IFT 6760A - Lecture 2
SVD and Orthogonality

Scribe(s): Genevieve Chafouleas, David Ferland
(adapted from Yann BOUTEILLER, Parviz HAGGI last year's notes) **Instructor:** Guillaume Rabusseau

1 Summary

In the previous lecture we reviewed the basics of linear algebra and introduced briefly the concept of Singular Value Decomposition of a matrix, often called SVD. In this lecture we will first look into the concepts of orthogonality and orthogonal projections, and then we will expand on the concept of SVD and introduce QR decomposition.

2 Orthogonality and Projections

Definition 1 (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 $\mathbb{R}^n$*
- *The inner product induces a norm on $\mathbb{R}^n$: $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$*

Definition 2 (Orthogonal vectors)**.** *We say that $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n are orthogonal iff \langle \mathbf{a}, \mathbf{b} \rangle = 0$.*

Definition 3 (Orthogonal basis)**.** *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*

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \begin{cases} 1 & \forall \ i = j \\ 0 & \forall \ i \neq j \end{cases} \quad \text{; for all } \ i, j \in \{1, 2, .., k\}$$

This means the vectors $\mathbf{u}_1, .., \mathbf{u}_k$ are all of norm 1 and pairwise orthogonal.

Definition 4 (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 $\mathbf{U}^T \mathbf{U} = \mathbf{I}$.*

Remark. *$\mathbf{U}\mathbf{U}^T$ 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 $\mathbf{U}\mathbf{U}^T = \mathbf{I}$*

Definition 5 (Orthogonal complement)**.** *If $\mathcal{U}$ is a subspace of $\mathbb{R}^n$, the* orthogonal complement *of $\mathcal{U}$ is defined as*

$$\mathcal{U}^\perp = \{\mathbf{v} \in \mathbb{R}^n \mid \langle \mathbf{u}, \mathbf{v} \rangle = 0 \quad \forall \mathbf{u} \in \mathcal{U}\}$$

$\mathcal{U}^\perp$ is the set of vectors that are orthogonal to every vector in $\mathcal{U}$. A graphical illustration of the orthogonal complement can be found in figure 1.

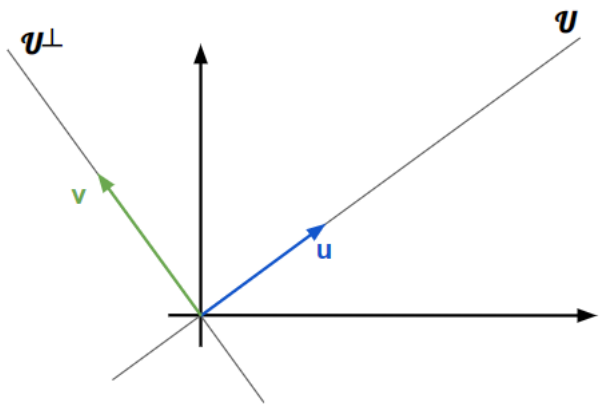


Figure 1: Illustration of the orthogonal complement of U

Property 6. *$\mathcal{U}^\perp$ is a sub-space of $\mathbb{R}^n$.*

Proof. $\mathcal{U}^\perp$ will be a subspace of $\mathbb{R}^n$ if and only if $\mathcal{U}^\perp$ is closed under addition and scalar multiplication, and that it contains the $\mathbf{0}$ vector. Let $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{U}^\perp$, $\mathbf{u} \in \mathcal{U}$, and c be a constant, then

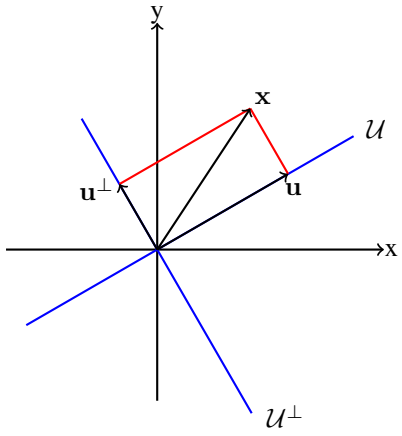
1. $(\mathbf{v}_1 + \mathbf{v}_2) \cdot \mathbf{u} = \mathbf{v}_1 \cdot \mathbf{u} + \mathbf{v}_2 \cdot \mathbf{u} = 0$, hence $\mathbf{v}_1 + \mathbf{v}_2 \in \mathcal{U}^\perp$.
2. $(c\mathbf{v}_1) \cdot \mathbf{u} = c(\mathbf{v}_1 \cdot \mathbf{u}) = c(0) = 0$, hence $c\mathbf{v}_1 \in \mathcal{U}^\perp$.
3. $\mathbf{0} \cdot \mathbf{u} = \mathbf{0}$, hence $\mathbf{0} \in \mathcal{U}^\perp$.

□

Property 7. *$\mathbb{R}^n = \mathcal{U} \oplus \mathcal{U}^\perp$, 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 $\mathcal{U}^\perp$.*

Proof. We know that both $\mathcal{U}$ and $\mathcal{U}^\perp$ are sub-spaces of $\mathbb{R}^n$, and that zero is an element of every sub-space. It is easily seen that 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 $\mathcal{U}^\perp$, we can prove that the inner product of $\mathbf{w}$ with itself must be 0 i.e. $\langle \mathbf{w}, \mathbf{w} \rangle = 0$ which is contradictory to the assumption that $\mathbf{w} \neq 0$).
Any 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 11 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 so it must belong to the orthogonal complement of $\mathcal{U}$.
We conclude that $\mathbb{R}^n = \mathcal{U} \oplus \mathcal{U}^\perp$, 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 $\mathcal{U}^\perp$. □

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



Property 8. *Let $\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_k \in \mathbb{R}^n$ be an orthonormal basis of a sub-space $\mathcal{U}$, and let*

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

Then, for all $\mathbf{x} \in \mathcal{U}$,

$$\mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{x}$$

and 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 $\mathbf{U}^T\mathbf{x}$ have the same norm (even though they are not of the same dimension):

$$\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = (\mathbf{U}\mathbf{U}^T\mathbf{x})^T\mathbf{x} = \mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{x} = (\mathbf{U}^T\mathbf{x})^T(\mathbf{U}^T\mathbf{x}) = \|\mathbf{U}^T\mathbf{x}\|^2$$

so,

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

□

Definition 9 (Linear map)**.** *A function $f : V \to W$ is said to be a linear map if for any two vectors $\mathbf{u}, \mathbf{v} \in V$ and any scalar $c \in \mathbb{F}$ the following two conditions are satisfied:*

- *Additivity: $f(\mathbf{u} + \mathbf{v}) = f(\mathbf{u}) + f(\mathbf{v})$*
- *homogeneity (of degree 1): $f(c\mathbf{u}) = cf(\mathbf{u})$*

Given a linear map f : $\mathbb{R}^n \to \mathbb{R}^n$, the kernel of f and the image of f are defiend by:

- Ker(f): $\{\mathbf{x} \in \mathbb{R}^n |\ f(\mathbf{x}) = \mathbf{0}\}$
- Im(f): $\{f(\mathbf{x}) |\ \mathbf{x} \in \mathbb{R}^n\}$

If $f : \mathbf{x} \mapsto \mathbf{M}\mathbf{x}$, then $\mathcal{R}(\mathbf{M}) = Im(f)$ and $\mathcal{N}(\mathbf{M}) = ker(f)$

Definition 10 (Orthogonal projection)**.** *Let $\mathbf{u}_1, \cdots \mathbf{u}_k \in \mathbb{R}^n$ an orthonormal basis of a subspace $\mathcal{U}$ and $\mathbf{U} \in \mathbb{R}^{n \times k}$ such that*

$$\mathbf{U} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k \\ | & & | \end{bmatrix}$$

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

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

A graphical illustration of the orthogonal projection can be found in figure 3.

Proof. We need to show that Π_u does not depend on the particular choice of an orthonormal basis of U, that is, we need to show that Π_u is well-defined, i.e, if $\mathbf{V} \in \mathbb{R}^{n\times k}$ is orthogonal and $\mathcal{R}(\mathbf{V}) = \mathcal{R}(\mathbf{U})$ then $\mathbf{U}\mathbf{U}^T = \mathbf{V}\mathbf{V}^T$. Let $\mathbf{v}_1, \cdots \mathbf{v}_k$ be another orthogonal basis of $\mathcal{U}$ and $\mathbf{V} \in \mathbb{R}^{n\times k}$ such that

$$\mathbf{V} = \begin{bmatrix} | & & | \\ \mathbf{v}_1 & \cdots & \mathbf{v}_k \\ | & & | \end{bmatrix}$$

Since $\mathcal{R}(\mathbf{V}) = \mathcal{R}(\mathbf{U})$ then there exists $\mathbf{P} \in \mathbb{R}^{k\times k}$ such that $\mathbf{V} = \mathbf{U}\mathbf{P}$.
Since:

- $\mathbf{P}$ is square
- $\mathbf{I} = \mathbf{V}^T\mathbf{V} = \mathbf{P}^T \underbrace{\mathbf{U}^T\mathbf{U}}_{\mathbf{I}} \mathbf{P} = \mathbf{P}^T\mathbf{P}$

we have $\mathbf{P}\mathbf{P}^T = \mathbf{I}$. Thus, $\mathbf{V}\mathbf{V}^T = \mathbf{U}\mathbf{P}\mathbf{P}^T\mathbf{U}^T = \mathbf{U}\mathbf{U}^T$ □

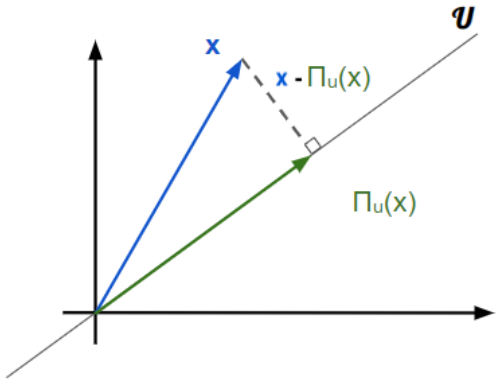


Figure 3: Illustration of the orthogonal projection of $\mathbf{x}$ onto $\mathbf{U}$

Property 11. *Based on the above definitions of orthogonality and orthogonal complement, we can summarize the following properties of orthogonal projection:*

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

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 and the fact that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ when $\mathbf{U}$ is an orthogonal matrix. Knowing that $\Pi_{\mathcal{U}}(\mathbf{x}) = \mathbf{U}\mathbf{U}^T\mathbf{x}$, we evaluate the expression $\Pi_{\mathcal{U}}^2(\mathbf{x})$:
  $\Pi_{\mathcal{U}}^2(\mathbf{x}) = \Pi_{\mathcal{U}}(\Pi_{\mathcal{U}}(\mathbf{x})) = \Pi_{\mathcal{U}}(\mathbf{U}\mathbf{U}^T\mathbf{x}) = \mathbf{U}\underbrace{(\mathbf{U}^T\mathbf{U})}_{\mathbf{I}}\mathbf{U}^T\mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{x} = \Pi_{\mathcal{U}}(\mathbf{x})$.

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

  To prove this we use the definition of projection in terms of matrices and the fact that $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ when $\mathbf{U}$ is an orthogonal matrix. Knowing that $\Pi_{\mathcal{U}}(\mathbf{x}) = \mathbf{U}\mathbf{U}^T\mathbf{x}$, we evaluate the expression:

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

- $\text{Im}(\Pi_{\mathcal{U}}) = \mathcal{U}$.

  We first show that $\text{Im}(\Pi_{\mathcal{U}}) \subset \mathcal{U}$. Let $\mathbf{v} \in \text{Im}(\Pi_{\mathcal{U}})$, this means that there exists $\mathbf{x}$ such that $\mathbf{v} = \Pi_{\mathcal{U}}(\mathbf{x}) = \mathbf{U}\mathbf{U}^T\mathbf{x} = \mathbf{U}(\mathbf{U}^T\mathbf{x})$, showing that $\mathbf{v}$ is a linear combination of the columns of $\mathbf{U}$, hence $\mathbf{v} \in \mathcal{U}$ since $\mathcal{R}(\mathbf{U}) = \mathcal{U}$ by definition.

  The inclusion $\text{Im}(\Pi_{\mathcal{U}}) \supset \mathcal{U}$ directly follows for Property 8: for all $\mathbf{x} \in \mathcal{U}$, we have $\mathbf{x} = \mathbf{U}\mathbf{U}^T\mathbf{x} = \Pi_{\mathcal{U}}(\mathbf{x}) \in \text{Im}(\Pi_{\mathcal{U}})$.

- $\text{Ker}(\Pi_{\mathcal{U}}) = \mathcal{U}_{\perp}$.

  Let $\mathbf{x} \in \text{Ker}(\Pi_{\mathcal{U}})$, and $\mathbf{u} \in \mathcal{U}$, we first show that their dot product is 0:

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

  This shows that $\text{Ker}(\Pi_{\mathcal{U}})$ is a subspace of the orthogonal complement of $\mathcal{U}$.

  To show that $\mathcal{U}^{\perp}$ is a subspace of $\text{Ker}(\Pi_{\mathcal{U}})$, we nedd to show that if $\langle \mathbf{x}, \mathbf{u} \rangle = 0$ for all $\mathbf{u} \in \mathcal{U}$, then $\mathbf{x} \in ker(\Pi_{\mathcal{U}})$:

$$\langle \mathbf{x}, \mathbf{u} \rangle = 0 \text{ for all } \mathbf{u} \in \mathcal{U} \Rightarrow \mathbf{U}^T\mathbf{x} = 0 \Rightarrow \mathbf{U}\mathbf{U}^T\mathbf{x} = 0 \Rightarrow \mathbf{x} \in ker(\Pi_{\mathcal{U}})$$

- $\forall \mathbf{x}: \ \|\Pi_{\mathcal{U}}(\mathbf{x})\| \leq \|\mathbf{x}\|$.

  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 $\mathbf{v}$, which gives us $\mathbf{x} = \Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v}$. Developing the expression of the norm of $\mathbf{x}$ we find:

$$\|\mathbf{x}\|^2 = \|\Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v}\|^2 = \langle \Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v}, \Pi_{\mathcal{U}}(\mathbf{x}) + \mathbf{v} \rangle = \langle \Pi_{\mathcal{U}}(\mathbf{x}), \Pi_{\mathcal{U}}(\mathbf{x}) \rangle + \langle \Pi_{\mathcal{U}}(\mathbf{x}), \mathbf{v} \rangle + \langle \mathbf{v}, \Pi_{\mathcal{U}}(\mathbf{x}) \rangle + \langle \mathbf{v}, \mathbf{v} \rangle$$

  Since the vector $\mathbf{v}$ is in the orthogonal complement of $\mathcal{U}$, it must be orthogonal to all the vectors on $\mathcal{U}$, therefore

$$\langle \mathbf{v}, \Pi_{\mathcal{U}}(\mathbf{x}) \rangle = \langle \Pi_{\mathcal{U}}(\mathbf{x}), \mathbf{v} \rangle = 0$$

  and $\|\mathbf{x}\|^2 = \|\Pi_{\mathcal{U}}(\mathbf{x})\|^2 + \underbrace{\|\mathbf{v}\|^2}_{\geq 0} \Rightarrow \|\Pi_{\mathcal{U}}(\mathbf{x})\|^2 \leq \|\mathbf{x}\|^2 \Rightarrow \|\Pi_{\mathcal{U}}(\mathbf{x})\| \leq \|\mathbf{x}\|$

□

Property 12. *Let $\mathcal{U}$ be a subspace of $\mathbb{R}^n$. Then, for any $\mathbf{x} \in \mathbb{R}^n$ we have*

$$\arg\min_{\mathbf{u}\in\mathcal{U}} \|\mathbf{u} - \mathbf{x}\| = \Pi_{\mathcal{U}}(\mathbf{x}).$$

Which means the orthogonal projection of $\mathbf{x}$ onto $\mathcal{U}$ is the closest point to $\mathbf{x}$ in $\mathcal{U}$.

Proof. We want to show that for all $\mathbf{x} \in \mathbb{R}^n$, we have $\arg\min_{v\in\mathbf{U}} \|\mathbf{x} - \mathbf{v}\| = \Pi_u(\mathbf{x})$, where $\mathbf{v} \in \mathcal{U}$.
We first show that the vectors $\mathbf{x} - \Pi_{\mathcal{U}}(\mathbf{x})$ and $\mathbf{v} - \Pi_{\mathcal{U}}(\mathbf{x})$ are orthogonal.

$$\begin{aligned}
\langle \mathbf{x} - \Pi_{\mathcal{U}}(\mathbf{x}), \mathbf{v} - \Pi_{\mathcal{U}}(\mathbf{x})\rangle &= \langle \mathbf{x} - \Pi_{\mathcal{U}}(\mathbf{x}), \mathbf{v}\rangle - \langle \mathbf{x} - \Pi_{\mathcal{U}}(\mathbf{x}), \Pi_{\mathcal{U}}(\mathbf{x})\rangle \\
&= \langle \mathbf{x} - \Pi_{\mathcal{U}}(\mathbf{x}), \mathbf{v}\rangle - 0 && \text{Using } 2^{\text{nd}} \text{ element of property 11} \\
&= \langle \mathbf{x}, \mathbf{u}\rangle - \langle \Pi_{\mathcal{U}}(\mathbf{x}), \mathbf{v}\rangle \\
&= \mathbf{x}^T\mathbf{v} - (\mathbf{U}\mathbf{U}^T\mathbf{x})^T\mathbf{v} \\
&= \mathbf{x}^T\mathbf{v} - \mathbf{x}^T\mathbf{U}\mathbf{U}^T\mathbf{v} \\
&= 0
\end{aligned}$$

Note that the last equality is due to the fact that vectors in $\mathcal{U}$ are projected onto themselves ($\mathbf{u} = \mathbf{U}\mathbf{U}^T\mathbf{u}$).
Knowing that $\mathbf{x} - \Pi_{\mathcal{U}}(\mathbf{x})$ and $\mathbf{v} - \Pi_{\mathcal{U}}(\mathbf{x})$ are orthogonal, we can then use the Pythagorean theorem to express $\|\mathbf{x} - \mathbf{v}\|^2$ as $\|\mathbf{x} - \mathbf{v}\|^2 = \|\mathbf{x} - \Pi_u(\mathbf{x})\|^2 + \|\mathbf{v} - \Pi_u(\mathbf{x})\|^2$. Therefore, the minimum distance between $\mathbf{x}$ and $\mathbf{v}$ is when $\mathbf{v}$ is the orthogonal projection of $\mathbf{x}$ onto $\mathcal{U}$ which gives us $\|\mathbf{v} - \Pi_u(\mathbf{x})\|^2 = 0$ as shown in figure 4. □

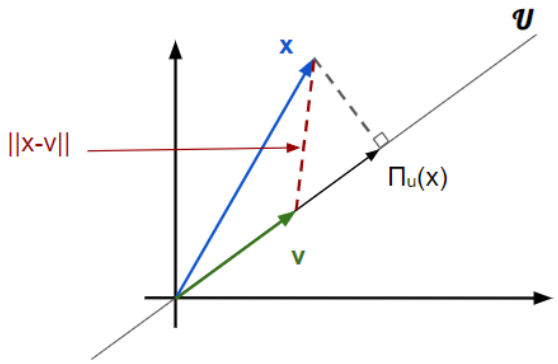


Figure 4: Illustration of the distance between $\mathbf{x}$ and $\mathbf{v}$

3 Singular Value Decomposition

Theorem 13 (SVD)**.** *Any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with rank$(\mathbf{A}) = r$ can be decomposed 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 with positive entries 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 full Singular-Value Decomposition (SVD) of $\mathbf{A}$.

Property 14. *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 $\mathbf{u}_1, \ldots, \mathbf{u}_m \in \mathbb{R}^m$ and $\mathbf{v}_1, \ldots, \mathbf{v}_m \in \mathbb{R}^m$ respectively. Then the following hold:*

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

In order to prove the above properties we can rewrite the SVD as follows:

$$\underset{(m\times n)}{\mathbf{A}} = \underset{(m\times m)}{\mathbf{U}}\underset{(m\times n)}{\mathbf{\Sigma}}\underset{(n\times n)}{\mathbf{V}^T} = \begin{bmatrix}\underset{(m\times r)}{\mathbf{U_1}} & \underset{(m\times m-r)}{\mathbf{U_2}}\end{bmatrix}\begin{bmatrix}\underset{(r\times r)}{\mathbf{\Lambda}} & \underset{(r\times n-r)}{\mathbf{0}} \\ \underset{(m-r\times r)}{\mathbf{0}} & \underset{(m-r\times n-r)}{\mathbf{0}}\end{bmatrix}\begin{bmatrix}\underset{(r\times n)}{\mathbf{V_1}^T} \\ \underset{(n-r\times n)}{\mathbf{V_2}^T}\end{bmatrix} \tag{1}$$

We can solve for Σ of the SVD as follows:

$$\mathbf{\Sigma} = \begin{bmatrix}\underset{(r\times r)}{\mathbf{\Lambda}} & \underset{(r\times n-r)}{\mathbf{0}} \\ \underset{(m-r\times r)}{\mathbf{0}} & \underset{(m-r\times n-r)}{\mathbf{0}}\end{bmatrix} = \begin{bmatrix}\underset{(r\times m)}{\mathbf{U_1}^T} \\ \underset{(m-r\times m)}{\mathbf{U_2}^T}\end{bmatrix}\underset{(m\times n)}{\mathbf{A}}\begin{bmatrix}\underset{(n\times r)}{\mathbf{V_1}} & \underset{(n\times n-r)}{\mathbf{V_2}}\end{bmatrix} = \begin{bmatrix}\mathbf{U_1}^T\mathbf{A}\mathbf{V_1} & \mathbf{U_1}^T\mathbf{A}\mathbf{V_2} \\ \mathbf{U_2}^T\mathbf{A}\mathbf{V_1} & \mathbf{U_2}^T\mathbf{A}\mathbf{V_2}\end{bmatrix} \tag{2}$$

Here are the following proofs for the properties of the SVD[1]

Proof.

(i) $\mathbf{u}_1, \ldots, \mathbf{u}_r$ forms a basis a basis of $\mathcal{R}(\mathbf{A})$. We will prove $\mathcal{R}(\mathbf{A}) = \mathcal{R}(\mathbf{U_1})$ where $\mathbf{u}_1, \ldots, \mathbf{u}_r$ are the columns of $\mathbf{U_1}$. We have $\mathbf{A} = \mathbf{U_1}\mathbf{\Lambda}\mathbf{V_1}^T$. Since $\mathbf{\Lambda}$ is invertible and $\mathbf{V_1}^T\mathbf{V_1} = \mathbf{I}_r$, we have $\mathbf{U_1} = \mathbf{A}\mathbf{V_1}\mathbf{\Lambda}^{-1}$. Suppose $\mathbf{v} \in \mathcal{R}(\mathbf{U_1})$. Then, there exists $\mathbf{a}$ such that $\mathbf{v} = \mathbf{U_1}\mathbf{a} = \mathbf{A}\mathbf{V_1}\mathbf{\Lambda}^{-1}\mathbf{a}$. Therefore, $\mathbf{v} \in \mathcal{R}(\mathbf{A})$. It follows that $\mathcal{R}(\mathbf{A}) \supset \mathcal{R}(\mathbf{U_1})$. Conversely, if $\mathbf{v} \in \mathcal{R}(\mathbf{A})$, there exists $\mathbf{x}$ such that $\mathbf{v} = \mathbf{A}\mathbf{x} = \mathbf{U_1}\mathbf{\Lambda}\mathbf{V_1}\mathbf{x} \in \mathcal{R}(\mathbf{U_1})$, showing that $\mathcal{R}(\mathbf{A}) \subset \mathcal{R}(\mathbf{U_1})$.

(ii) $\mathbf{u}_{r+1}, \ldots, \mathbf{u}_m$ forms a basis of $\mathcal{N}(\mathbf{A}^T)$. We need to show that the vector columns of $\mathbf{U_2}$ form a basis of $\mathcal{N}(\mathbf{A}^T)$. From equation 2, we can extract the following:
$\begin{bmatrix}\mathbf{U_2}^T\mathbf{A}\mathbf{V_1} & \mathbf{U_2}^T\mathbf{A}\mathbf{V_2}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{0}\end{bmatrix} \implies \mathbf{U_2}^T\mathbf{A}\begin{bmatrix}\mathbf{V_1}^T & \mathbf{V_2}^T\end{bmatrix} = \mathbf{0} \implies \mathbf{U_2}^T\mathbf{A}\mathbf{V} = \mathbf{0} \implies \mathbf{U_2}^T\mathbf{A} = \mathbf{0} \implies \mathbf{A}^T\mathbf{U_2} = \mathbf{0}$
This implies $\mathcal{R}(\mathbf{U_2}) \subset \mathcal{N}(\mathbf{A}^T)$.
Conversely, if $\mathbf{x} \in \mathcal{N}(\mathbf{A}^T)$ we have $\mathbf{0} = \mathbf{A}^T\mathbf{x} = \mathbf{V_1}\mathbf{\Lambda}\mathbf{U_1}^T\mathbf{x}$. Since $\mathbf{V_1}$ is orthogonal and $\mathbf{\Lambda}$ is invertible, this implies $\mathbf{U_1}^T\mathbf{x} - = \mathbf{0}$ and therefore $\mathbf{x} \in \mathcal{R}(\mathbf{U_1})^\top = \mathcal{R}(\mathbf{U_2})$. Hence $\mathcal{N}(\mathbf{A}^T) \subset \mathcal{R}(\mathbf{U_2})$.

The proofs for the null space and row space of $\mathbf{A}$ are similar. □

Definition 15 (Thin SVD)**.** *If $m \geq n$, Theorem 13 can be rewritten as:*

$$\underset{(m\times n)}{\mathbf{A}} = \underset{(m\times m)}{\mathbf{U}}\underset{(m\times n)}{\mathbf{\Sigma}}\underset{(n\times n)}{\mathbf{V}^T} = \underset{(m\times m)}{\mathbf{U}}\begin{bmatrix}\underset{(n\times n)}{\mathbf{\Sigma}} \\ \mathbf{0}\end{bmatrix}\underset{(n\times n)}{\mathbf{V}^T} = \underset{(m\times n)}{\mathbf{U_n}}\underset{(n\times n)}{\mathbf{\Sigma}}\underset{(n\times n)}{\mathbf{V}^T}$$

where:

- $\mathbf{\Lambda} \in \mathbb{R}^{r\times r}$ *is diagonal with* $\mathbf{\Lambda}_{i,i} \neq 0$
- $\mathbf{U_n}$ *only keep the first n columns of* $\mathbf{U}$
- $\mathbf{U}^T\mathbf{U} = \mathbf{V}^T\mathbf{V} = \mathbf{I_r}$ *(*$\mathbf{U}$ *and* $\mathbf{V}$ *are both orthogonal matrices)*

The form $\mathbf{A} = \mathbf{U_n}\mathbf{\Sigma}\mathbf{V}^T$ is called the thin Singular-Value Decomposition *(thin SVD) of $\mathbf{A}$. If $m \leq n$ then rows of $\mathbf{V}^T$ or ommitted instead of columns of $\mathbf{U}$.*

[1] https://jekyll.math.byuh.edu/courses/m343/handouts/svdbasis.pdf

Definition 16 (Compact SVD)**.** *Given that the $\mathbf{A} = r$, Theorem 13 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* $\mathbf{\Lambda}_{i,i} \neq 0$, $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r > 0$
- $\widetilde{\mathbf{U}}^T\widetilde{\mathbf{U}} = \widetilde{\mathbf{V}}^T\widetilde{\mathbf{V}} = \mathbf{I_r}$ *($\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ are both orthogonal matrices)*

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

4 The QR decomposition

In this section, we briefly present the QR decomposition which can be used to solve the linear least squares problem for example.

Theorem 17. *Any matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ can be written as $\mathbf{A} = \mathbf{QR}$ where $\mathbf{Q} \in \mathbb{R}^{m\times m}$ is orthogonal and $\mathbf{R} \in \mathbb{R}^{m\times n}$ is upper triangular. This decomposition of $\mathbf{A}$ is called the QR decomposition.*
If $m > n$ then the reduced (thin) QR decomposition *of $\mathbf{A}$ is defined as:*

$$\mathbf{A} = \begin{bmatrix}\mathbf{Q}_1 & \mathbf{Q}_2\end{bmatrix}\begin{bmatrix}\mathbf{R}_1 \\ \\ 0\end{bmatrix} = \mathbf{Q}_1\mathbf{R}_1$$

where $\mathbf{Q}_1 \in \mathbb{R}^{m\times n}$ is orthogonal and $\mathbf{R}_1 \in \mathbb{R}^{n\times n}$ is upper triangular.

Remark 18. *If $\mathbf{U} \in \mathbb{R}^{n\times k}$ and $\operatorname{rank}(\mathbf{U}) = k$ (i.e full rank) then its thin QR decomposition $\mathbf{U} = \mathbf{QR}$ is such that:*

- $\mathcal{R}(\mathbf{Q}) = \mathcal{R}(\mathbf{U})$
- $\mathbf{R}$ *is invertible*

where $\mathbf{Q} \in \mathbb{R}^{n\times k}$ and $\mathbf{R} \in \mathbb{R}^{k\times k}$

If $\mathbf{u}_1, \cdots \mathbf{u}_k \in \mathbb{R}^n$ is a basis of $\mathcal{U}$, which is **not necessarily orthonormal**, and $\mathbf{U} \in \mathbb{R}^{n\times k}$ such that

$$\mathbf{U} = \begin{bmatrix} | & & | \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k \\ | & & | \end{bmatrix}$$

then we have the following property: $\Pi_u(\mathbf{x}) = \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{x}$

Proof. In order to show the previous property, let us consider the thin QR decomposition of $\mathbf{U}$, i.e, $\mathbf{U} = \mathbf{QR}$ where $\mathbf{Q} \in \mathbb{R}^{n\times k}$ is orthogonal and $\mathbf{R} \in \mathbb{R}^{k\times k}$ is upper triangular and invertible. The invertability follows from the previous remark since $\mathbf{U}$ is full rank. We have

$$\begin{aligned}\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{x} &= \mathbf{Q}\underbrace{\mathbf{R}(\mathbf{R}^T\mathbf{Q}^T\mathbf{Q}\mathbf{R})^{-1}\mathbf{R}^T}_{\mathbf{I}}\mathbf{Q}^T\mathbf{x} \\ &= \mathbf{Q}\mathbf{Q}^T\mathbf{x} \\ &= \Pi_u(\mathbf{x})\end{aligned} \tag{3}$$

□