Linear Algebra

Vectors, Matrices and SVD

Arun K. Tangirala

Department of Chemical Engineering, IIT Madras

Opening remarks

Elementary math introduces us to scalars, which are essentially useful for handling one-dimensional variables or data. Three aspects of analysis are usually familiar:

1. **Elementary operations:** Addition, multiplication, etc.
2. **Comparisons and strengths:** Equality, magnitude, etc.
3. **Functions of scalars:** Squares, cubic roots, etc.

Examples: mass, temperature, height, cost, salary, population

Opening remarks

Elementary math introduces us to scalars, which are essentially useful for handling one-dimensional variables or data. Three aspects of analysis are usually familiar:

1. **Elementary operations:** Addition, multiplication, etc.
2. **Comparisons and strengths:** Equality, magnitude, etc.
3. **Functions of scalars:** Squares, cubic roots, etc.

Examples: mass, temperature, height, cost, salary, population

However, scalars are highly insufficient for descriptions of several processes and many common mathematical tasks in data science, geometry, vision, etc.

Vectors

Definition 1: Vector is a quantity (variable) that has a direction (angle) and magnitude

Definition 2: Vector is an array or a list of scalars (usually numbers)

Definition 3: Vector is a point in n-dimensional space

Vectors

Definition 1: Vector is a quantity (variable) that has a direction (angle) and magnitude

Definition 2: Vector is an array or a list of scalars (usually numbers)

Definition 3: Vector is a point in n-dimensional space

Examples

1. Wind velocity - horizontal or vertical component
2. Displacement - has a magnitude and an angle
3. Height and weight of an individual in a single list
4. Attributes of a data set (max. value, no. of non-zero elements, etc.) in a single list
5. ...

Representations of Vectors

Vectors are usually written as an (row or column) array or as a point in the Cartesian space:

Examples: $\mathbf{x} = \begin{bmatrix} 4.5 \\ -2.9 \end{bmatrix}$, $\mathbf{z} = \begin{bmatrix} 32 & 14.5 & -23.7 \end{bmatrix}$, $\mathbf{y} = (4.5, -2.9)$ $\underline{v} = \begin{bmatrix} 5 \\ 62 \end{bmatrix}$: $\alpha \underline{v}$

Scalar

Vector space

A space comprising vectors (defined over a field $\mathcal{F}$) with the two operations of vector addition and scalar (in the field $\mathcal{F}$) multiplication, together with the eight axioms. is called a vector space $\mathcal{V}$.

Note: The phrase "vector space" can also comprise functions or any other elements that satisfy the definition

$\underline{x} + \underline{y} = \underline{y} + \underline{x}$ $\underline{x} + \underline{0} = \underline{x}$

1. The axioms include associativity and commutativity of addition, existence of additive identity and inverse elements, compatibility of scalar multiplication, existence of multiplicative identity and distributivity of multiplication w.r.t. addition.

Operations

$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ (column vector)

1. **Addition:** Vectors (of the same dimension) are added element-wise: $\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}$

2. **Scalar multiplication:** Multiplication of $\mathbf{x}$ with $\alpha \in \mathbb{R}$ is denoted by $\alpha\mathbf{x} = \begin{bmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_n \end{bmatrix}$

3. **Transpose:** A column (or a row) vector is transposed to a row (or a column) vector by interchanging the row and column dimensions. For instance, $\mathbf{x}^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$

Vector Norm

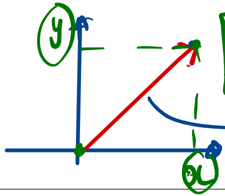
Norms measure the property or size of a vector (e.g., length, distance, sparsity). They are highly useful in data science - e.g., "fitting" models, quantifying errors.

Definition

The p-norm (or the l_p-norm) of a vector $\mathbf{x} \in \mathbb{R}^N$ is defined as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$$

$$||\mathbf{x}||_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p} \tag{1}$$



$\begin{bmatrix} x \\ y \end{bmatrix}$ $\quad p \geq 1$

2-norm (length of vector)

Vector Norm

Norms measure the property or size of a vector (e.g., length, distance, sparsity). They are highly useful in data science - e.g., "fitting" models, quantifying errors.

Definition

The p-norm (or the l_p-norm) of a vector $\mathbf{x} \in \mathbb{R}^N$ is defined as

$p \geq 1$

$p = 2$

$$||\mathbf{x}||_p = \left(\sum_{i=1}^{N} |x_i|^p \right)^{1/p} \tag{1}$$

2-norm

$p = 0.5$

- Norms satisfy (i) $||\mathbf{x}|| = 0 \iff \mathbf{x} = 0$, (ii) homogeneity, $||\alpha \mathbf{x}|| = |\alpha| ||\mathbf{x}||$ and (iii) triangular inequality, $||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$.
- In fact, for any $0 \leq p < 1$, the function in (1) is not a proper norm (why?).
- All vectors that satisfy $||\mathbf{x}||_p = 1$ constitute what is known as a **unit ball.**

2-norm, 1-norm and ∞-norm

2-norm

The 2-norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$||\mathbf{x}||_2 = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \tag{2}$$

and measures the vector length / distance of the point in $\mathbb{R}^n$ from the origin.

1-norm

The 1-norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

p

$$||\mathbf{x}||_1 = |x_1| + |x_2| + \cdots + |x_n| \tag{3}$$

and measures the distance traversed from the origin as travelled by a taxi.

∞-norm

The ∞-norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is defined as

$p \to \infty$

$$||\mathbf{x}||_\infty = \max\{|x_1|, |x_2|, \cdots, |x_n|\} \tag{4}$$

and measures the peak value of element in the vector

$$\underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{100} \end{bmatrix} \quad ; \quad \underline{\hat{y}} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_{100} \end{bmatrix}$$

Observations

Predictions

$$\underline{e} = \underline{y} - \underline{\hat{y}} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_{100} \end{bmatrix}$$

$\underline{e}$ to be "small"

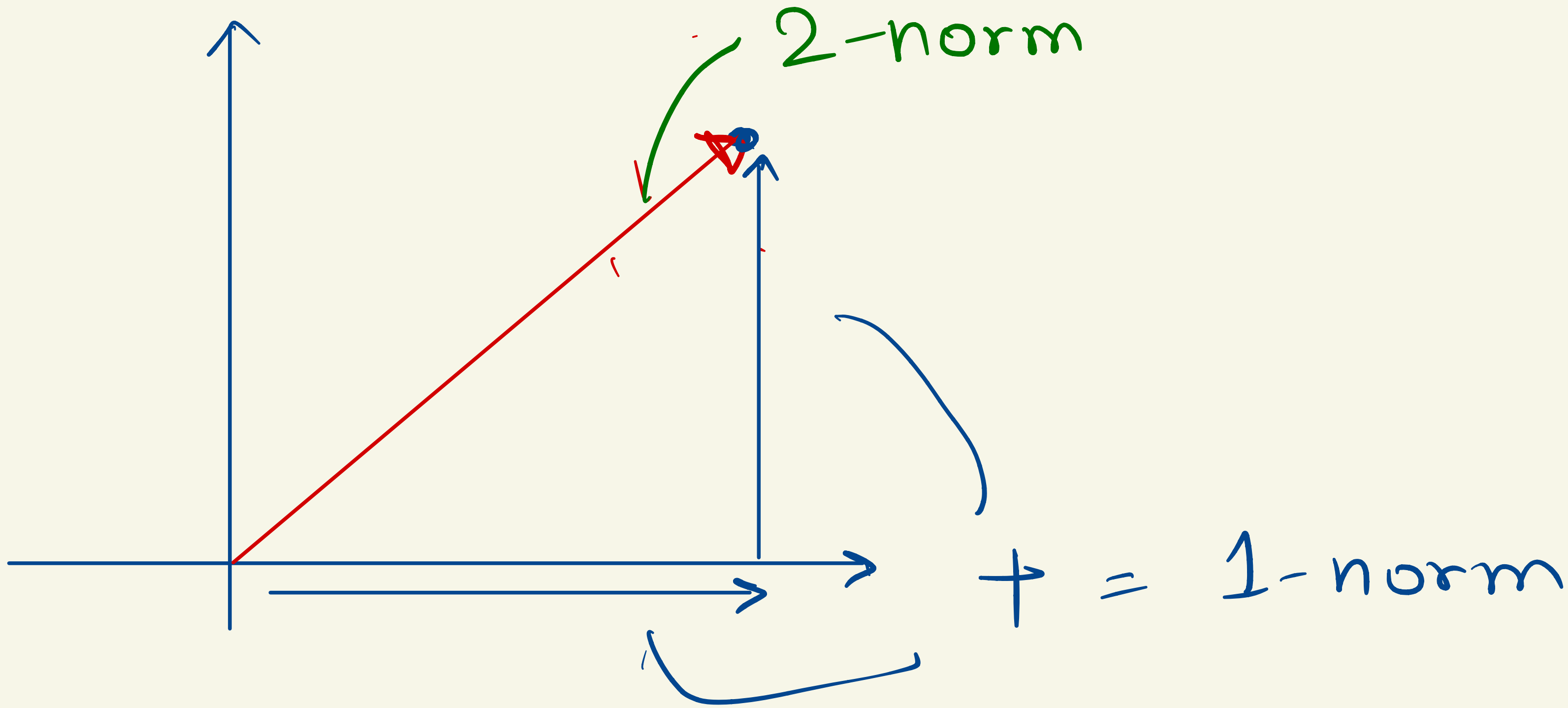
Squared Euclidean distance

① 2-norm of $\underline{e}$ be low $\quad \|\underline{e}\|_2^2 = e_1^2 + e_2^2 + \dots + e_{100}^2$

Least squares ← minimizes

Taxi-driver's distance $\quad \|\underline{e}\|_1 = |e_1| + |e_2| + \dots + |e_{100}|$

Least absolute distance



2-norm
+ = 1-norm

Zero "norm"

$\hat{y}[k] = \sum_{i=1}^{100} \theta_i x_i[k]$ | $\min \sum_{k=1}^{1000} e^2[k]$

$|x_i|^1, \; |x_i|^{0.1}, \; |x_i|^{0.001}, \; |x_i|^{0.00001}$ $+\|\theta\|_0$

Definition

The function $||\mathbf{x}||_0$, referred to as l_0-norm of $\mathbf{x}$, is defined as

limit

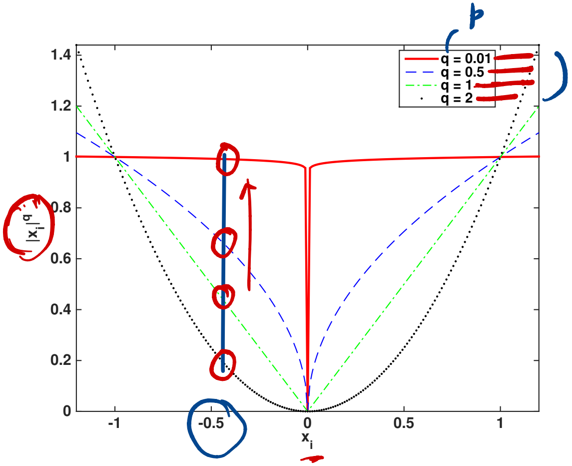
$$||\mathbf{x}||_0 = \boxed{\lim_{p\to 0} ||\mathbf{x}||_p^p} \tag{5}$$

$$= \lim_{p\to 0} \sum_{i=1}^{N} |x_i|^p = \boxed{\sum_{i=1}^{N} \lim_{p\to 0} |x_i|^p} \tag{6}$$

For every x_i, when $p \to 0$, $|x_i|^p \to \mathcal{I}(x_i)$, **the indicator function.**

0-norm: No. of non-zero elements

Zero "norm"



Plot of $|x_i|^q$ vs. x_i for different q.

Dot product

The "dot product" (or the scalar product) of two **real-valued** $N \times 1$ vectors $\mathbf{v}$ and $\mathbf{w}$ is defined as

$$\mathbf{v}.\mathbf{w} = \sum_{n=1}^{N} v_i w_i = \mathbf{w}^T\mathbf{v} = \mathbf{v}^T\mathbf{w} \tag{7}$$

Dot product

The "dot product" (or the scalar product) of two **real-valued** $N \times 1$ vectors $\mathbf{v}$ and $\mathbf{w}$ is defined as

$$\mathbf{v}.\mathbf{w} = \sum_{n=1}^{N} v_i w_i = \mathbf{w}^T\mathbf{v} = \mathbf{v}^T\mathbf{w} \tag{7}$$

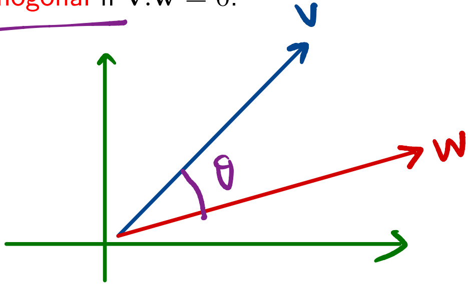
- The dot product is a special case of inner product.
- The length of a vector, denoted by $||.||$, is the square root of $\mathbf{v}.\mathbf{v}$: $||\mathbf{v}|| = \sqrt{\mathbf{v}.\mathbf{v}}$
- Dot product is also the squared 2-norm of $\mathbf{v}$, denoted by $||\mathbf{v}||_2^2$ (or sometimes simply $||\mathbf{v}||^2$).

MATLAB: `dot`

- **Geometrical relation:** If $\mathbf{v}$ and $\mathbf{w}$ are at an angle θ w.r.t. each other, then

$$\cos\theta = \frac{\mathbf{v}.\mathbf{w}}{||\mathbf{v}||_2||\mathbf{w}||_2} \tag{8}$$

Thus, two vectors are orthogonal if $\mathbf{v}.\mathbf{w} = 0$.



- **Geometrical relation:** If $\mathbf{v}$ and $\mathbf{w}$ are at an angle θ w.r.t. each other, then

$$\cos\theta = \frac{\mathbf{v}.\mathbf{w}}{||\mathbf{v}||||\mathbf{w}||} \tag{8}$$

  Thus, two vectors are orthogonal if $\mathbf{v}.\mathbf{w} = 0$.

- **Schwartz inequality:** $|\mathbf{v}.\mathbf{w}| \leq ||\mathbf{v}||||\mathbf{w}||$ (also satisfied by inner products)

Dot product contd.

- **Geometrical relation:** If $\mathbf{v}$ and $\mathbf{w}$ are at an angle θ w.r.t. each other, then

$$\cos\theta = \frac{\mathbf{v}.\mathbf{w}}{||\mathbf{v}||||\mathbf{w}||} \tag{8}$$

Thus, two vectors are orthogonal if $\mathbf{v}.\mathbf{w} = 0$.

- **Schwartz inequality:** $|\mathbf{v}.\mathbf{w}| \leq ||\mathbf{v}||||\mathbf{w}||$ (also satisfied by inner products)
- **For complex-valued vectors**, the order of operation matters, i.e. $\mathbf{v}.\mathbf{w} = \mathbf{v}^H\mathbf{w} \neq \mathbf{w}.\mathbf{v}$, where H indicates complex conjugate transpose. Thus,

$$\cos\theta = \frac{\text{Re}(\mathbf{v}.\mathbf{w})}{||\mathbf{v}||||\mathbf{w}||} \tag{9}$$

Inner product

The inner product between two vectors, denoted by, $\langle \mathbf{v}, \mathbf{w} \rangle$ is a product that satisfies:

1. **Conjugate symmetry:** $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle^\star$ (* denotes complex conjugate).
2. **Positive-definiteness:** $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and $\langle \mathbf{v}, \mathbf{v} \rangle = 0 \implies \mathbf{v} = 0$
3. **Linearity in the first argument:** $\langle \alpha\mathbf{v}, \mathbf{w} \rangle = \alpha\langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle \mathbf{v} + \mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{z}, \mathbf{w} \rangle$

Inner product

The inner product between two vectors, denoted by, $\langle \mathbf{v}, \mathbf{w} \rangle$ is a product that satisfies:

1. **Conjugate symmetry:** $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle^\star$ (* denotes complex conjugate).
2. **Positive-definiteness:** $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$, and $\langle \mathbf{v}, \mathbf{v} \rangle = 0 \implies \mathbf{v} = 0$
3. **Linearity in the first argument:** $\langle \alpha\mathbf{v}, \mathbf{w} \rangle = \alpha\langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle \mathbf{v} + \mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{z}, \mathbf{w} \rangle$

- It is easy to verify that the dot product satisfies all the above properties
- Example of an inner product in the $\mathbb{R}^2$ space that is NOT a dot product:

$$\underline{\langle \mathbf{v}, \mathbf{w} \rangle = v_1 w_1 - v_1 w_2 - v_2 w_1 + 4 v_2 w_2}. \leftarrow IP$$

$$v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}; \quad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$v \cdot w = v_1 w_1 + v_2 w_2 \leftarrow DP$$

The inner (dot) product between two **functions** $f(t)$ and $g(t)$ in the Hilbert space is given by

$$< f, g >= \int_{-\infty}^{\infty} f(t)g^{\star}(t)\, dt \tag{10}$$

Note: We shall use the terms "dot product" and "inner product" interchangeably, unless specified otherwise.

Linear independence

$y = \alpha_1 x_1 + \alpha_2 x_2$

Does there exist a relation bet. y, x_1 & x_2?

$y - \alpha_1 x_1 - \alpha_2 x_2 = 0$

OR $c_1 y + c_2 x_1 + c_3 x_2 = 0$

Linear Independence

The set of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$ is said to be **linearly independent** if and only if their linear combination yields zero under trivial conditions, i.e.,

$$\sum_{n=1}^{p} c_i \mathbf{v}_i = 0 \iff c_i = 0, \; \forall i = 1, \cdots, p \tag{11}$$

- Linear dependence implies one or more of the vectors can be expressed as a linear combination of the others.

$\boxed{c_1 v_1 + c_2 v_2 + c_3 v_3 = 0}$

$\boxed{v_3 = \alpha_1 v_1 + \alpha_2 v_2}$

$$\underline{v_1} = \begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix}; \qquad \underline{v_2} = \begin{bmatrix} v_{21} \\ v_{22} \end{bmatrix};$$

Suppose $\underline{v_3} = 2\underline{v_1} + \underline{v_2}$

Then $\{\underline{v_1}, \underline{v_2}, \underline{v_3}\}$ are <u>not</u> linearly independent

$$\underline{v_2} = \underline{v_3} - 2\underline{v_1}$$

$$\underline{v_1} = \frac{1}{2}v_3 - \frac{1}{2}v_2$$

Span and Basis

The span of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$ is the linear combination of those vectors. Any vector $\mathbf{w}$ in the spanned space W can be expressed as

$$\mathbf{w} = \sum_{i=1}^{p} \alpha_i \mathbf{v}_i \tag{12}$$

V_1 : Daily Salary

V_2 : No. of working hours

V_3 : Sales of (product)

V_4 : Profit (daily)

$$\left\{ \underline{V_1}, \underline{V_2}, \underline{V_3}, \underline{V_4} \right\}$$

365×1

1. Are these vectors linearly related?
2. How many of these can be treated as "dependent"?

Span and Basis

The span of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_p$ is the linear combination of those vectors. Any vector $\mathbf{w}$ in the spanned space W can be expressed as

$$\mathbf{w} = \sum_{i=1}^{p} \alpha_i \mathbf{v}_i \tag{12}$$

Basis

A set of vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p$ is said to be the basis for a vector space W if

1. The vectors are linearly independent (not too many vectors)
2. They span the space W (not too few vectors)

The number of vectors in the basis set for a vector space W is said to be its **dimension**.

① $\underline{v}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$; $v_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

i. Are they linearly indep.? ✓

ii. Do they span R^2?

(can a $\underline{\text{linear combination}}$ of $\underline{v_1}$ & v_2 produce any 2-D real vector?)

$\alpha_1 v_1 + \alpha_2 v_2$

Basis for R^2

② $v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$; $v_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$; $v_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

$\underline{\text{Linear dependence}}$

③ $\underline{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$; $\underline{v}_2 = \begin{bmatrix} 2 \\ -1 \\ 1 \end{bmatrix}$: Do $\underline{v}_1$ & $\underline{v}_2$ qualify as basis for R^3?

i. Linear dependence? ✓:

ii. Do $\underline{v}_1$ & $\underline{v}_2$ span R^3?

Basis vectors (for some space) } Neither more nor less.

Projections

$\boxed{v = \alpha w}$ ✗ $\quad \boxed{v = \alpha w + \underline{e}}$

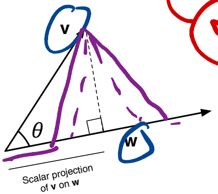
Idealistic. Reality

Orthogonal projection

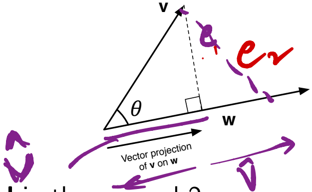
The **scalar** and **vector** orthogonal projection of a vector $\mathbf{v}$ onto another vector $\mathbf{w}$ are, respectively:

"prediction"

$$\boxed{\mathcal{P}_{\mathbf{w}}\mathbf{v} = \frac{\mathbf{v}.\mathbf{w}}{||\mathbf{w}||}}, \qquad \mathcal{P}_{\mathbf{w}}\mathbf{v} = \left(\frac{\mathbf{v}.\mathbf{w}}{||\mathbf{w}||}\right)\frac{\mathbf{w}}{||\mathbf{w}||} \tag{13}$$



α

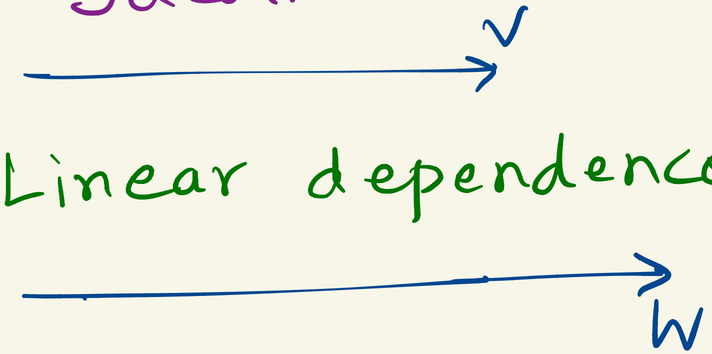


$||e_2|| > ||e_1||$

The **orthogonal** projection is unique and **optimal** in the squared 2-norm sense.

1

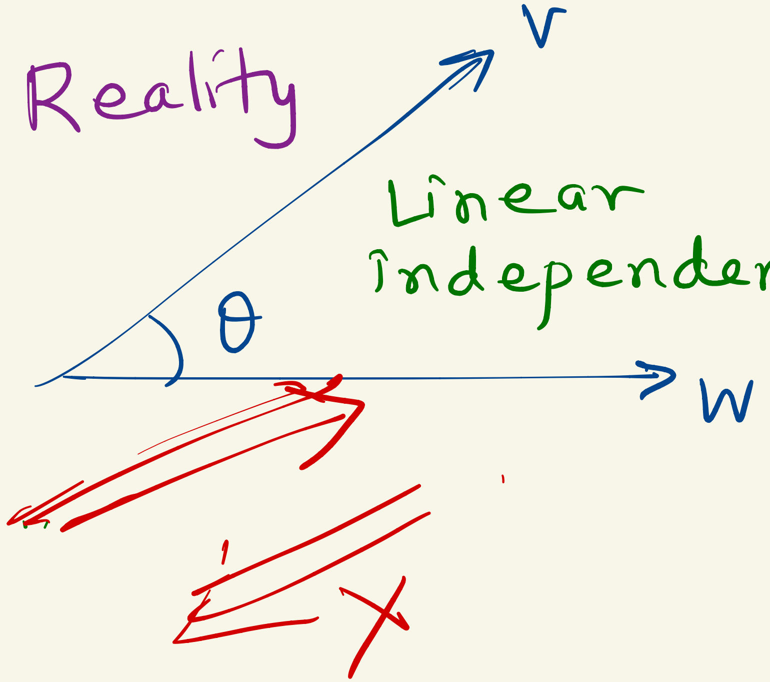
Idealism



Linear dependence

2

Reality

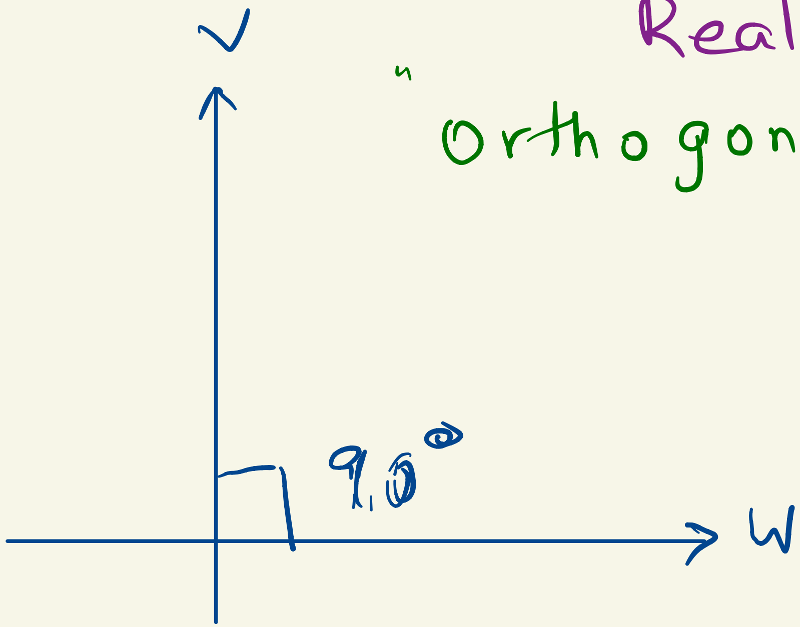


Linear independence

Reality

Orthogonal

3



$$\boxed{v_1 = \begin{bmatrix} h_1 \\ w_1 \end{bmatrix} ; \quad v_2 = \begin{bmatrix} h_2 \\ w_2 \end{bmatrix}}$$

OR

$$v_1 = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} ; \quad v_2 = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} \leftarrow$$