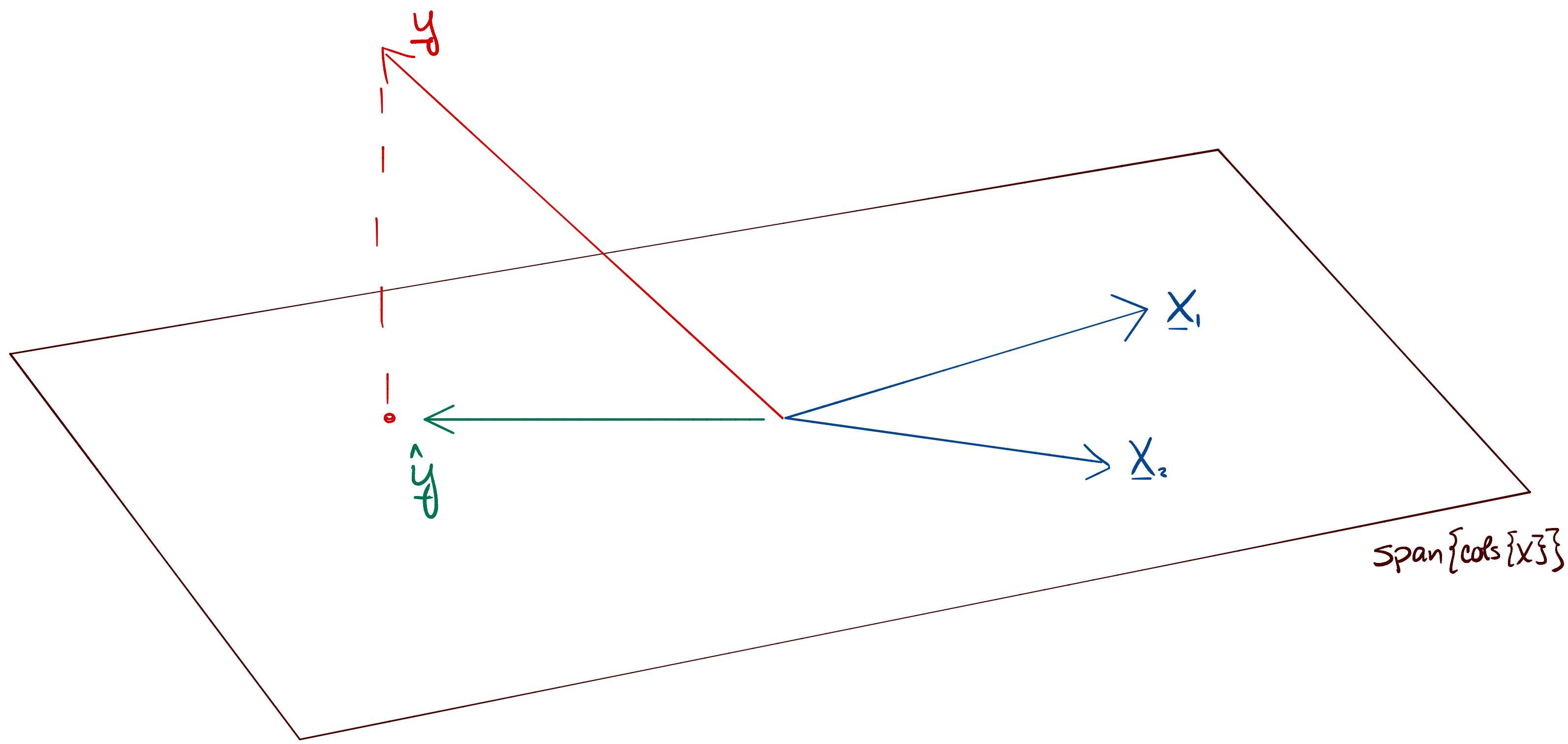
Lecture 5: Subspaces, Bases, and Projections

Mathematical Foundations of Machine Learning
University of Chicago

Recall our geometric picture of least squares:



The hyperplane $\text{span}\{\text{cols}\{X\}\}$ is called a <u>subspace</u>

The 2 columns of X in the image above <u>span</u> the subspace.

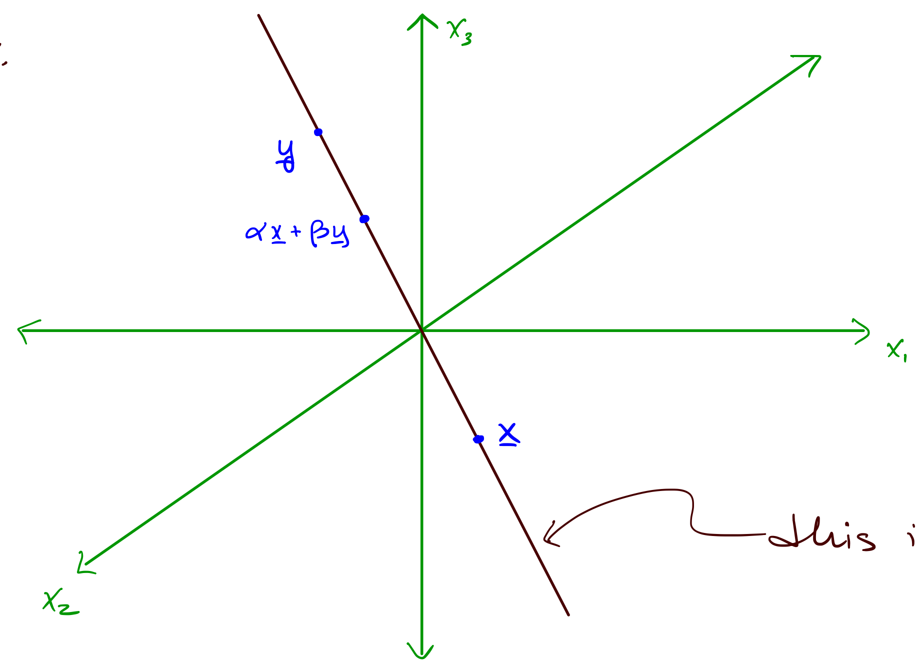
$\hat{y}$ is the <u>orthogonal projection</u> of y onto the subspace

Today we will discuss these concepts more formally.

Subspaces

Consider all points $\underline{x} \in \mathbb{R}^n$. A subspace is a subset of those points satisfying a few key properties. Specifically, let $\mathcal{S}$ be a subspace and let $\underline{x}$ and $\underline{y}$ be any two points in the subspace. Then for any scalars α and β, the weighted sum $\alpha\underline{x} + \beta\underline{y}$ must also be in the subspace.

Ex.



this is a subspace (1-dimensional subspace of $\mathbb{R}^3$)

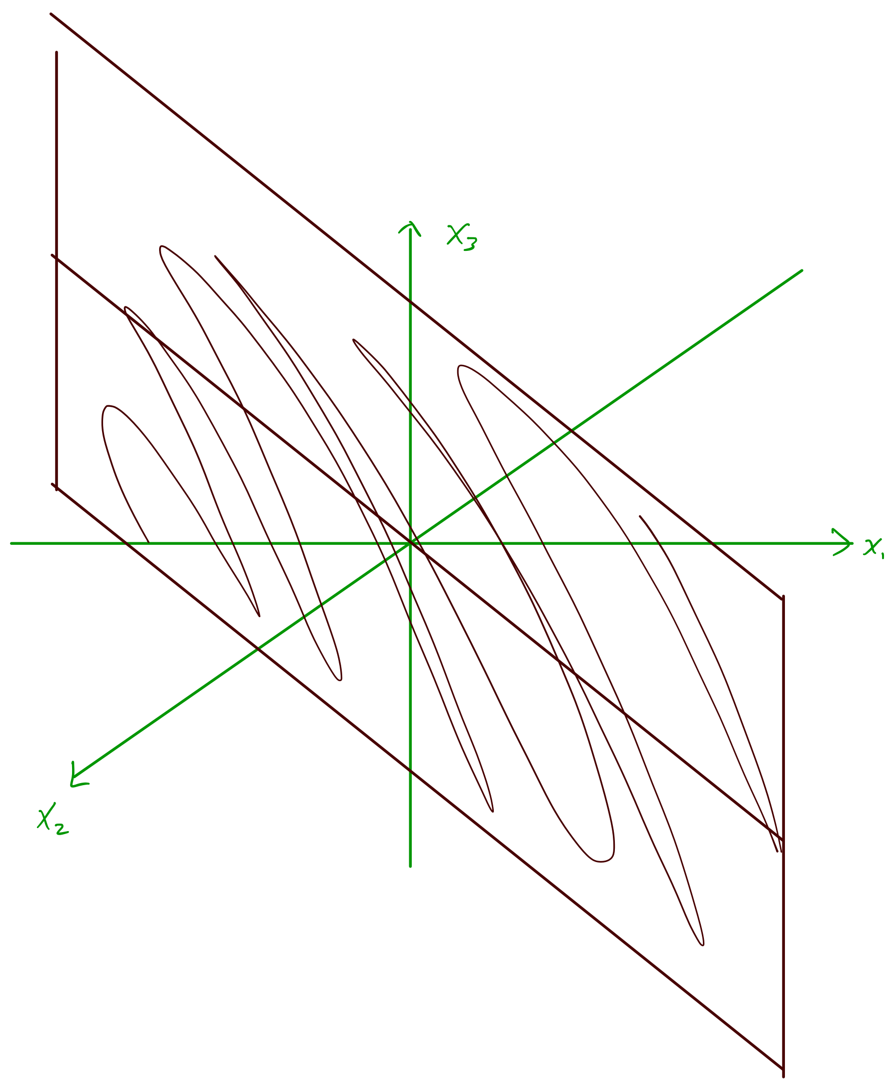
Q: is zero-vector in $\mathcal{S}$?

Q: can a subspace have an edge or boundary?

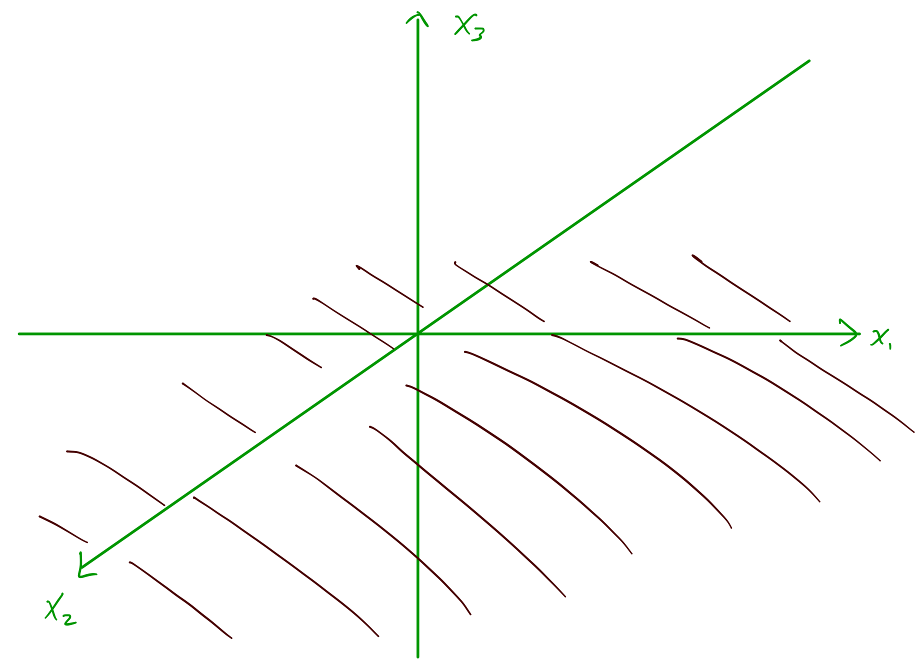
$$\mathcal{S} = \{ \underline{x} \in \mathbb{R}^3 : x_1 = x_2 = -x_3 \}$$

$$\underline{x} \in \mathcal{S} \iff \underline{x} = \alpha \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \text{ for some } \alpha \implies \mathcal{S} = \mathrm{span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$$

Ex: $n = 3$. $S = \{ \underline{x} \in \mathbb{R}^3 : x_1 = x_2 \}$



Ex. $S = \{ \underline{x} \in \mathbb{R}^3 : x_3 = 0 \}$



Ex. we are given points $\underline{X}_1, \underline{X}_2, \underline{X}_3 \in \mathbb{R}^n$

$$\mathcal{S} = \operatorname{Span}\{\underline{X}_1, \underline{X}_2, \underline{X}_3\} = \{\underline{y} \in \mathbb{R}^n : \underline{y} = w_1 \underline{X}_1 + w_2 \underline{X}_2 + w_3 \underline{X}_3 \text{ for some } w_1, w_2, w_3 \in \mathbb{R}\}$$

If $X = [\underline{X}_1 \ \ \underline{X}_2 \ \ \underline{X}_3]$, then $\operatorname{range}(X) = \operatorname{Span}(\operatorname{cols}(X))$

$$\underset{n\times p}{X} = \left.\underbrace{\begin{bmatrix} \quad & \quad & \quad \\ \quad & \quad & \quad \\ \quad & \quad & \quad \end{bmatrix}}_{\text{customer}}\right\}\text{movie} = \underset{= r \text{ representative taste profiles}}{\begin{bmatrix} U \\ n\times r \end{bmatrix}} \underbrace{\begin{bmatrix} V \\ r\times p \end{bmatrix}}_{r \text{ weights for each user}}$$

$\Longrightarrow$ model: each user's taste profile lies in a <u>subspace</u> spanned by the columns of U

Ex. $\mathbb{R}^n$ is a subspace

How to represent a subspace?

- as the span of a set of vectors (can be hard to interpret, hard to compute with, redundant)
- as the span of a set of linearly independent vectors (called subspace basis)
- as the span of a set of orthonormal vectors (called subspace orthonormal basis)
  (often people say orthogonal basis or orthobasis)

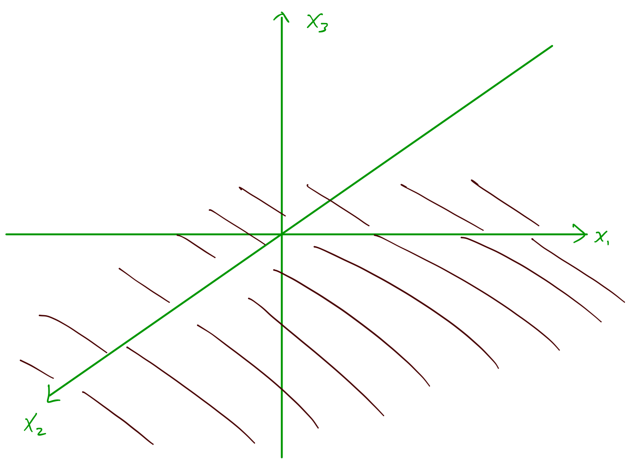
$\mathcal{S} = \text{span}\{\underline{u}_1, \underline{u}_2, \dots, \underline{u}_r\}$ where the $\underline{u}_i$'s are orthonormal

- orthogonal: $\underline{u}_i^T \underline{u}_j = 0$ if $i \neq j$
- normal: $\underline{u}_i^T \underline{u}_i = \|\underline{u}_i\| = 1$ for all i

Basis matrix $= U = [\underline{u}_1 \quad \underline{u}_2 \quad \cdots \quad \underline{u}_r]$

dimension of subspace $\dim(\mathcal{S}) = r$ = # vectors in basis of subspace

Ex. $\mathcal{S} = \{\underline{x} \in \mathbb{R}^n : x_3 = 0\}$



all $\underline{x} \in \mathcal{S}$ have the form $\underline{x} = \begin{bmatrix} \alpha \\ \beta \\ 0 \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$

basis $= \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$, $\dim(\mathcal{S}) = 2$

basis matrix $U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\text{rank}(U) = 2$ = subspace dimension

Properties of the orthonormal basis matrix $U = \begin{bmatrix} | & | & & | \\ \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_r \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times r}$

① $U^T U = I_{r \times r}$.

Proof: recall $U^T U = \begin{bmatrix} \underline{u}_1^T \underline{u}_1 & \underline{u}_1^T \underline{u}_2 & \cdots & \underline{u}_1^T \underline{u}_r \\ \underline{u}_2^T \underline{u}_1 & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \underline{u}_r^T \underline{u}_1 & & & \underline{u}_r^T \underline{u}_r \end{bmatrix}$ and $\underline{u}_i^T \underline{u}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$

② "Length preserving": for any vector $\underline{v} \in \mathbb{R}^r$, $U\underline{v}$ has length $\|U\underline{v}\| = \|\underline{v}\|$

Proof: $\|U\underline{v}\|_2^2 = \sum_{i=1}^{r} (v_i \underline{u}_i)^2$

squared length of $U\underline{v}$

$= (U\underline{v})^T (U\underline{v})$

$= \underline{v}^T U^T U \underline{v}$

$= \underline{v}^T \underline{v} = \|\underline{v}\|^2$ ⟵ squared length of $\underline{v}$.

How many LI vectors can lie in $\mathbb{R}^n$?

Let $\underline{e}_i \in \mathbb{R}^n$ be the length-n vector with all zeros except a 1 in the i^{th} location

i.e., the i^{th} column of the $n \times n$ identity matrix $I_{n \times n}$

These are called the canonical vectors.

Note that $\{\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n\}$ form a basis for $\mathbb{R}^n$ — any point in $\mathbb{R}^n$ can be written as a weighted sum of the e_i's, and they are all LI.

Also, we cannot have more than n LI vectors in $\mathbb{R}^n$ because if we consider the set

$\{\underline{e}_1, \underline{e}_2, \ldots, \underline{e}_n, \underline{a}\}$ for any $\underline{a}$, we can write $\underline{a}$ as a weighted sum of the $\underline{e}_i$'s

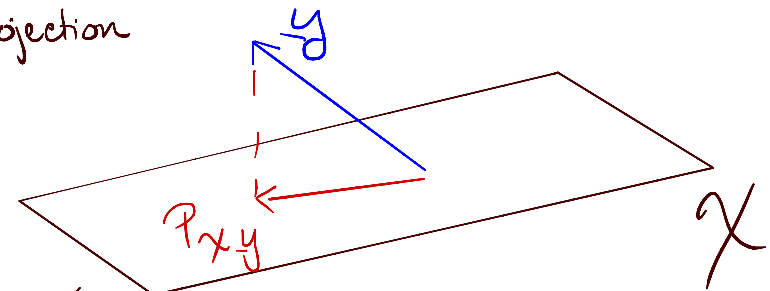
$$\underline{a} = a_1 \underline{e}_1 + a_2 \underline{e}_2 + \cdots + a_n \underline{e}_n$$

e.g. $$\begin{bmatrix} 2 \\ \pi \\ 10 \end{bmatrix} = 2 \cdot \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \pi \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + 10 \cdot \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Projection

The projection of a point y onto a set is the point in the set closest to y:

Let $\mathcal{X}$ be a set of points, and $P_{\mathcal{X}}\, y$ the projection



$$\hat{y} = P_{\mathcal{X}}\, y := \underset{\underline{x} \in \mathcal{X}}{\operatorname{argmin}} \|\underline{x} - \underline{y}\|_2^2$$

If $\mathcal{X}$ is a subspace spanned by the p columns of X,

then $\hat{y} = w_1 \underline{X}_1 + \cdots + w_p \underline{X}_p$ for some $w_1, \ldots, w_p$

$\Longrightarrow$ to find $\hat{y}$, 1st find w_i's, then compute $\hat{y} = X\hat{\underline{w}}$

$$\hat{y} = X\hat{\underline{w}}\,, \quad \hat{\underline{w}} = \underset{\underline{w}}{\operatorname{argmin}} \|X\underline{w} - y\|_2^2 \quad \text{— LEAST SQUARES!}$$

When the columns of X are linearly independent, then we know $\hat{\underline{w}} = (X^T X)^{-1} X^T y$

Therefore, $\hat{y} = X\hat{\underline{w}} = \underbrace{X(X^T X)^{-1} X^T}\, y$

This is called a
PROJECTION MATRIX,
denoted P_X

Properties of P_X:

- square
- $P_X^2 = P_X$

Orthogonal Subspace Bases and Least Squares

Let $X \in \mathbb{R}^{n \times p}$, $n \geq p$, X full-rank (p cols are linearly independent)

$\underline{y} \in \mathbb{R}^n$

$\underline{U}_1, \underline{U}_2, \ldots, \underline{U}_p$ be orthonormal basis vectors for the subspace $\mathcal{S} = \text{span}(\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p)$

Then $\hat{y} = X\hat{\underline{w}}$ where $\hat{\underline{w}} = \underset{\underline{w}}{\text{argmin}} \| y - X\underline{w} \|_2^2$ is given by $\hat{y} = UU^T y$ where $U = [\underline{U}_1, \ldots, \underline{U}_p]$

Proof:

$$\hat{y} = \underbrace{X(X^TX)^{-1}X^T}_{\substack{\text{projection matrix} \\ P_X - \text{projects vectors} \\ \text{onto subspace } \mathcal{S}}} y$$

now recall span (cols of X) = span(cols of U) $\Rightarrow P_X y = P_U y$.

but $P_U = U(U^TU)^{-1}U^T = UU^T$

$\Rightarrow \hat{y} = UU^T y$ $\longleftarrow$ no matrix inverse required!

Ex: consider set $\{ x \in \mathbb{R}^3 : x_3 = 5 \} = S$

$\underline{x}_1 = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}, \quad \underline{x}_2 = \begin{bmatrix} -3 \\ 4 \\ 5 \end{bmatrix} \Rightarrow \underline{x}_1, \underline{x}_2 \in S$

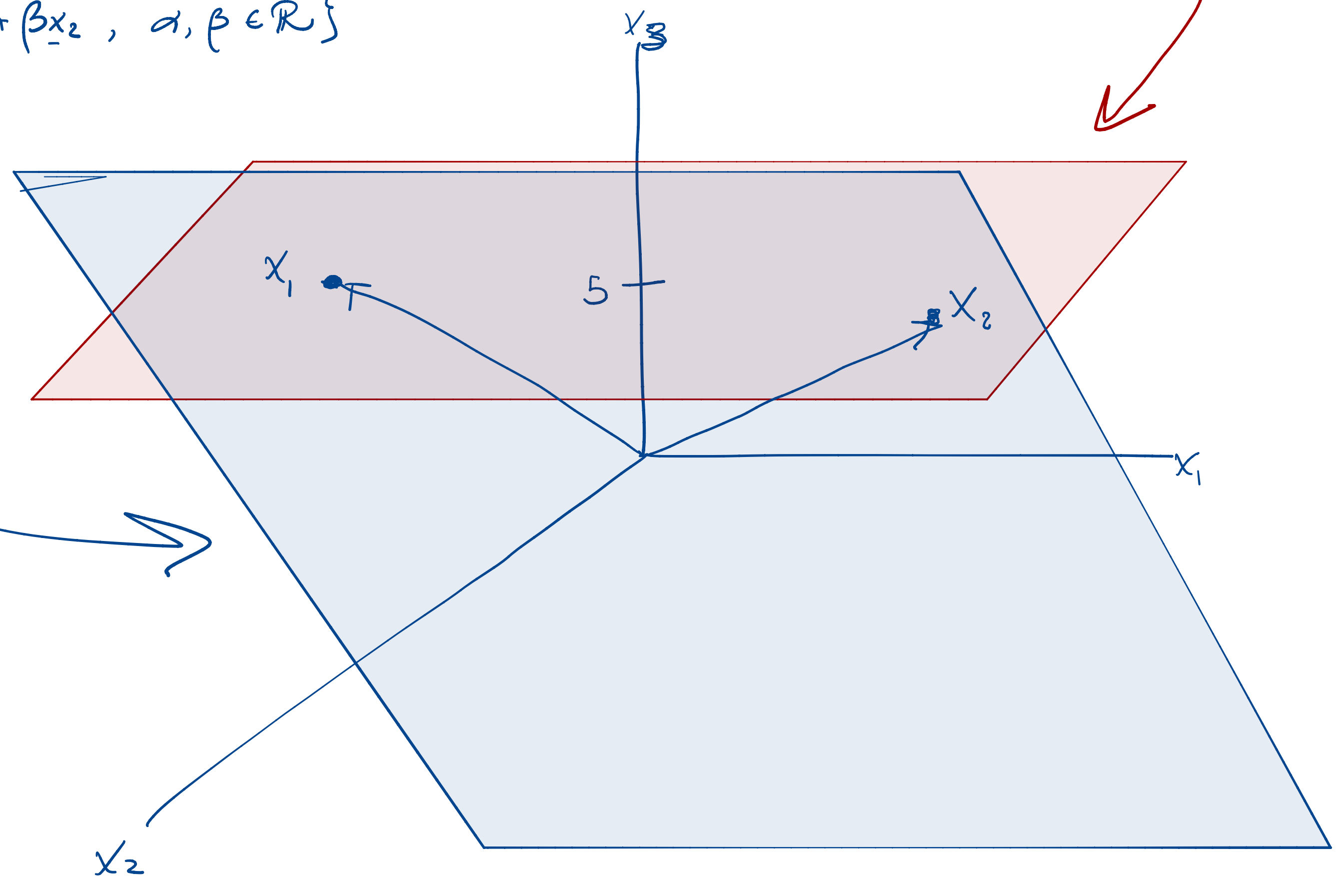
for S to be a subspace, all weighted sums of $\underline{x}_1$ and $\underline{x}_2$ must be in the subspace

including $0 \cdot \underline{x}_1 + 0 \underline{x}_2 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \notin S \Rightarrow S$ is not a subspace

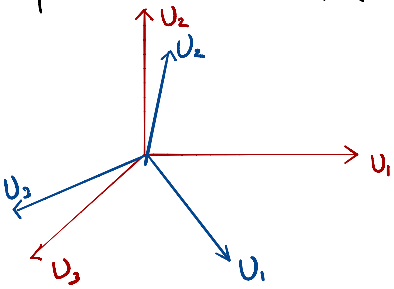
$\text{Span}(\underline{x}_1, \underline{x}_2)$ = set of all $\underline{x} = \alpha \underline{x}_1 + \beta \underline{x}_2$ for some α, β

$= \{ \underline{x} = \alpha \underline{x}_1 + \beta \underline{x}_2 , \ \alpha, \beta \in \mathbb{R} \}$

$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in \text{Span}(\underline{x}_1, \underline{x}_2)$



$p = 3 \Rightarrow$ basis matrix U must be 3×3



$$U = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I \Rightarrow U^T U = U U^T = I$$

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}$$ is orthonormal basis / U is orthogonal matrix $\Rightarrow U^T U = U U^T = I$

If U is an orthonormal basis matrix $U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_p \\ | & | & & | \end{bmatrix}$, then know $u_i^T u_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$

$$U^T U = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & \cdots & u_1^T u_p \\ u_2^T u_1 & u_2^T u_2 & \cdots & u_2^T u_p \\ \vdots & \vdots & \ddots & \vdots \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & 0 & \ddots & \\ & & & 1 \end{bmatrix} = I \Rightarrow U^T = U^{-1}$$

The above is for when U is a basis matrix for $\mathbb{R}^p$

If instead, U was a basis matrix for an r-dimensional subspace in $\mathbb{R}^p$,
then $U \in \mathbb{R}^{p \times r}$ (not square)

if U is orthonormal basis matrix for the subspace, then

$$\underbrace{U^T}_{r \times p} \underbrace{U}_{p \times r} = \underbrace{I}_{r \times r} \neq \underbrace{U U^T}_{p \times p}$$