IFT 6760A - Lecture 2
SVD and Orthogonality

Scribe(s): Yann BOUTEILLER, Parviz HAGGI **Instructor:** Guillaume Rabusseau

1 Summary

In the previous lecture we reviewed the basics of linear algebra. In this lecture we will first look into the concept of Singular Value Decomposition of a matrix, often called SVD, followed by the concepts of orthogonality and orthogonal projections.

2 Singular Value Decomposition

Theorem 1 (SVD)**.** *Any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ can be written as:*

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$$

where $\mathbf{U} \in \mathbb{R}^{m \times m}$ and $\mathbf{V} \in \mathbb{R}^{n \times n}$ are both orthogonal matrices (i.e. $\mathbf{U}^T\mathbf{U} = \mathbf{I_m}$ and $\mathbf{V}^T\mathbf{V} = \mathbf{I_n}$) and $\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ is a diagonal rectangular matrix such that $\mathbf{\Sigma}_{i,i} \neq 0$ if and only if $i \leq rank(\mathbf{A})$

The form $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is called the Singular-Value Decomposition (SVD) of $\mathbf{A}$.

Property 2. *Let r be the rank of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$. Let $\mathbf{U}$ and $\mathbf{V}$ be the matrices from the SVD of $\mathbf{A}$ with columns $\mathbf{u}_1, \dots, \mathbf{u}_m \in \mathbb{R}^m$ and $\mathbf{v}_1, \dots, \mathbf{v}_m \in \mathbb{R}^m$ respectively. Then the following hold:*

- *$\mathbf{u}_1, \dots, \mathbf{u}_r$ forms a basis a basis of $\mathcal{R}(\mathbf{A})$*
- *$\mathbf{u}_{r+1}, \dots, \mathbf{u}_m$ forms a basis of $\mathcal{N}(\mathbf{A}^T)$*
- *$\mathbf{v}_1, \dots, \mathbf{v}_r$ forms a basis a basis of $\mathcal{R}(\mathbf{A}^T)$*
- *$\mathbf{v}_{r+1}, \dots, \mathbf{v}_n$ forms a basis of $\mathcal{N}(\mathbf{A})$*

Definition 3 (Truncated SVD)**.** *Theorem 1 can be rewritten as:*

$$\underset{(m\times n)}{\mathbf{A}} = \underset{(m\times m)}{\mathbf{U}}\underset{(m\times n)}{\mathbf{\Sigma}}\underset{(n\times n)}{\mathbf{V}^T} = \begin{bmatrix} \underset{(m\times r)}{\widetilde{\mathbf{U}}} & \underset{(m\times m-r)}{\widetilde{\mathbf{U}}_\perp} \end{bmatrix} \begin{bmatrix} \underset{(r\times r)}{\mathbf{\Lambda}} & \underset{(r\times n-r)}{\mathbf{0}} \\ \underset{(m-r\times r)}{\mathbf{0}} & \underset{(m-r\times n-r)}{\mathbf{0}} \end{bmatrix} \begin{bmatrix} \underset{(r\times n)}{\widetilde{\mathbf{V}}^T} \\ \underset{(n-r\times n)}{\widetilde{\mathbf{V}}_\perp^T} \end{bmatrix} = \underset{(m\times r)}{\widetilde{\mathbf{U}}}\underset{(r\times r)}{\mathbf{\Lambda}}\underset{(r\times n)}{\widetilde{\mathbf{V}}^T}$$

Where:
$\mathbf{\Lambda} \in \mathbb{R}^{r \times r}$ is diagonal with $\mathbf{\Lambda}_{i,i} \neq 0$
$\widetilde{\mathbf{U}}^T\widetilde{\mathbf{U}} = \widetilde{\mathbf{V}}^T\widetilde{\mathbf{V}} = \mathbf{I_r}$ ($\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ are both orthogonal matrices)

The form $\mathbf{A} = \widetilde{\mathbf{U}}\mathbf{\Lambda}\widetilde{\mathbf{V}}^T$ is called the Truncated Singular-Value Decomposition (Truncated SVD) of $\mathbf{A}$.

Remark : In truncated SVD, $\widetilde{\mathbf{U}}$ is such that $\widetilde{\mathbf{U}}^T\widetilde{\mathbf{U}} = \mathbf{I_r}$. However, it is not true in general that $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T = \mathbf{I_m}$. If $r < m$, we actually know that $\widetilde{\mathbf{U}}\widetilde{\mathbf{U}}^T \neq \mathbf{I_m}$ because $\mathbf{I_m}$ is of rank m, and cannot be linearly generated by $\widetilde{\mathbf{U}}$ of rank $r < m$.

$\mathbf{U}$ is orthogonal, so we have:

$$\underset{(m\times m)}{\mathbf{U}^T\mathbf{U}} = \begin{bmatrix} \underset{(R\times m)}{\widetilde{\mathbf{U}}^T} \\ \underset{(m-R\times m)}{\widetilde{\mathbf{U}}_\perp^T} \end{bmatrix} \begin{bmatrix} \underset{(m\times R)}{\widetilde{\mathbf{U}}} & \underset{(m\times m-R)}{\widetilde{\mathbf{U}}_\perp} \end{bmatrix} = \begin{bmatrix} \underset{(r\times r)}{\widetilde{\mathbf{U}}^T\widetilde{\mathbf{U}}} & \underset{(r\times m-r)}{\widetilde{\mathbf{U}}^T\widetilde{\mathbf{U}}_\perp} \\ \underset{(m-r\times r)}{\widetilde{\mathbf{U}}_\perp^T\widetilde{\mathbf{U}}} & \underset{(m-r\times m-r)}{\widetilde{\mathbf{U}}_\perp^T\widetilde{\mathbf{U}}_\perp} \end{bmatrix} = \begin{bmatrix} \underset{(r\times r)}{\mathbf{I_r}} & \underset{(r\times m-r)}{\mathbf{0}} \\ \underset{(m-r\times r)}{\mathbf{0}} & \underset{(m-r\times m-r)}{\mathbf{I_{m-r}}} \end{bmatrix}$$

The same holds for $\mathbf{V}^T\mathbf{V}$ with n instead of m.

3 Orthogonality and Projections

Definition 4 (The dot product)**.** *The dot product on $\mathbb{R}^n$ has the following properties:*

- *The dot product $\langle \mathbf{a}, \mathbf{b}\rangle := \mathbf{a}^T\mathbf{b}$ is an inner product on $\mathbb{R}^n$.*
- *Two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ are orthogonal iff $\langle \mathbf{a}, \mathbf{b}\rangle = 0$.*
- *The inner product induces a norm on $\mathbb{R}^n$: $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$.*
- *A basis $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k \in \mathbb{R}^n$ of a sub-space $\mathcal{U}$ is orthonormal iff* $\begin{cases} \langle u_i, u_j\rangle = 0 & \forall\ i \neq j \\ \sqrt{\langle u_i, u_i\rangle} = \|u_i\| = 1 & \forall\ i \end{cases}$

Definition 5 (Orthogonal matrix)**.** *A matrix $\mathbf{U} \in \mathbb{R}^{n\times k}$ is orthogonal if its columns form an orthonormal basis of $\mathcal{R}(\mathbf{U})$. Equivalently $\mathbf{U}$ is orthogonal if and only if $\mathbf{U}^T\mathbf{U} = \mathbf{I}$.*

Note: $\mathbf{U}\mathbf{U}^T$ is not necessarily equal to the identity matrix: $\mathbf{U}\mathbf{U}^T \neq \mathbf{I}$. However, if the matrix $\mathbf{U}$ is square i.e. $\mathbf{U} \in \mathbb{R}^{n\times n}$, then $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ implies $\mathbf{U}\mathbf{U}^T = \mathbf{I}$

Remark: if $\mathbf{U} \in \mathbb{R}^{n\times k}$ is orthogonal, then $\|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$. This is easily seen that this follows from the orthogonality of the matrix $\mathbf{U}$: $\|\mathbf{U}\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{U}^T\mathbf{U}\mathbf{x} = \mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|^2$.

Property 6. *If $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k \in \mathbb{R}^n$ is an orthonormal basis of a sub-space $\mathcal{U}$,*

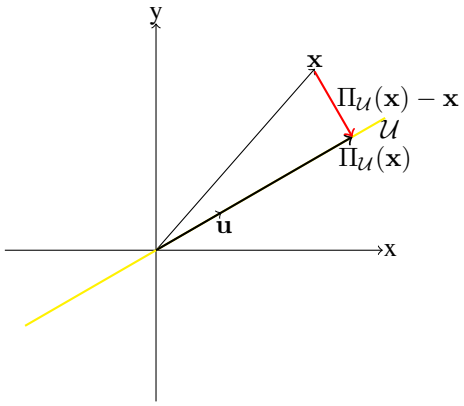
$$\mathbf{U} = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n\times k} \Rightarrow \forall \mathbf{x} \in \mathcal{U},\ \mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{x}$$

Consequently $\|\mathbf{x}\| = \|\mathbf{U}^T\mathbf{x}\|$.

Note: Even though $\mathbf{U}\mathbf{U}^T$ is not the identity matrix, it acts as such on the sub-space $\mathcal{U}$.

Proof. A vector $\mathbf{x}$ in the sub-space $\mathcal{U}$ can, by definition, be written as a linear combination of its basis vectors.
$\mathbf{x} \in \mathcal{U} \Rightarrow \exists\ \mathbf{a} \in \mathbb{R}^k\ : \mathbf{x} = \mathbf{U}\mathbf{a}$
To prove that $\mathbf{U}\mathbf{U}^T$ acts as an identity matrix, we look at what it does to $\mathbf{x}$. Using the fact that $\mathbf{x}$ can be written as a linear combination of the basis vectors of the sub-space $\mathcal{U}$, and that $\mathbf{U}$ is defined as an orthogonal matrix, we have:
$\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{U}\mathbf{a} = \mathbf{U}\mathbf{a} = \mathbf{x}$.
A consequence of this property is that $\mathbf{x}$ and $\mathbf{U}^T\mathbf{x}$ are vectors of the same size:
$\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = (\mathbf{U}\mathbf{U}^T\mathbf{x})^T\mathbf{x} = \mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{x} = (\mathbf{U}^T\mathbf{x})^T(\mathbf{U}^T\mathbf{x}) = \|\mathbf{U}^T\mathbf{x}\|^2$ □

Figure 1: Orthogonal projection of a vector $\mathbf{x}$ onto a sub-space $\mathcal{U}$



Definition 7 (Orthogonal Projection)**.** *Let $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k \in \mathbb{R}^n$ be an orthonormal basis of some sub-space $\mathcal{U}$ and*

$$U = \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_k \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{n \times k}$$

The orthogonal projection onto $\mathcal{U}$ is defined as:

$$\begin{aligned} \Pi_{\mathcal{U}} : &\mathbb{R}^n \to \mathbb{R}^n \\ &\mathbf{x} \to \mathbf{U}\mathbf{U}^T\mathbf{x} \end{aligned}$$

This is well defined and unique. In other words, if there is another orthogonal matrix $\mathbf{V} \in \mathbb{R}^{n \times k}$ such that $\mathcal{R}(\mathbf{V}) = \mathcal{U}$, then we can prove that the two projections are the same: $\mathbf{U}\mathbf{U}^T = \mathbf{V}\mathbf{V}^T$.

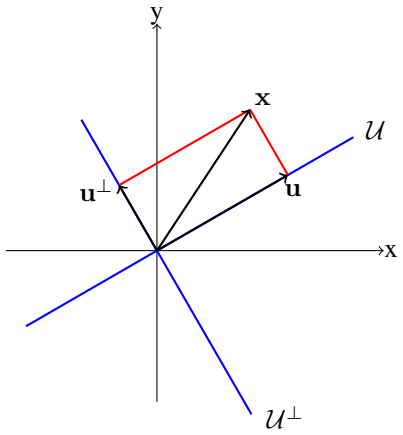
Proof. Since the columns of $\mathbf{U}$ and $\mathbf{V}$ consist of the orthonormal basis vectors of the sub-space $\mathcal{U}$, they span the same sub-space i.e. they have the same range: $\mathcal{R}(\mathbf{V}) = \mathcal{R}(\mathbf{U})$.
But if this is true, then there must exist a linear map $\mathbf{P}$ between the two matrices:
$\exists \mathbf{P} \in \mathbb{R}^{k \times k}$ such that $\mathbf{V} = \mathbf{U}\mathbf{P}$
But since the matrices $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, we have:
$\mathbf{I} = \mathbf{V}^T\mathbf{V} = \mathbf{P}^T\mathbf{U}^T\mathbf{U}\mathbf{P} = \mathbf{P}^T\mathbf{P}$
So we have proven that $\mathbf{P}$ is also orthogonal. Since $\mathbf{P}$ is a square matrix, we have that: $\mathbf{P}^T\mathbf{P} = \mathbf{P}\mathbf{P}^T = \mathbf{I}$
Inserting this identity matrix below, we prove that orthogonal projection is unique.
$\mathbf{V}\mathbf{V}^T = \mathbf{U}\mathbf{P}\mathbf{P}^T\mathbf{U}^T = \mathbf{U}\mathbf{U}^T$ □

Definition 8 (Orthogonal Complement)**.** *If $\mathcal{U}$ is a sub-space of $\mathbb{R}^n$, the orthogonal complement of $\mathcal{U}$ is defined as:* $\mathcal{U}^\perp = \{\mathbf{v} \in \mathbb{R}^n | \langle \mathbf{u}, \mathbf{v} \rangle = 0 \ \forall \mathbf{u} \in \mathcal{U}\}$.

Property 9. *$\mathcal{U}^\perp$ is a sub-space of $\mathbb{R}^n$ and $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{U}^\perp$.*

Proof. Let $\mathbf{u}, \mathbf{v} \in \mathcal{U}^\perp$. For any $\mathbf{w} \in \mathcal{U}$ we have $\langle \mathbf{w}, \mathbf{u} + \mathbf{v} \rangle = \langle \mathbf{w}, \mathbf{u} \rangle + \langle \mathbf{w}, \mathbf{v} \rangle = 0 + 0 = 0$, so $\mathbf{u} + \mathbf{v} \in \mathcal{U}^\perp$.
Let $\mathbf{u} \in \mathcal{U}^\perp$ and $\alpha \in \mathbb{R}$. Then for any $\mathbf{w} \in \mathcal{U}$ we have $\langle \mathbf{w}, \alpha\mathbf{u} \rangle = \alpha \langle \mathbf{w}, \mathbf{u} \rangle = \alpha 0 = 0$, so $\alpha\mathbf{u} \in \mathcal{U}^\perp$.
Therefore $\mathcal{U}^\perp$ is a sub-space.
We also know that zero must be a part of every sub-space. In fact, zero is the only common element of these two sub-spaces (Assuming that there is a non-zero vector $\mathbf{w}$ that belongs to both of the sub-spaces $\mathcal{U}$ and $\mathcal{U}^\perp$, using the above definition of $\mathcal{U}^\perp$, it is easily seen that the inner product of $\mathbf{w}$ with itself is 0 i.e. $\langle \mathbf{w}, \mathbf{w} \rangle = 0$ which is contradictory to the assumption that $\mathbf{w} \neq 0$.) A vector $\mathbf{x} \in \mathbb{R}^n$ can be written as $\mathbf{x} = \Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v}$. It can be proven (see property 10 below) that the vector $\mathbf{v} = \mathbf{x} - \Pi_{\mathcal{U}}(\mathbf{x})$ must be orthogonal to $\Pi_{\mathcal{U}}(x) \in \mathcal{U}$ and that it must, therefore, belong to the orthogonal complement of $\mathcal{U}$. So $\mathbb{R}^n = \mathcal{U} + \mathcal{U}^\perp$.

Figure 2: Splitting a vector into its components on two sub-spaces $\mathcal{U}$ and $\mathcal{U}^\perp$



Since 0 is the only vector in common of $\mathcal{U}$ and $\mathcal{U}_\perp$, $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{U}^\perp$, i.e. every vector in $\mathbb{R}^n$ can be decomposed uniquely into a sum of two vectors, one belonging to $\mathcal{U}$ and the other to $\mathcal{U}^\perp$. □

Property 10. *Based on the above definitions of orthogonality and orthogonal complement, we can summarize the following properties of orthogonal projection:*

- $\Pi_{\mathcal{U}}^2 = \Pi_{\mathcal{U}}$
- $\forall \mathbf{x}: \quad \langle \Pi_{\mathcal{U}}(\mathbf{x}), \Pi_{\mathcal{U}}(\mathbf{x}) - \mathbf{x} \rangle = 0$
- $Im(\Pi_{\mathcal{U}}) = \mathcal{U}$
- $Ker(\Pi_{\mathcal{U}}) = \mathcal{U}^\perp$
- $\forall \mathbf{x}: \quad \|\Pi_{\mathcal{U}}(\mathbf{x})\| \leq \|\mathbf{x}\|$

Proof. These properties can be proven as follows :

- $\Pi_{\mathcal{U}}^2 = \Pi_{\mathcal{U}}$:
  To prove this we use the definition of projection in terms of matrices: Knowing that $\mathbf{U}\mathbf{U}^T\mathbf{x}$ is a projection, we have: $\Pi_{\mathcal{U}}^2(\mathbf{x}) = \Pi_{\mathcal{U}}(\mathbf{U}\mathbf{U}^T\mathbf{x}) = \mathbf{U}\mathbf{U}^T\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{x} = \Pi_{\mathcal{U}}(\mathbf{x})$.
  Note that in this calculation we used $\mathbf{U}^T\mathbf{U} = \mathbf{I}$.

- $\forall \mathbf{x}: \quad \langle \Pi_{\mathcal{U}}(\mathbf{x}), \Pi_{\mathcal{U}}(\mathbf{x}) - \mathbf{x} \rangle = 0$:

$$\begin{aligned} \langle \Pi_{\mathcal{U}}(\mathbf{x}), \Pi_{\mathcal{U}}(\mathbf{x}) - \mathbf{x} \rangle &= \langle \mathbf{U}\mathbf{U}^T\mathbf{x}, \mathbf{U}\mathbf{U}^T\mathbf{x} - \mathbf{x} \rangle \\ &= \mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{U}\mathbf{U}^T\mathbf{x} - \mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{x} \\ &= 0 \end{aligned}$$

- $\text{Im}(\Pi_{\mathcal{U}}) = \mathcal{U}$:
  $\forall \mathbf{x} \in \mathbb{R}^n$ we have that $\Pi_{\mathcal{U}}(x) \in \mathcal{U}$ because since $\mathbf{U}$ is a basis of $\mathcal{U}$, $\mathbf{U}\mathbf{U}^T\mathbf{x} \in \mathcal{U}$, so $\text{Im}(\Pi_{\mathcal{U}}) \subset \mathcal{U}$.
  If $\mathbf{y} \in \mathcal{U}$ then $\Pi_{\mathcal{U}}(\mathbf{y}) = \mathbf{U}\mathbf{U}^T\mathbf{y} = \mathbf{y}$ (see property 6), so $\text{Im}(\Pi_{\mathcal{U}}) \supset \mathcal{U}$.
  So $\text{Im}(\Pi_{\mathcal{U}}) = \mathcal{U}$

- $\text{Ker}(\Pi_{\mathcal{U}}) = \mathcal{U}_\perp$:
  Let $\mathbf{x} \in \text{Ker}(\Pi_{\mathcal{U}})$. Then $\Pi_{\mathcal{U}}(\mathbf{x}) = 0$, so $\mathbf{U}\mathbf{U}^T\mathbf{x} = 0$.
  We know from property 9 that $\mathbf{x}$ can be uniquely decomposed into $\mathbf{x} = \mathbf{x}' + \mathbf{x}'_\perp$ where $\mathbf{x}' \in \mathcal{U}$ and $\mathbf{x}'_\perp \in \mathcal{U}_\perp$.
  So $0 = \mathbf{U}\mathbf{U}^T\mathbf{x}' + \mathbf{U}\mathbf{U}^T\mathbf{x}'_\perp = \mathbf{x}' + \mathbf{U}\mathbf{U}^T\mathbf{x}'_\perp$.
  Since $\mathbf{x}'_\perp \in \mathcal{U}_\perp$, $\mathbf{U}\mathbf{U}^T\mathbf{x}'_\perp = 0$, so $\mathbf{x}' = 0$.

Therefore $\mathbf{x} = \mathbf{x'}_{\perp} \in \mathcal{U}_{\perp}$, so $\text{Ker}(\Pi_{\mathcal{U}}) \subset \mathcal{U}_{\perp}$.
Let $\mathbf{x} \in \mathcal{U}_{\perp}$, then $\forall \mathbf{u} \in \mathcal{U}, \langle \mathbf{u}, \mathbf{x} \rangle = 0$, so $\mathbf{U}\mathbf{U}^T\mathbf{x} = 0$. So $\text{Ker}(\Pi_{\mathcal{U}}) \supset \mathcal{U}_{\perp}$.
So $\text{Ker}(\Pi_{\mathcal{U}}) = \mathcal{U}_{\perp}$.

- $\forall \mathbf{x}: \quad \|\Pi_{\mathcal{U}}(\mathbf{x})\| \leq \|\mathbf{x}\|$:
We know that any vector $\mathbf{x}$ can be uniquely decomposed into two components: one is $\Pi_{\mathcal{U}}(\mathbf{x}) \in \mathcal{U}$, and the other is in $\mathcal{U}_{\perp}$. For simplicity we call the other component $\mathbf{v}$.
$\mathbf{x} = \Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v}$
$\|\mathbf{x}\|^2 = \|\Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v}\|^2 = \langle \Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v}, \Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v} \rangle = \langle \Pi_{\mathcal{U}}(\mathbf{x}), \Pi_{\mathcal{U}}(\mathbf{x}) \rangle + \langle \Pi_{\mathcal{U}}(\mathbf{x}), \mathbf{v} \rangle + \langle \mathbf{v}, \Pi_{\mathcal{U}}(\mathbf{x}) \rangle + \langle \mathbf{v}, \mathbf{v} \rangle$
Since the vector $\mathbf{v}$ is in the orthogonal complement of $\mathcal{U}$, it must be orthogonal to all the vectors on $\mathcal{U}$.
$\|\mathbf{x}\|^2 = \|\Pi_{\mathcal{U}}(\mathbf{x})\|^2 + \|\mathbf{v}\|^2 \Rightarrow \|\Pi_{\mathcal{U}}(\mathbf{x})\|^2 \leq \|\mathbf{x}\|^2 \Rightarrow \|\Pi_{\mathcal{U}}(\mathbf{x})\| \leq \|\mathbf{x}\|$

□

Property 11. *Let $\mathcal{U}$ be a subspace of $\mathbb{R}^n$. Then, for any $\mathbf{x} \in \mathbb{R}^n$ we have $\arg\min_{\mathbf{u} \in \mathcal{U}} \|\mathbf{u} - \mathbf{x}\| = \Pi_{\mathcal{U}}(\mathbf{x})$. I.e. the closest point to $\mathbf{x}$ which belongs to $\mathcal{U}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathcal{U}$.*