$ must be > 0

  Take $\boldsymbol{x} = \boldsymbol{e_i}$, which is all 0, other than the ith entry being 1. Then $\boldsymbol{e_i}^\top \boldsymbol{A} \boldsymbol{e_i} = a_{i,i}$, which has to be greater than 0.

- all principal minors of $\boldsymbol{A}$ must be positive definite.

  For $1 \leq k \leq n$, $\boldsymbol{A}^{(k)} \in \mathbb{R}^{k \times k}$ denotes the top left square submatrix of $\boldsymbol{A}$.

$$\boldsymbol{x} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix} \qquad \text{where } \boldsymbol{y} \in \mathbb{R}^k, \boldsymbol{0} \in \mathbb{R}^{n-k}$$

$$\boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}^\top \boldsymbol{A} \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}$$

$$= \boldsymbol{y}^\top \boldsymbol{A}^{(k)} \boldsymbol{y}$$

  So for any $\boldsymbol{y} \neq \boldsymbol{0}$, $\boldsymbol{x} \neq \boldsymbol{0}$, thus $\boldsymbol{y}^\top \boldsymbol{A}^{(k)} \boldsymbol{y} > 0 \Leftrightarrow \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} > 0$

- $|a_{i,j}| < \max\{a_{i,i}, a_{j,j}\}$

  For any entry not on the diagonal, there are two corresponding entries on the diagonal (one corresponding to the row, the other to the column);

$$\boldsymbol{A} = \begin{bmatrix} \ddots & & & & \\ & a_{i,i} & \cdots & a_{i,j} & \\ & & \ddots & \vdots & \\ & & & a_{j,j} & \\ & & & & \ddots \end{bmatrix}$$

$$\text{take } \boldsymbol{x} = \boldsymbol{e}_i + \boldsymbol{e}_j$$

$$\begin{aligned} \boldsymbol{x}^\top \boldsymbol{A} \boldsymbol{x} &= (\boldsymbol{e}_i \pm \boldsymbol{e}_j)^\top \boldsymbol{A} (\boldsymbol{e}_i \pm \boldsymbol{e}_j) \\ &= \left(\boldsymbol{e}_i^\top \pm \boldsymbol{e}_j^\top \right) \boldsymbol{A} (\boldsymbol{e}_i \pm \boldsymbol{e}_j) \\ &= \boldsymbol{e}_i^\top \boldsymbol{A} \boldsymbol{e}_i + \boldsymbol{e}_j^\top \boldsymbol{A} \boldsymbol{e}_j \pm \boldsymbol{e}_i^\top \boldsymbol{A} \boldsymbol{e}_j \pm \boldsymbol{e}_j^\top \boldsymbol{A} \boldsymbol{e}_i \\ &= a_{i,i} + a_{j,j} \pm 2a_{i,j} \qquad\qquad\qquad\qquad \boldsymbol{A} \text{ is symmetric} \\ &> 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow \end{aligned}$$

$$-a_{i,j} < \frac{a_{i,i} + a_{j,j}}{2} \qquad\qquad\qquad \text{diagonal terms are positive}$$

$$\text{or } a_{i,j} < \frac{a_{i,i} + a_{j,j}}{2} \qquad\qquad\qquad\qquad\qquad\qquad\qquad \Rightarrow$$

$$|a_{i,j}| < \max\{a_{i,i}, a_{j,j}\}$$

- all positive eigenvalues

For an example, take the following;

$$\boldsymbol{A} = \begin{bmatrix} 1 & -1 & 3 \\ -1 & 10 & -1 \\ -3 & -1 & 2 \end{bmatrix}$$

However, we see that $|-3| = 3 > 2 = \max\{1, 2\}$, hence $\boldsymbol{A}$ is not positive definite

Least Square Method

For a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, where n is commonly the dimension of the data, and m the number of samples (such that each row is an entry), we can write $\boldsymbol{A}$ as;

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a^1} \\ \boldsymbol{a^2} \\ \vdots \\ \boldsymbol{a^m} \end{bmatrix} \text{(all rows)}$$

The goal of the LSM is to find a $x \in \mathbb{R}^n$, such that $||Ax - b||_2$ is minimised (which is equivalent to minimising $||Ax - b||_2^2$, as it is positive). From previous results, we can take $b = b_r + b_n$, therefore we can minimise the following;

$$||Ax - b_r - b_n||_2^2 = ((Ax - b_r) - b_n) \cdot ((Ax - b_r) - b_n)$$

$$= (Ax - b_r) \cdot (Ax - b_r) + b_n \cdot b_n - 2 \overbrace{(Ax - b_r)}^{\in \mathrm{Im}(A)} \cdot \underbrace{b_n}_{\in \ker(A^\top)} \quad (1)$$

$$= (Ax - b_r) \cdot (Ax - b_r) + b_n \cdot b_n \quad (2)$$

(1) the dot product is bilinear in both variables, hence we can expand out

(2) range space of A is perpendicular to the null space of $A^\top$, hence the dot product is 0

Therefore, we want to minimise $(Ax - b_r) \cdot (Ax - b_r) + b_n \cdot b_n$. However, we have no control over b_n, therefore we want to solve for $Ax = b_r$. This will always have a solution, as $b_r \in \mathrm{Im}(A)$, which means that such a x exists by definition of the range space. We then claim the following;

$$Ax = b_r \Leftrightarrow A^\top Ax = A^\top b$$

Therefore, if the left hand side of the double implication has a solution, which we know it does, then the right hand side must also have a solution. The proof for this is done in both directions;

proving "$\Rightarrow$"

$$\text{suppose } Ax = b_r \qquad\qquad\qquad \Rightarrow$$
$$A^\top Ax = A^\top b_r \qquad\qquad\qquad \text{pre-multiply}$$
$$= A^\top (b_r + b_n) \qquad\qquad \text{in null space, hence } A^\top b_n = 0$$
$$= A^\top b \qquad\qquad\qquad\qquad \blacksquare$$

proving "$\Leftarrow$"

$$\text{suppose } A^\top Ax = A^\top b \qquad\qquad\qquad\qquad \Rightarrow$$
$$A^\top Ax - A^\top (b_r + b_n) = 0 \qquad\qquad\qquad\qquad \Rightarrow$$
$$A^\top (Ax - b_r - b_n) = 0 \qquad\qquad\qquad\qquad \Rightarrow$$
$$Ax - b_r - b_n \in \ker(A^\top) \qquad \text{property of adding in subspace}, \Rightarrow$$
$$Ax - b_r \in \ker(A^\top) \qquad\qquad\qquad\qquad \text{and}$$
$$Ax - b_r \in \mathrm{Im}(A^\top) \qquad \text{adding two elements of the range space}, \Rightarrow$$
$$Ax - b_r = 0 \qquad\qquad\qquad\qquad \Rightarrow$$
$$Ax = b_r \qquad\qquad\qquad\qquad \blacksquare$$

An example of using this is as follows (note that $Ax = b$ is inconsistent);

$$A = \begin{bmatrix} 2 & 2 \\ 1 & 2 \\ 2 & 0 \end{bmatrix}$$

$$b = \begin{bmatrix} 0 \\ 5 \\ -1 \end{bmatrix}$$

find $\boldsymbol{x}$ such that $\boldsymbol{Ax} = \boldsymbol{b_r}$

$$\boldsymbol{A^\top A x} = \boldsymbol{A^\top b} \qquad \Rightarrow$$

$$\begin{bmatrix} 9 & 6 \\ 6 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 \\ 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 5 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ 10 \end{bmatrix}$$

we know this is positive definite, and therefore there is a solution

$$x_1 = -1$$
$$x_2 = 2$$

6th February 2020

Using Cholesky Decomposition for LSM

While it's easy in some cases to solve with an inverse, we have a symmetric positive definite matrix in $\boldsymbol{A^\top A}$. Therefore we have the following steps;

- goal is to solve $\boldsymbol{A^\top A x} = \boldsymbol{A^\top b}$
- let $\boldsymbol{A^\top A} = \boldsymbol{LL^\top}$, therefore we have $\boldsymbol{LL^\top x} = \boldsymbol{A^\top b}$
- let $\boldsymbol{y} = \boldsymbol{L^\top x}$
- find $\boldsymbol{y}$ with forward substitution in $\boldsymbol{Ly} = \boldsymbol{A^\top b}$
- find $\boldsymbol{x}$ with backward substitution in $\boldsymbol{y} = \boldsymbol{L^\top x}$

First Application (Linear Regression) + Example

A simple application for this is to consider m samples of pairs (a_i, b_i). We want to find an affine line $y = s_1 x + s_0$, such that;

$$\sum_{i=1}^{m} e_i^2 \text{ is minimised, where the error term } e_i = |(s_1 a_i + s_0) - b_i|$$

In this case, we only have two unknowns, and therefore we can write the following;

$$\boldsymbol{A} = \begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_m \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$\boldsymbol{z} = \begin{bmatrix} s_0 \\ s_1 \end{bmatrix}$$

$$\boldsymbol{Az} = \boldsymbol{b} \qquad \Rightarrow$$

$$
\begin{bmatrix} 1 & a_1 \\ 1 & a_2 \\ \vdots & \vdots \\ 1 & a_m \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \end{bmatrix} = \begin{bmatrix} s_0 + a_1 s_1 \\ s_0 + a_2 s_1 \\ \vdots \\ s_0 + a_m s_1 \end{bmatrix}
$$

$$
= \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}
$$

However, being able to solve it exactly is unlikely, especially for large datasets. In a real example, we have the following;

$$(a_1, b_1) = (1,6), (2,5), (3,7), (4,10)$$

$$
\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}
$$

$$
\mathbf{b} = \begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix}
$$

$$
\mathbf{A}^\top \mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}
$$

$$
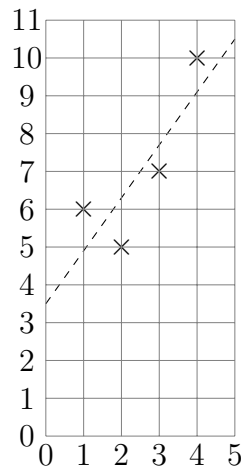= \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix}
$$

$$
\mathbf{A}^\top \mathbf{b} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix}
$$

$$
= \begin{bmatrix} 28 \\ 77 \end{bmatrix}
$$

by doing Gaussian elimination, we get the following results

$$s_0 = \frac{7}{2}$$

$$s_1 = \frac{7}{5}$$

Higher Dimensions

In order to generalise this to higher dimensions, for each (a_i, b_i), we now have a vector $\boldsymbol{a_i} \in \mathbb{R}^n$, and a scalar $b_i \in \mathbb{R}$, for $i = 1, \ldots, m$. The affine line is now no longer as simple, and is therefore written as;

$$y = \boldsymbol{x}^\top \boldsymbol{s} + s_0 = \boldsymbol{x} \cdot \boldsymbol{s} + s_0 = s_0 + \sum_{j=1}^{n} x_j s_j, \text{ where } \boldsymbol{s} \in \mathbb{R}^n, \text{ and } s_0 \in \mathbb{R}.$$

Setting this out, we have the following;

$$\boldsymbol{a_i} = \begin{bmatrix} a_{i,1} \\ \vdots \\ a_{i,n} \end{bmatrix} \qquad\qquad \text{for } i = 1, \ldots, m$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & a_{1,1} & \cdots & a_{1,n} \\ 1 & a_{2,1} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & a_{2,1} & \cdots & a_{m,n} \end{bmatrix}$$

$$\boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$\boldsymbol{z} = \begin{bmatrix} s_0 \\ \boldsymbol{s} \end{bmatrix} \qquad\qquad \boldsymbol{s} \in \mathbb{R}^n$$

Second Application (Polynomial Regression)

Another application of this is to suppose data is (y_i, t_i), where $y_i \in \mathbb{R}$ and $t_i \in \mathbb{R}$ (considered as "time" - but is any continuous variable). We hypothesise the following;

$$y(t) = \sum_{j=1}^{n} P_j f_j(t)$$

where P_js are the parameters of the model, and f_js are basic functions, for **example**; $f_1(t) = 1$, $f_2(t) = t, \ldots, f_n(t) = t^{n-1}$. This is called **polynomial regression**. We know (we're given) f_js, but we want to find the P_js.

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$$
$$\boldsymbol{A} = (a_{i,j})$$
$$a_{i,j} = f_j(t_i)$$
$$\boldsymbol{x} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix}$$

An example of this is to determine height (h), and the coefficient of gravity (g). We hypothesise that the distance with respect to time, away from the start, is

$$y(t) = h - g\frac{t^2}{2}$$

We assume that h and g are the unknown parameters, and define the basic functions as follows (note that the pairs (y_i, t_i) can be obtained with experiments);

$$f_1(t) = 1$$

$$f_2(t) = -\frac{t^2}{2}$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & -\frac{t_1^2}{2} \\ \vdots & \vdots \\ 1 & -\frac{t_m^2}{2} \end{bmatrix}$$

$$\boldsymbol{x} = \begin{bmatrix} h \\ g \end{bmatrix}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

This can now be solved with the standard LSM.

QR Decomposition

We define $\boldsymbol{Q} \in \mathbb{R}^{m \times n}$ to be an orthogonal matrix if $\boldsymbol{Q} = [\boldsymbol{q_1}, \ldots, \boldsymbol{q_n}]$, the column vectors are unit vectors with respect to the ℓ_2-norm, and are pairwise orthogonal. This is only possible if $n \leq m$. It's important to note that $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}_n$, and $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}_m$.

The goal is to find $\boldsymbol{Q}$ and $\boldsymbol{R}$ such that $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, where $\boldsymbol{Q}$ is an orthogonal matrix, and $\boldsymbol{R}$ is an upper triangular matrix. It's trivial to see that $\boldsymbol{R} = \boldsymbol{Q}^\top \boldsymbol{A}$, due to the previously mentioned property. This is used to easily solve the following (since $\boldsymbol{R}$ is upper triangular, we can employ backward substitution);
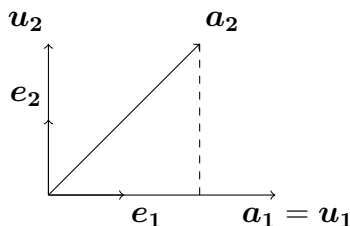
$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \Rightarrow \boldsymbol{Q}\boldsymbol{R}\boldsymbol{x} = \boldsymbol{b} \Rightarrow \boldsymbol{R}\boldsymbol{x} = \boldsymbol{Q}^\top \boldsymbol{b}$$

Gram-Schmidt Process

Suppose we have a set of linearly independent vectors $\boldsymbol{a_1}, \ldots, \boldsymbol{a_n} \in \mathbb{R}^m$. The Gram-Schmidt process is as follows;

$$\boldsymbol{u_1} = \boldsymbol{a_1} \qquad\qquad \boldsymbol{e_1} = \frac{\boldsymbol{u_1}}{||\boldsymbol{u_1}||}$$

$$\boldsymbol{u_2} = \boldsymbol{a_2} - (\boldsymbol{e_1} \cdot \boldsymbol{a_2})\boldsymbol{e_1} \qquad\qquad \boldsymbol{e_2} = \frac{\boldsymbol{u_2}}{||\boldsymbol{u_2}||}$$

$$\boldsymbol{u_3} = \boldsymbol{a_3} - (\boldsymbol{e_1} \cdot \boldsymbol{a_3})\boldsymbol{e_1} - (\boldsymbol{e_2} \cdot \boldsymbol{a_3})\boldsymbol{e_2} \qquad\qquad \boldsymbol{e_2} = \frac{\boldsymbol{u_3}}{||\boldsymbol{u_3}||}$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

$$\boldsymbol{u_n} = \boldsymbol{a_n} - \sum_{j=1}^{n-1}(\boldsymbol{e_j} \cdot \boldsymbol{a_n})\boldsymbol{e_j} \qquad\qquad \boldsymbol{e_n} = \frac{\boldsymbol{u_n}}{||\boldsymbol{u_n}||}$$

Geometrically, we can visualise it as the following;

In general, for $1 \leq j \leq n$, we have

$$\boldsymbol{a_j} = (\boldsymbol{e_1} \cdot \boldsymbol{a_j})\boldsymbol{e_1} + (\boldsymbol{e_2} \cdot \boldsymbol{a_j})\boldsymbol{e_2} + \cdots + (\boldsymbol{e_j} \cdot \boldsymbol{a_j})\boldsymbol{e_j}$$

First Method for QR Decomposition (Gram-Schmidt Process)

Take a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, where $\boldsymbol{A} = [\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}]$, assuming that $\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}$ are linearly independent. We then apply Gram-Schmidt to $\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}$ as described above to get $\boldsymbol{Q}_0 = (\boldsymbol{e_1}, \ldots, \boldsymbol{e_n})$, where. Note that for $1 \le j \le n$, $\boldsymbol{e_j} \cdot \boldsymbol{e_j} = 1$, and for $j \ne k$, $\boldsymbol{e_j} \cdot \boldsymbol{e_k} = 0$. Let $\boldsymbol{Q} = [\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}]$, which is orthogonal. While we can see that $\boldsymbol{R} = \boldsymbol{Q}^\top \boldsymbol{A}$, we can write this as;

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{e_1} \cdot \boldsymbol{a_1} & \boldsymbol{e_1} \cdot \boldsymbol{a_2} & \boldsymbol{e_1} \cdot \boldsymbol{a_3} & \cdots & \boldsymbol{e_1} \cdot \boldsymbol{a_n} \\ & \boldsymbol{e_2} \cdot \boldsymbol{a_2} & \boldsymbol{e_2} \cdot \boldsymbol{a_3} & \cdots & \boldsymbol{e_2} \cdot \boldsymbol{a_n} \\ & & \boldsymbol{e_3} \cdot \boldsymbol{a_3} & \cdots & \boldsymbol{e_3} \cdot \boldsymbol{a_n} \\ & & & \ddots & \vdots \\ & & & & \boldsymbol{e_n} \cdot \boldsymbol{a_n} \end{bmatrix} \in \mathbb{R}^{n \times n}$$

We claim that the diagram below is valid when $\boldsymbol{A}$ represents the linear map $f : \mathbb{R}^n \to \mathbb{R}^m$, in the standard basis.

$$\begin{array}{ccc} \mathbb{R}^n_{\boldsymbol{E_n}} & \xrightarrow{\;f_{\boldsymbol{E_m E_n}} = \boldsymbol{A}\;} & \mathbb{R}^m_{\boldsymbol{E_m}} \\ & & \\ f_{\boldsymbol{Q_0 E_n}} = \boldsymbol{R} \searrow & & \downarrow \mathrm{Id}_{\boldsymbol{E_m Q_0}} = \boldsymbol{Q} \\ & \mathrm{Im}(\boldsymbol{A})_{\boldsymbol{Q_0}} & \end{array}$$

12th February 2020

Proof of Diagram

Continuing on from last week's lecture, we aim to prove two claims in the diagram;

- $f_{\boldsymbol{Q_0 E_n}} = \boldsymbol{R}$

  The matrix $\boldsymbol{R}$ represents the linear map f from $\mathbb{R}^n$ (in the standard basis) to $\mathbb{R}^m$ in the new basis $(\boldsymbol{Q}_0)$.

  To begin, we observe how the standard basis is mapped under $\boldsymbol{A}$;

$$j^{\text{th}} \to \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \xmapsto{\quad \boldsymbol{A}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \boldsymbol{a_j}\quad} \boldsymbol{a_j} = (\boldsymbol{e_1} \cdot \boldsymbol{a_j})\boldsymbol{e_1} + (\boldsymbol{e_2} \cdot \boldsymbol{a_j})\boldsymbol{e_2} + \cdots + (\boldsymbol{e_j} \cdot \boldsymbol{a_j})\boldsymbol{e_j} + 0\boldsymbol{e_{j+1}} + \cdots + 0\boldsymbol{e_n}$$

  Looking at these coefficients with respect to the basis formed by $(\boldsymbol{e_1}, \ldots, \boldsymbol{e_n})$, which are highlighted in violet, it's clear that it forms the jth column of $\boldsymbol{R}$.

- $\mathrm{Id}_{\boldsymbol{E_m Q_0}} = \boldsymbol{Q}$

  This is a change of basis in the image space of $\boldsymbol{A}$.

  This is trivial to show, as we want to look at what happens to the jth vector in the new basis $\boldsymbol{Q}_0$. To write each of these matrices in the new (standard) basis, we simply do the following;

$$\boldsymbol{e_j} = e_{1,j} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + e_{2,j} \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \cdots + e_{m,j} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}$$

Trivially, we get $\boldsymbol{Q} = [\boldsymbol{e_1}, \ldots, \boldsymbol{e_n}]$, as expected.

Note on Basis Change

To find the matrix representing $f_{\boldsymbol{DB}}$, which takes a vector in basis $\boldsymbol{B}$ to a vector in basis $\boldsymbol{D}$, we look at what happens to each basis vector.

$$\boldsymbol{b_j} \mapsto f(\boldsymbol{b_j}) = \delta_1 \boldsymbol{d_1} + \delta_2 \boldsymbol{d_1} + \cdots + \delta_m \boldsymbol{d_m}$$

Then the jth column of the matrix is given as;

$$(f_{\boldsymbol{DB}})_j = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_m \end{bmatrix}$$

Householder Operator

I can't draw the geometric intuition nicely, at all, look at Panopto timestamp 36:37. However, imagine there is a plane, and there is some point $\boldsymbol{x}$ that exists. That point is orthogonally projected onto the plane, with the orthogonal projection matrix $\boldsymbol{P}$, such that the vector between the point $\boldsymbol{x}$ and $\boldsymbol{Px}$ is perpendicular to the plane. Now, take the perpendicular unit vector for the plane as $\boldsymbol{u}$, therefore the length of $\boldsymbol{x}$ to $\boldsymbol{Px}$ is $\boldsymbol{u} \cdot \boldsymbol{x}$. Imagine the plane as a mirror, and the image of $\boldsymbol{x}$ is on the other side of the plane, let that point be $H(\boldsymbol{x}) = \boldsymbol{Hx}$. Therefore, we can work out $\boldsymbol{H}$ as follows;

$$H(\boldsymbol{x}) = \boldsymbol{x} - 2(\boldsymbol{u} \cdot \boldsymbol{x})\boldsymbol{u} = \boldsymbol{x} - 2\boldsymbol{x}\boldsymbol{u}\boldsymbol{u}^\top = \underbrace{\left(\boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^\top\right)}_{H} \boldsymbol{x}$$

We expect $\boldsymbol{H}^2$ to give us the identity, as if "reflect" a reflected point, it should give the original;

$$\begin{aligned} \boldsymbol{H}^2 &= \left(\boldsymbol{I} - 2\boldsymbol{u}\boldsymbol{u}^\top\right) \\ &= \boldsymbol{I} - 4\boldsymbol{u}\boldsymbol{u}^\top + 4\boldsymbol{u}\left(\boldsymbol{u}^\top \boldsymbol{u}\right)\boldsymbol{u}^\top \\ &= \boldsymbol{I} - 4\boldsymbol{u}\boldsymbol{u}^\top + 4\boldsymbol{u}\boldsymbol{u}^\top \qquad\qquad \ell_2\text{-norm of unit vector is 1} \\ &= \boldsymbol{I} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \blacksquare \end{aligned}$$

Note that there are two distinct eigenvalues of $\boldsymbol{H}$; $\lambda_1 = 1$ has multiplicity $m_1 = m - 1$, where m is the dimension of the space. This is for each of the vectors $\boldsymbol{v_i}$ that span the plane, as they are mapped to themselves ($\boldsymbol{Hv_i} = H(\boldsymbol{v_i}) = \boldsymbol{v_i}$). There is also an additional eigenvalue $\lambda_2 = -1$, for $\boldsymbol{u}$, as we can see that $\boldsymbol{Hu} = H(\boldsymbol{u}) = -\boldsymbol{u}$.

Second Method for QR Decomposition (Householder Operator)

Take the matrix $\boldsymbol{A} = [\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}]$. The Householder operator must preserve the length, in the following mapping;

$$H\left(\begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix}\right) = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$
$$\pm\lambda = \|\boldsymbol{a_1}\|$$

We want to send the vector $\boldsymbol{a_1}$ to the vector $||\boldsymbol{a_1}||\boldsymbol{e_1}$, where $\boldsymbol{e_1}$ is all 0, other than the first element.

$$\boldsymbol{u} = \frac{\boldsymbol{a_1} - ||\boldsymbol{a_1}||\boldsymbol{e_1}}{||\boldsymbol{a_1} - ||\boldsymbol{a_1}||\boldsymbol{e_1}||}$$

$$\boldsymbol{H_1} = \boldsymbol{I} = 2\boldsymbol{u}\boldsymbol{u}^\top$$

$$\boldsymbol{H_1}\boldsymbol{A} = \boldsymbol{H_1}[\boldsymbol{a_1}, \ldots, \boldsymbol{a_n}]$$

$$= [\boldsymbol{H_1}\boldsymbol{a_1}, \ldots, \boldsymbol{H_1}\boldsymbol{a_n}]$$

$$= \begin{bmatrix} ||\boldsymbol{a_1}|| & \times \\ \boldsymbol{0} & \boldsymbol{A_2} \end{bmatrix} \qquad \times \text{ is an unknown row, } \boldsymbol{A_2} \text{ is the bottom right submatrix}$$

Applying $\boldsymbol{H}$ to $\boldsymbol{A}$ gets rid of all entries in the first column, other than the first element. This is an iterative process, where every iteration brings us closer to an upper triangular matrix.

$$\boldsymbol{H_n} \cdots \boldsymbol{H_1}\boldsymbol{A} = \boldsymbol{R}$$

By reversing this, we can get the following result;

$$\boldsymbol{A} = \boldsymbol{H_1} \cdots \boldsymbol{H_n}\boldsymbol{R}$$

Using the Householder operator to do QR decomposition isn't examinable.