Lecture 6 : Finding Orthonormal Bases

We are in an n-dimensional space.

We have p points, $\underline{x}_1, \underline{x}_2, \ldots, \underline{x}_p \in \mathbb{R}^n$

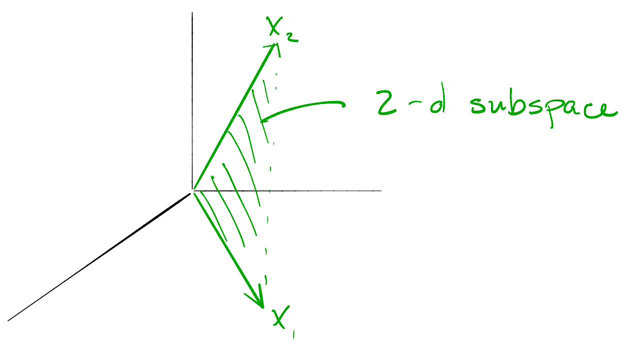
\# basis vectors ($\underline{u}_i$'s) = dim of subspace spanned by $\underline{x}_i$'s

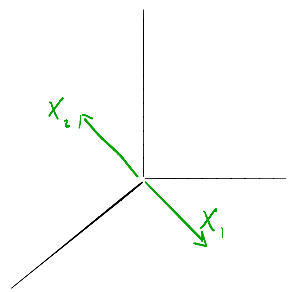
Case 1: $p \leq n$

Span$(x_1, \ldots, x_p)$ = a subspace with dimension p (or less)

it's p if the x_i's are linearly independent

otherwise, it's $< p$





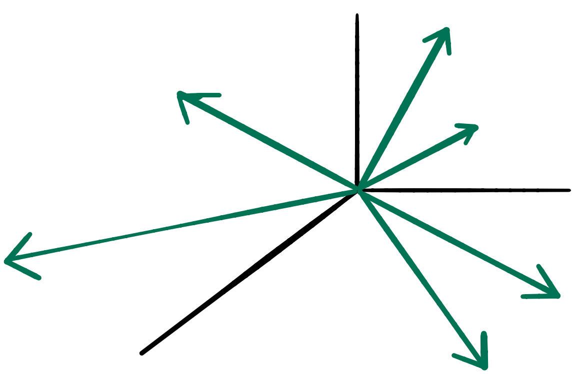
x_1, x_2 are Lin dependent

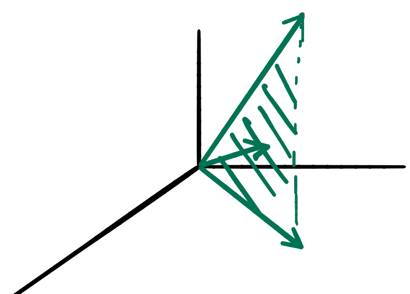
$\Rightarrow$ span a 1-dim subspace

Case 2: $n < p$

span$(\underline{x}_1, \ldots, \underline{x}_p)$ = a subspace with dimension n (or less)

$\Longrightarrow$ the x_i's are linearly dependent





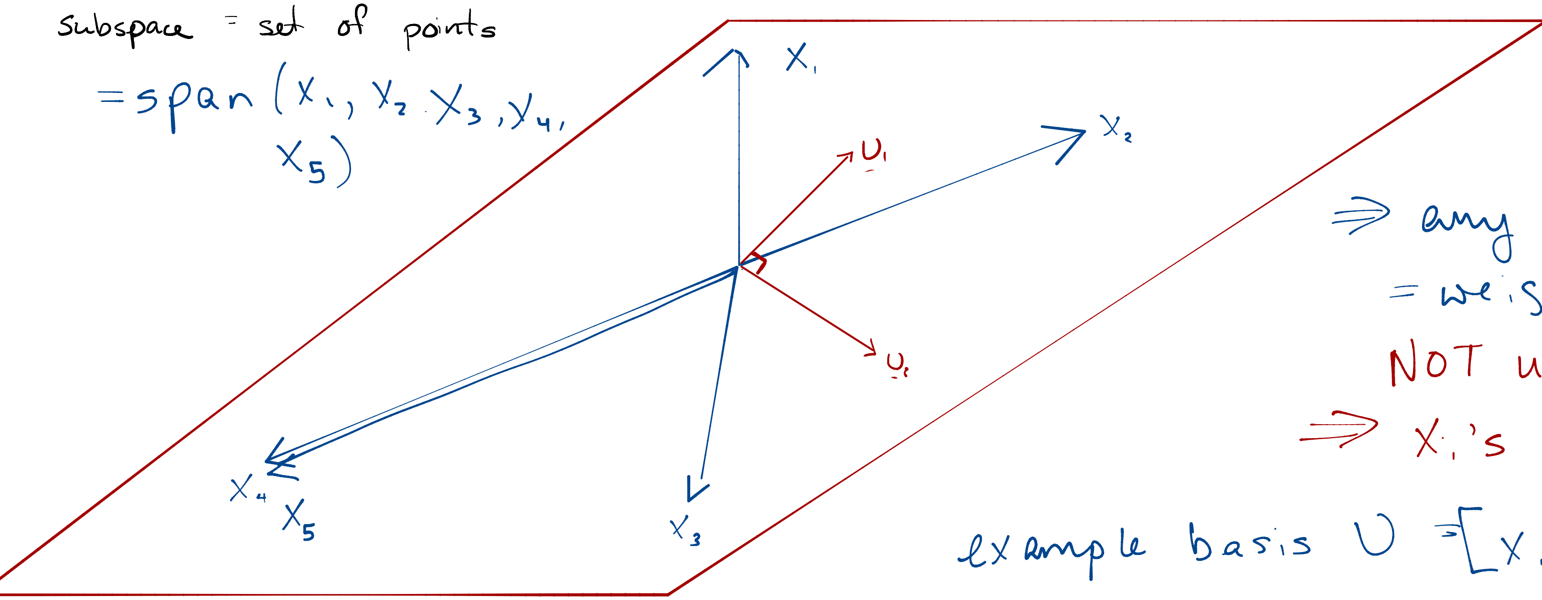
Question: given a set of vectors spanning a subspace, how to we find an orthonormal subspace basis?

A: Gram Schmidt orthogonalization (lec 6)

alt: can also use singular value decomposition (lec 7-8)

subspace = set of points

$= \text{span}(x_1, x_2, x_3, x_4, x_5)$



$\Rightarrow$ any pt in subspace

= weighted sum of x_i's

NOT unique

$\Rightarrow$ x_i's are not a basis

example basis $U = [x_1 \; x_2]$

$\text{span}(x_1, x_2) = \text{span}(x_1, \dots, x_5)$

x_1 and x_2 are not orthonormal

u_1 and u_2 are orthonormal and $\text{span}(u_1, u_2)$

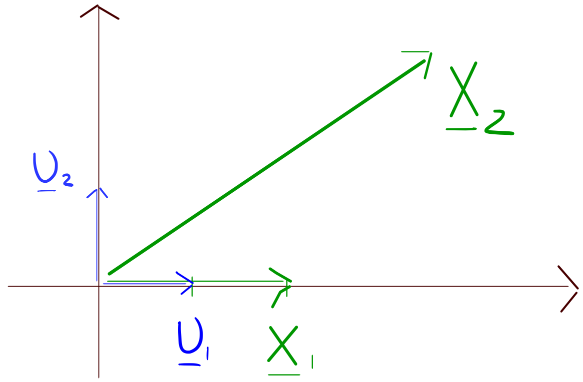
$= \text{span}(x_1, x_2)$

$\tilde{U} = [u_1 \; u_2]$ = orthonormal basis matrix for subspace

Gram - Schmidt Orthogonalization

A process for taking a generic set of vectors $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p$ and finding an orthonormal basis for their span.

Ex. $\underline{X}_1 = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, $\underline{X}_2 = \begin{bmatrix} a \\ b \end{bmatrix}$



$$\underline{U}_1 = \frac{\underline{X}_1}{\|\underline{X}_1\|_2} = \frac{\underline{X}_1}{2} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

write $\underline{X}_2$ as a weighted $\underline{U}_1$ + residual (making residual as small as possible)

$$\underline{X}_2 = a \cdot \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\underline{U}_1} + \begin{bmatrix} 0 \\ b \end{bmatrix} \Rightarrow \text{resid } \underline{X}_2' = \begin{bmatrix} 0 \\ b \end{bmatrix} \Longrightarrow \underline{U}_2 = \frac{\underline{X}_2'}{\|\underline{X}_2'\|_2} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Ex. $\underline{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$, $\underline{x}_2 = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$

$$\underline{u}_1 = \frac{\underline{x}_1}{\|\underline{x}_1\|_2} = \frac{\underline{x}_1}{\sqrt{2}} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$$

now write $\underline{x}_2$ as weighted $\underline{u}_1$ + residual, where residual is as small as possible (smallest norm $\|\cdot\|_2$)

Q: what is the projection of $\underline{x}_2$ onto $\underline{u}_1$?

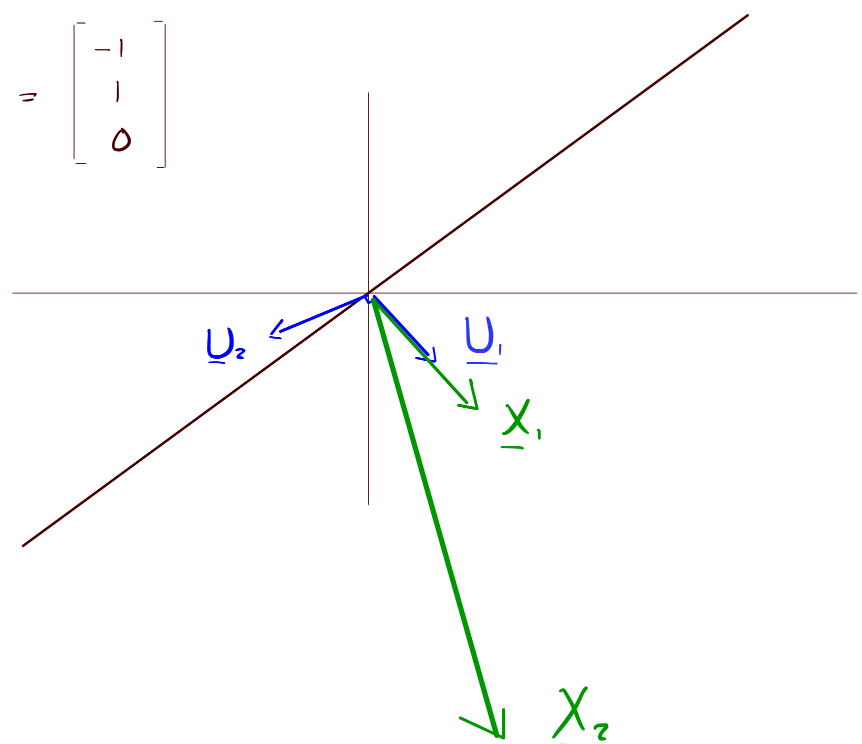
$$\hat{w} = \underset{w}{\mathrm{argmin}} \; \|\underline{x}_2 - \underline{u}_1 w\|_2^2 = \underline{u}_1^T \underline{x}_2, \quad P_{\underline{u}_1} \underline{x}_2 = \underline{u}_1 \underline{u}_1^T \underline{x}_2$$

$$\Rightarrow \underline{x}_2 = \underline{u}_1 \hat{w} + \text{resid} = \underline{u}_1 (\underline{u}_1^T \underline{x}_2) + \text{resid.} \Rightarrow \text{resid} = \underline{x}_2' = \underline{x}_2 - \underline{u}_1 (\underline{u}_1^T \underline{x}_2)$$

$$\underline{u}_1^T \underline{x}_2 = 1/\sqrt{2} + 3/\sqrt{2} = 4/\sqrt{2}$$

$$\underline{x}_2' = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \frac{4}{\sqrt{2}} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

$$\underline{u}_2 = \frac{\underline{x}_2'}{\|\underline{x}_2'\|_2} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}$$



Ex $\quad \underline{x}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \quad \underline{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix}$

$$\underline{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} / \sqrt{5}$$

want to write

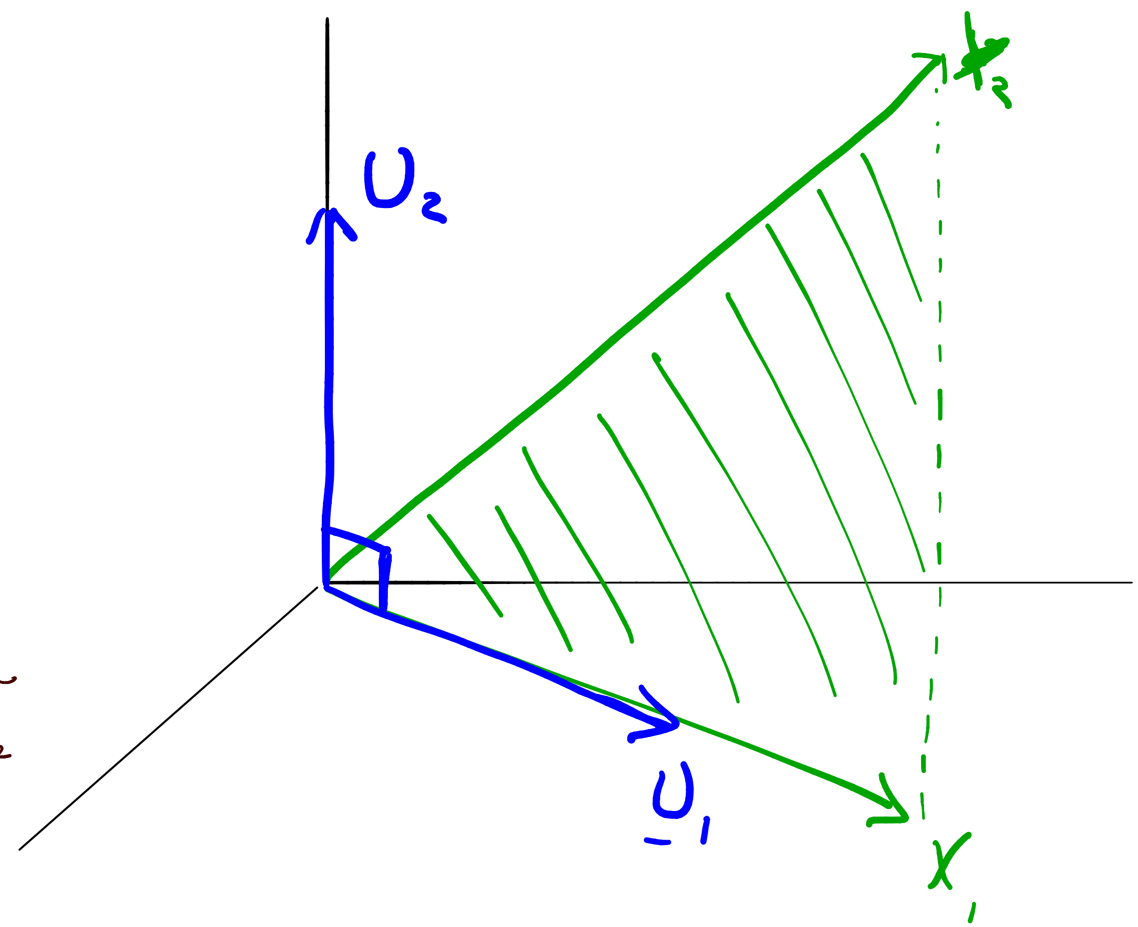
$$\underline{x}_2 = \underline{u}_1 v_1 + \underline{u}_2 v_2$$

$$v_1 = \underline{u}_1^T \underline{x}_2 = [2 \ 1 \ 0] \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} / \sqrt{5}$$

$$= \frac{5}{\sqrt{5}} = \sqrt{5}$$

$$\Rightarrow \underline{x}_2 - \underline{u}_1 v_1 = \begin{bmatrix} 2 \\ 1 \\ 5 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\underline{u}_2} \cdot \underbrace{5}_{v_2}$$

unit norm



Gram-Schmidt Orthogonalization Algorithm

input: $X = [\underline{x}_1, \underline{x}_2, \cdots, \underline{x}_p] \in \mathbb{R}^{n \times p}$

output: $U = [\underline{u}_1, \underline{u}_2, \ldots, \underline{u}_r] \in \mathbb{R}^{n \times r}$, $\quad r = \text{rank}(X) \leq \min(n,p)$

$= \dim(\mathcal{S})$ if $\mathcal{S}$ is subspace spanned by $\underline{x}_i$'s

Preprocess: delete any $\underline{x}_i = \underline{0}$

when $j = 2$

$$x_2' = x_2 - \underbrace{(u_1^T x_2)}_{v_1} u_1$$

Initialize: $\underline{u}_1 = \underline{x}_1 / \|\underline{x}_1\|_2$

For $j = 2, 3, \ldots, p$

$\underline{x}_j' = \underline{x}_j$ − best representation of $\underline{x}_j$ as weighted sum of $\underline{u}_1, \ldots, \underline{u}_{j-1}$

$$= \underbrace{\underline{x}_j - \sum_{i=1}^{j-1} (\underline{u}_i^T \underline{x}_j) \underline{u}_i}$$

(if $\hat{U}_{j-1} = [\underline{u}_1, \ldots, \underline{u}_{j-1}]$, then $\underline{x}_j' = \underline{x}_j - \underbrace{\hat{U}_{j-1} \hat{U}_{j-1}^T \underline{x}_j}$)

projection of $\underline{x}_j$ onto space spanned by cols of $\hat{U}_{j-1}$

$$\underline{u}_j = \begin{cases} \underline{x}_j' / \|\underline{x}_j'\|_2 & \text{if } \underline{x}_j' \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

end

remove zero-valued $\underline{u}_j$'s.

Ex. $X = \begin{bmatrix} -1 & 1 & 2 \\ 1 & -1 & 2 \end{bmatrix}$

$$\underline{U}_1 = \frac{\underline{X}_1}{\|\underline{X}_1\|_2} = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$\underline{X}_2' = \underline{X}_2 - (\underline{U}_1^T \underline{X}_2)\,\underline{U}_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} - (-2/\sqrt{2}) \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longrightarrow \underline{U}_2 = 0$$

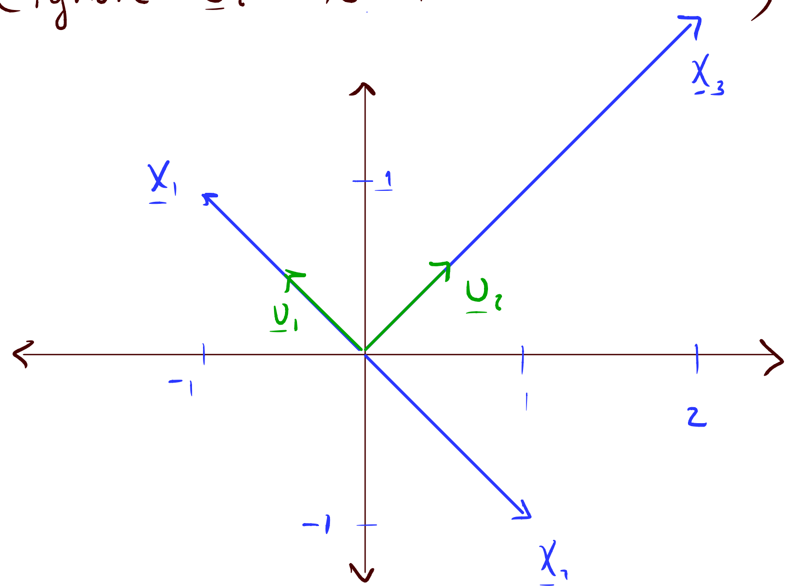
$$\underline{X}_3' = \underline{X}_3 - (\underline{U}_1^T \underline{X}_3)\,\underline{U}_1 + (\underline{U}_2^T \underline{X}_3)\,\underline{U}_2$$

$$= \underline{X}_3 - 0 - 0$$

$$\Rightarrow \underline{U}_3 = \frac{\underline{X}_3}{\|\underline{X}_3\|_2} = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$\Rightarrow U = [\underline{U}_1 \;\; \underline{U}_3]$ (ignore $\underline{U}_2$ b/c it is zero vector)

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



Gram-Schmidt finds a basis spanning a set of vectors $\underline{X}_1, \dots, \underline{X}_n$

The first G-S basis vector is $\underline{U}_1 = \underline{X}_1 / \|X_1\|_2$ — that is, if we shuffle the $\underline{X}_i$'s, we will get a different basis.

The Singular Value Decomposition is an alternative tool for finding the basis for a set of points. It is invariant to the order of the points.

Basic idea: $\underline{U}_1$ is the 1d subspace that is closest to all the $\underline{X}_i$'s (ie. best 1d subspace fit)

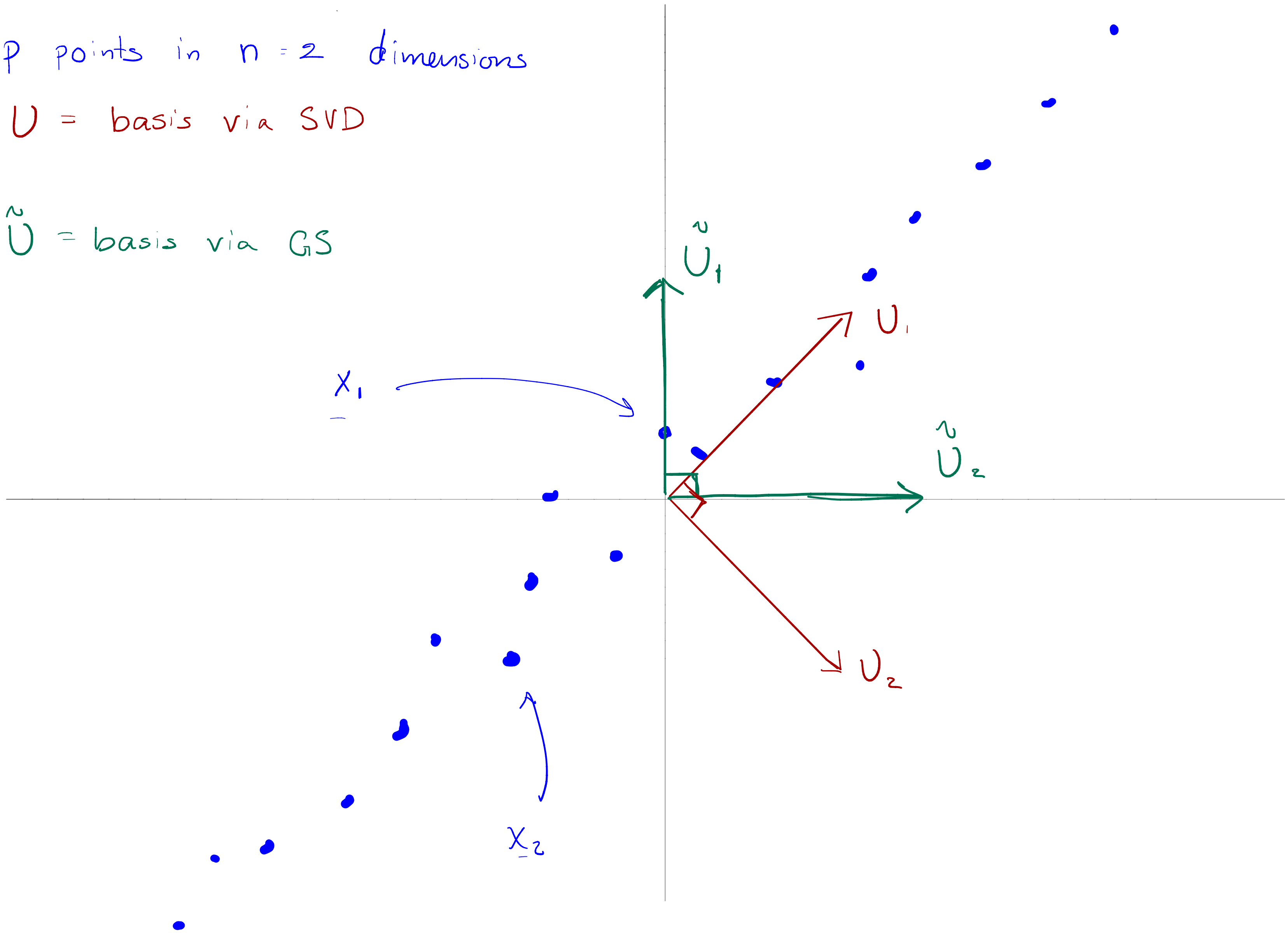
$\underline{U}_2$ is the 1d subspace that is closest to all the residuals, $\underline{X}_i - \underline{U}_1 \underline{U}_1^T \underline{X}_1$

$\underline{U}_3$ is the 1d subspace that is closest to all the residuals $\underline{X}_i - \underline{U}_1 \underline{U}_1^T \underline{X}_1 - \underline{U}_2 \underline{U}_2^T \underline{X}_i$

p points in $n = 2$ dimensions

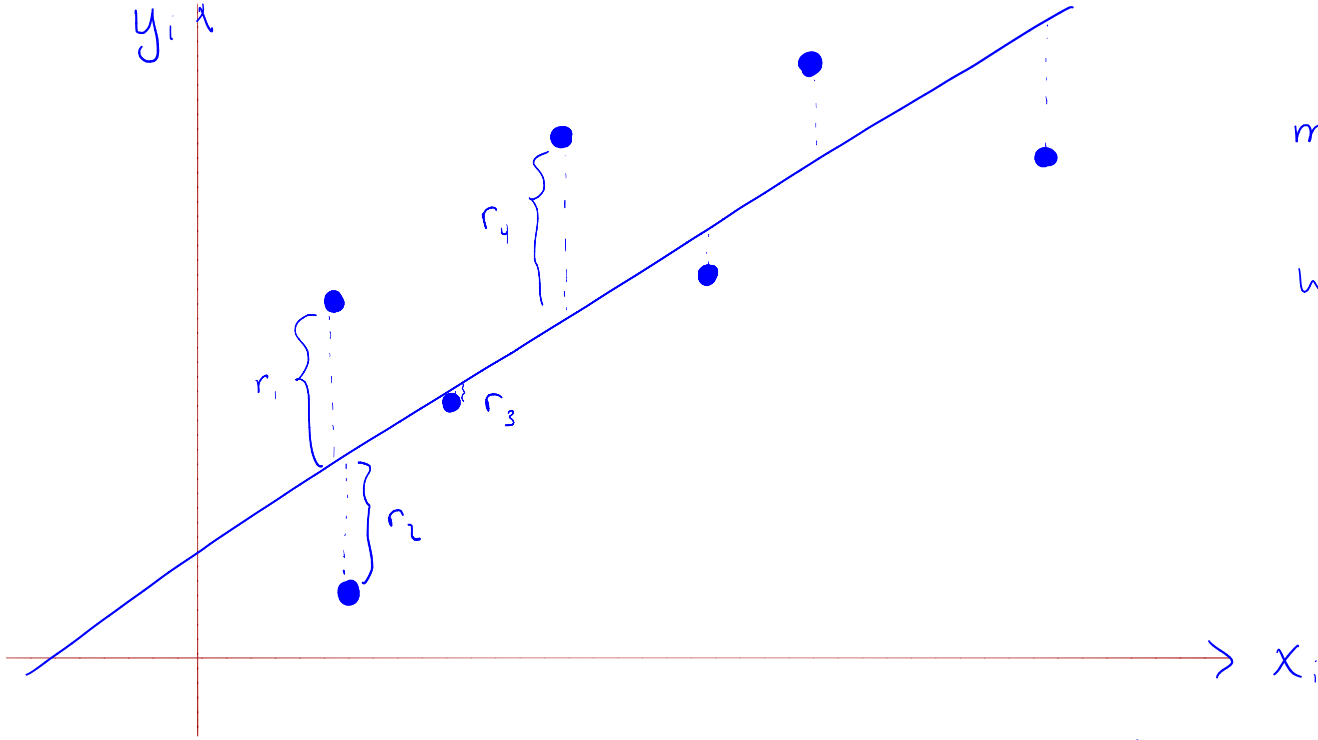
U = basis via SVD

$\tilde{U}$ = basis via GS



$x_i \in \mathbb{R}$

L.S.



$$\text{minimize} \sum_i r_i^2$$

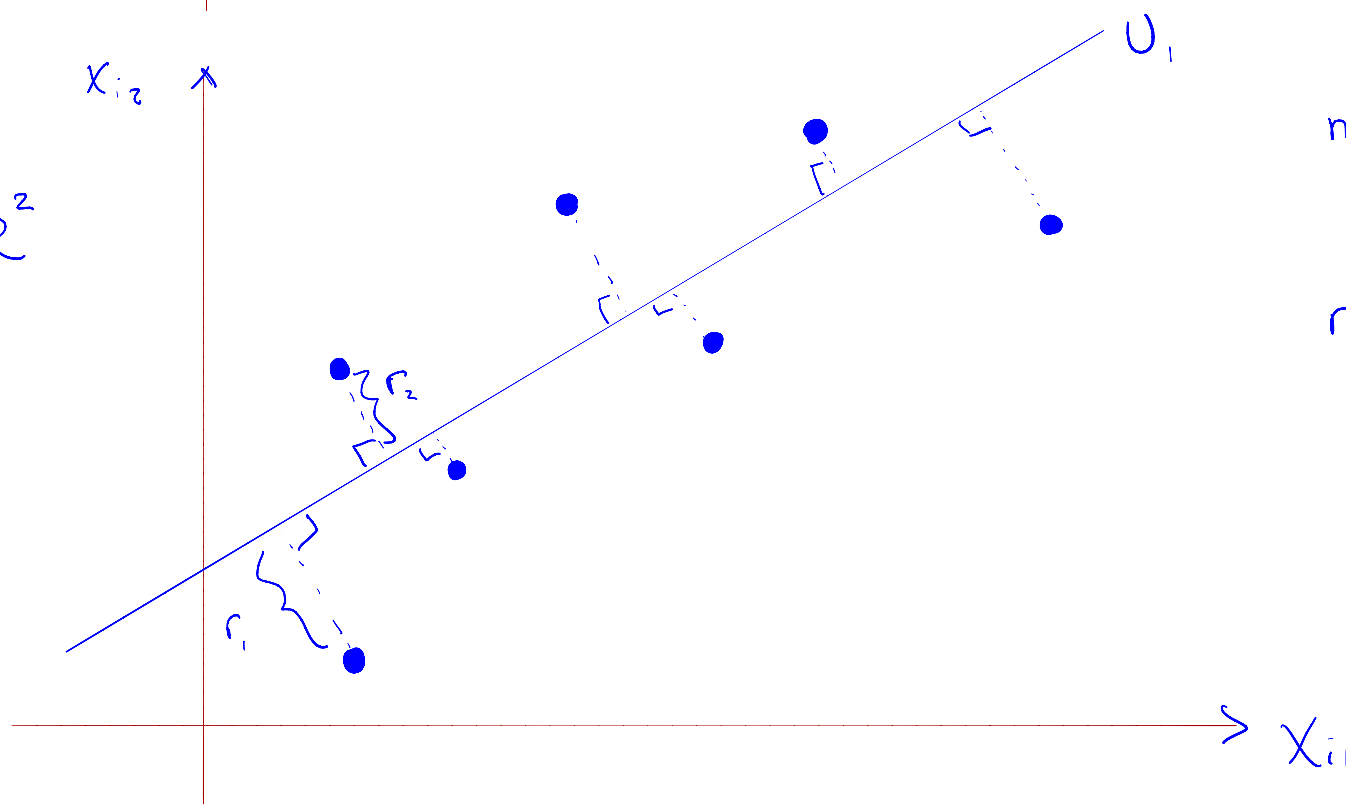
write

$$y_i \approx \underbrace{a x_i + b}_{\underline{w}} \quad \leftarrow$$

$$r_i = y_i - (a x_i + b)$$

$x_i \in \mathbb{R}^2$

SVD



$$\text{minimize} \sum_i \|r_i\|_2^2$$

$$r_i = \underline{x}_i - P_{U_1} \underline{x}_i$$