Fundamentals of linear algebra

- Scalar product and distance
- Orthogonality
- Orthogonal basis
- Orthogonal projections and decompositions
- Gram-Schmidt process
- Orthogonal matrices

Definition:

A real linear space E is called an Euclidian space, if there is an operation of *scalar product* defined for this space, such that

$\forall\,\{\mathbf{x},\,\mathbf{y},\,\mathbf{z}\}\in E$ and $\forall c\in\mathbb{R}$:

- $\mathbf{x}\cdot\mathbf{y}=\mathbf{y}\cdot\mathbf{x}$
- $(\mathbf{x}+\mathbf{y})\cdot\mathbf{z}=\mathbf{x}\cdot\mathbf{z}+\mathbf{y}\cdot\mathbf{z}$
- $(c\,\mathbf{x})\cdot\mathbf{y}=c\,(\mathbf{x}\cdot\mathbf{y})=\mathbf{x}\cdot(c\,\mathbf{y})$
- $\mathbf{x}\cdot\mathbf{x}\geqslant 0$, and
- $\mathbf{x}\cdot\mathbf{x}=0$ if and only if $\mathbf{x}=\mathbf{0}$

Definition: Elements $\mathbf{x}$ and $\mathbf{y}$ in E are called *orthogonal* to each other if $\mathbf{x}\cdot\mathbf{y}=0$. The notation is: $\mathbf{x}\perp\mathbf{y}$.

1. A set of all functions continuous on some $t \in [a, b]$ interval, is a Euclidian space with scalar product defined as

$$x(t) \cdot y(t) = \int_a^b x(t)y(t)\,\mathrm{d}t$$

2. For $\mathbb{P}_2$, scalar product can be also introduced as

$$p_1(t) \cdot p_2(t) = p_1(-1)p_2(-1) + p_1(0)p_2(0) + p_1(1)p_2(1)$$

3. In the space of continuous in $t \in [-\pi, \pi]$ interval functions with standard scalar product $f_1(t) \cdot f_2(t) = \int_{-\pi}^{\pi} f_1(t)f_2(t)\,\mathrm{d}t$, functions $\sin(t)$ and $\cos(t)$ are orthogonal.

Definition: For $\mathbf{v}$, $\mathbf{w} \in \mathbb{R}^n$ the scalar product (sometimes called also *inner product*, or *dot product*) is defined as: $\mathbf{v} \cdot \mathbf{w} = \sum_{i=1}^{n} v_i w_i$

Note: $\forall\, \mathbf{v} \in \mathbb{R}^n, \quad \mathbf{v} \perp \mathbf{0}$ because $\mathbf{0} \cdot \mathbf{v} = 0 \;\forall\, \mathbf{v}$

Note: Vectors in $\mathbb{R}^n$ can be written as: $\mathbf{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$

Their transposes are: $\mathbf{v}^\mathsf{T} = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix}, \quad \mathbf{w}^\mathsf{T} = \begin{bmatrix} w_1 & \dots & w_n \end{bmatrix}$

The result of $\mathbf{v}^\mathsf{T}\mathbf{w}$ is a 1×1 matrix, that is, a scalar:

$$\mathbf{v} \cdot \mathbf{w} \equiv \mathbf{v}^\mathsf{T}\mathbf{w} = \begin{bmatrix} v_1 & \dots & v_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = v_1 w_1 + v_2 w_2 + \ldots + v_n w_n$$

Definitions:

- The length, or *norm*, of $\mathbf{v}$ is: $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{\mathbf{v}^{\mathsf{T}}\mathbf{v}}$

  $\forall\, c, \quad \|c\mathbf{v}\| = |c|\, \|\mathbf{v}\|$

- A vector with a unit norm (length) is called a *unit vector*.

  If we divide a non-zero vector $\mathbf{v}$ by its length $\|\mathbf{v}\|$ we will obtain a *unit* (or *normalised*) vector $\mathbf{u}$ with unit norm:

$$\|\mathbf{u}\| = \left\| \frac{\mathbf{v}}{\|\mathbf{v}\|} \right\| = \frac{\|\mathbf{v}\|}{\|\mathbf{v}\|} = 1$$

Definition: The distance between two vectors $\mathbf{v}$ and $\mathbf{w}$ in $\mathbb{R}^n$ is

$$\mathrm{dist}(\mathbf{v}, \mathbf{w}) = \|\mathbf{v} - \mathbf{w}\|$$

Pythagorean theorem: If and only if $\mathbf{v} \perp \mathbf{w}$,

$$\|\mathbf{v} + \mathbf{w}\|^2 = \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2$$

Definition: If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$ then $\mathbf{z}$ is said to be orthogonal to W.

The set of all vectors that are orthogonal to W is called the *orthogonal complement* of W and is denoted by $W^\perp$.

Note: $W^\perp$ is a subspace of $\mathbb{R}^n$ and $\dim W^\perp + \dim W = n$.

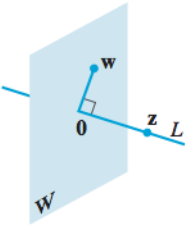
Example in $\mathbb{R}^3$:

Let W be a plane through the origin and L be a line through the origin and perpendicular to W.

If $\mathbf{z} \in L$ and $\mathbf{w} \in W$ then $\mathbf{z} \cdot \mathbf{w} = 0$.

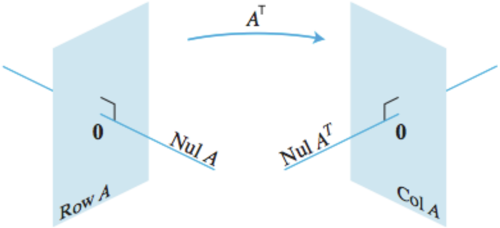
L consists of all vectors orthogonal to $\mathbf{w} \in W$ and W consists of all vectors orthogonal to $\mathbf{z} \in L$, so

$$L = W^\perp \quad \text{and} \quad W = L^\perp$$



Theorem: Let $\mathbf{A}$ be an $m \times n$ matrix. Then

$$(\mathrm{Row}\,\mathbf{A})^{\perp} = \mathrm{Nul}\,\mathbf{A} \qquad \text{and} \qquad (\mathrm{Col}\,\mathbf{A})^{\perp} = \mathrm{Nul}\big(\mathbf{A}^{\mathsf{T}}\big)$$



Proof: If $\mathbf{x} \in \mathrm{Nul}\,\mathbf{A}$, then $\mathbf{x}$ is orthogonal to each row of $\mathbf{A}$ because each row multiplied by $\mathbf{x}$ yields 0:

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Since the rows of $\mathbf{A}$ span $\mathrm{Row}\,\mathbf{A}$, $\mathbf{x}$ is orthogonal to $\mathrm{Row}\,\mathbf{A}$.

Conversely, if $\mathbf{x}$ is orthogonal to $\mathrm{Row}\,\mathbf{A}$, then $\mathbf{x}$ is certainly orthogonal to each row of $\mathbf{A}$, and hence $\mathbf{A}\mathbf{x} = \mathbf{0}$.

Thus $(\mathrm{Row}\,\mathbf{A})^{\perp} = \mathrm{Nul}\,\mathbf{A}$ is true for any matrix; so also for $\mathbf{A}^{\mathsf{T}}$;

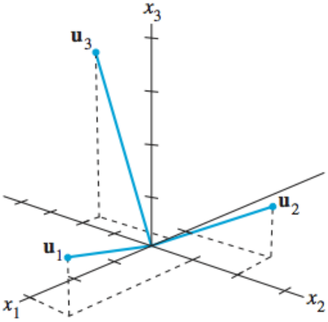
but $\mathrm{Row}\,\mathbf{A}^{\mathsf{T}} = \mathrm{Col}\,\mathbf{A}$, thus $(\mathrm{Col}\,\mathbf{A})^{\perp} = \Big(\mathrm{Row}\big(\mathbf{A}^{\mathsf{T}}\big)\Big)^{\perp} = \mathrm{Nul}\,\mathbf{A}^{\mathsf{T}}$.

Definition: A set of vectors $\{\mathbf{v}_1, \ldots \mathbf{v}_p\}$ in $\mathbb{R}^n$ is called an *orthogonal set* if each pair of vectors from the set is orthogonal:

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \qquad \forall\ i \neq j$$

Example: Show that this set is orthogonal:

$$\mathbf{u}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{u}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \ \mathbf{u}_3 = \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}.$$



Solution: Consider all the possible pairs:

$\mathbf{u}_1 \cdot \mathbf{u}_2 = 3 \cdot (-1) + 1 \cdot 2 + 1 \cdot 1 = 0$
$\mathbf{u}_1 \cdot \mathbf{u}_3 = 3 \cdot (-1) + 1 \cdot (-4) + 1 \cdot 7 = 0$
$\mathbf{u}_2 \cdot \mathbf{u}_3 = -1 \cdot (-1) + 2 \cdot (-4) + 1 \cdot 7 = 0$

Each pair is orthogonal, thus $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal set.

Theorem: If $S = \{\mathbf{v}_1, \ldots \mathbf{v}_p\} \in \mathbb{R}^n$ is an orthogonal set of non-zero vectors, then S is a linearly independent set.

Proof: Consider the equation for linear independence:

$$\mathbf{0} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_p\mathbf{v}_p$$

Multiply this relation by $\mathbf{v}_1$ on each side:

$$\begin{aligned} \mathbf{0} \cdot \mathbf{v}_1 &= (c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_p\mathbf{v}_p) \cdot \mathbf{v}_1 \\ 0 &= c_1(\mathbf{v}_1 \cdot \mathbf{v}_1) + c_2(\mathbf{v}_2 \cdot \mathbf{v}_1) + \ldots + c_p(\mathbf{v}_p \cdot \mathbf{v}_1) \\ 0 &= c_1(\mathbf{v}_1 \cdot \mathbf{v}_1) \end{aligned}$$

only the first term on the right remains since $\mathbf{v}_1 \perp \{\mathbf{v}_2, \ldots \mathbf{v}_p\}$. However $\mathbf{v}_1 \neq \mathbf{0}$ so $\mathbf{v}_1 \cdot \mathbf{v}_1 \neq 0$ and thus we must have $c_1 = 0$. Similarly, multiplying by $\mathbf{v}_2, \ldots \mathbf{v}_p$, we find $c_2, \ldots c_p$ are all zero. Thus S is linearly independent.

Definition: An *orthogonal basis* for a subspace V of $\mathbb{R}^n$ is such a basis for V which is an orthogonal set.

Coordinates with respect to an orthogonal basis are easily found:

Theorem: Let $\{\mathbf{v}_1, \dots \mathbf{v}_p\}$ be an orthogonal basis for a subspace V of $\mathbb{R}^n$. For each $\mathbf{x} \in V$, the linear combination

$$\mathbf{x} = c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p \qquad \text{has the weights} \qquad c_i = \frac{\mathbf{x} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$$

Proof: Compute all the scalar products, as in the previous proof

$$\mathbf{x} \cdot \mathbf{v}_i = (c_1\mathbf{v}_1 + \ldots + c_p\mathbf{v}_p) \cdot \mathbf{v}_i = c_i(\mathbf{v}_i \cdot \mathbf{v}_i)$$

Since $\mathbf{v}_i \neq \mathbf{0}$ (why?) then $\mathbf{v}_i \cdot \mathbf{v}_i \neq 0$ and so $c_i = \dfrac{\mathbf{x} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}$

Orthogonal basis, example

Example: Find $[\mathbf{y}]_{\mathcal{B}}$ in the orthogonal basis $\mathcal{B} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$:

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} -1 \\ -4 \\ 7 \end{bmatrix}; \qquad \mathbf{y} = \begin{bmatrix} 6 \\ 1 \\ -8 \end{bmatrix}$$

Solution: To find the coordinates $[\mathbf{y}]_{\mathcal{B}}$ we compute

$$\begin{aligned} \mathbf{y} \cdot \mathbf{v}_1 &= 6 \cdot 3 + 1 \cdot 1 - 8 \cdot 1 = 11, & \mathbf{v}_1 \cdot \mathbf{v}_1 &= 3^2 + 1^2 + 1^2 = 11 \\ \mathbf{y} \cdot \mathbf{v}_2 &= -6 \cdot 1 + 1 \cdot 2 - 8 \cdot 1 = -12, & \mathbf{v}_2 \cdot \mathbf{v}_2 &= (-1)^2 + 2^2 + 1^2 = 6 \\ \mathbf{y} \cdot \mathbf{v}_3 &= -6 \cdot 1 - 1 \cdot 4 - 8 \cdot 7 = -66, & \mathbf{v}_3 \cdot \mathbf{v}_3 &= (-1)^2 + (-4)^2 + 7^2 = 66 \end{aligned}$$

Thus $\quad \mathbf{y} = \dfrac{11}{11}\mathbf{v}_1 - \dfrac{12}{6}\mathbf{v}_2 - \dfrac{66}{66}\mathbf{v}_3 \quad$ and $\quad [\mathbf{y}]_{\mathcal{B}} = \begin{pmatrix} 1 \\ -2 \\ -1 \end{pmatrix}$

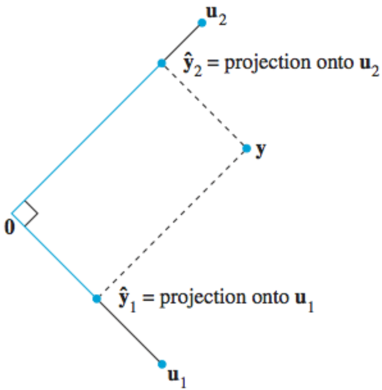
It is quite easy to find coordinates in an orthogonal basis.
For a non-orthogonal case we would need to solve a linear system.

For two orthogonal basis vectors $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^2$:

$$\mathbf{y} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2$$

with

$$c_k = \frac{\mathbf{y} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k}$$



The first term is the projection of $\mathbf{y}$ onto the line $\mathrm{Span}\{\mathbf{u}_1\}$, and the second term is the projection of $\mathbf{y}$ onto the line $\mathrm{Span}\{\mathbf{u}_2\}$.

Orthogonal projections

Given a non-zero vector $\mathbf{u} \in \mathbb{R}^n$, consider decomposing another vector $\mathbf{y} \in \mathbb{R}^n$ into the sum of two vectors, such that

$$\mathbf{y} = \check{\mathbf{y}} + \mathbf{z}, \qquad \check{\mathbf{y}} = \alpha\mathbf{u}, \qquad \mathbf{z} \perp \mathbf{u}$$

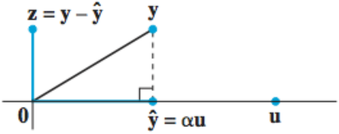
Consider $\mathbf{z} = \mathbf{y} - \alpha\mathbf{u}$, which is orthogonal to $\mathbf{u}$ if and only if

$$0 = \mathbf{z} \cdot \mathbf{u} = (\mathbf{y} - \alpha\mathbf{u}) \cdot \mathbf{u} = \mathbf{y} \cdot \mathbf{u} - \alpha(\mathbf{u} \cdot \mathbf{u})$$

Hence

$$\alpha = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}, \qquad \check{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}}{\mathbf{u} \cdot \mathbf{u}}\,\mathbf{u}$$

$\check{\mathbf{y}}$ is the *orthogonal projection* of $\mathbf{y}$ onto $\mathbf{u}$, and $\mathbf{z}$ is called the component orthogonal to $\mathbf{u}$.



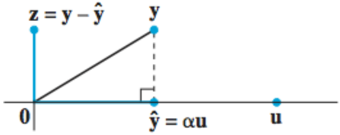
The orthogonal projection $\check{\mathbf{y}}$ does not depend on the length of $\mathbf{u}$. Indeed, if we replace $\mathbf{u}$ by $\mathbf{u}' = k\mathbf{u}$, then

$$\check{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u}'}{\mathbf{u}'\cdot\mathbf{u}'}\,\mathbf{u}' = \frac{\mathbf{y}\cdot(k\mathbf{u})}{(k\mathbf{u})\cdot(k\mathbf{u})}\,(k\mathbf{u}) = \frac{k\,(\mathbf{y}\cdot\mathbf{u})}{k^2\,(\mathbf{u}\cdot\mathbf{u})}\,k\,\mathbf{u} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\,\mathbf{u}.$$

Definition:

$$\check{\mathbf{y}} \equiv \mathrm{proj}_L\,\mathbf{y} = \frac{\mathbf{y}\cdot\mathbf{u}}{\mathbf{u}\cdot\mathbf{u}}\,\mathbf{u}$$

is the *orthogonal projection* of $\mathbf{y}$ onto $L = \mathrm{Span}\{\mathbf{u}\}$



(subspace $L = \mathrm{Span}\{\mathbf{u}\}$ is a line through $\mathbf{u}$ and $\mathbf{0}$).

Example: Find orthogonal projection of $\mathbf{y}$ onto $\mathbf{v}$ and decompose $\mathbf{y}$ as the sum of $\check{\mathbf{y}} \in \mathrm{Span}\{\mathbf{v}\}$ and $\mathbf{z} \perp \mathbf{v}$, given

$$\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}, \qquad \mathbf{v} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}.$$

Solution: First we compute $\mathbf{y} \cdot \mathbf{v}$ and $\mathbf{v} \cdot \mathbf{v}$:

$$\mathbf{y} \cdot \mathbf{v} = \mathbf{y}^{\mathsf{T}}\mathbf{v} = \begin{bmatrix} 7 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40,$$

$$\mathbf{v} \cdot \mathbf{v} = \mathbf{u}^{\mathsf{T}}\mathbf{v} = \begin{bmatrix} 4 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20.$$

Then the orthogonal projection and the orthogonal component are

$$\check{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v}}{\mathbf{v} \cdot \mathbf{v}}\,\mathbf{v} = \frac{40}{20}\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

$$\mathbf{z} = \mathbf{y} - \check{\mathbf{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

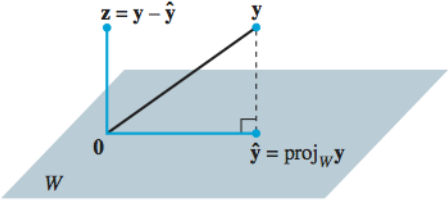
Theorem: (*the orthogonal decomposition theorem*)

Let W be a subspace of $\mathbb{R}^n$.

Then $\forall\, \mathbf{y} \in \mathbb{R}^n$ there is a unique decomposition

$$\mathbf{y} = \check{\mathbf{y}} + \mathbf{z},$$

where $\check{\mathbf{y}} \in W$ and $\mathbf{z} \in W^{\perp}$.



If $\{\mathbf{v}_1, \ldots \mathbf{v}_p\}$ is any orthogonal basis in W,

$$\check{\mathbf{y}} = \sum_{i=1}^{p} \frac{\mathbf{y} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}\, \mathbf{v}_i \qquad \text{and} \qquad \mathbf{z} = \mathbf{y} - \check{\mathbf{y}}.$$

$\check{\mathbf{y}} \equiv \mathrm{proj}_W\, \mathbf{y}$ is called the *orthogonal projection* of $\mathbf{y}$ onto W.

Notes:

- The uniqueness of the decomposition indicates that the orthogonal projection $\check{\mathbf{y}}$ depends only on W but not on a particular orthogonal basis used in W.
- If $\mathbf{y} \in W$ then $\operatorname{proj}_W \mathbf{y} = \mathbf{y}$.

Example: Let

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

The set $\{\mathbf{v}_1, \mathbf{v}_2\}$ is an orthogonal basis for $W = \operatorname{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$

(check: $\mathbf{v}_1 \cdot \mathbf{v}_2 = -4 + 5 - 1 = 0$ so the vectors are orthogonal).

Decompose $\mathbf{y}$ into a vector in W and a vector orthogonal to W.

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Solution:

$$\check{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{y} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2.$$

$$\mathbf{y} \cdot \mathbf{v}_1 = 2 + 10 - 3 = 9 \qquad \mathbf{y} \cdot \mathbf{v}_2 = -2 + 2 + 3 = 3$$

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 4 + 25 + 1 = 30 \qquad \mathbf{v}_2 \cdot \mathbf{v}_2 = 4 + 1 + 1 = 6$$

$$\check{\mathbf{y}} = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

and

$$\mathbf{y} - \check{\mathbf{y}} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}.$$

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad \check{\mathbf{y}} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}, \quad \mathbf{y} - \check{\mathbf{y}} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}.$$

The theorem ensures that $(\mathbf{y} - \check{\mathbf{y}}) \in W^{\perp}$.

We can verify that $(\mathbf{y} - \check{\mathbf{y}}) \cdot \mathbf{v}_1 = 0$ and $(\mathbf{y} - \check{\mathbf{y}}) \cdot \mathbf{v}_2 = 0$.

So the decomposition is

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}.$$

Definition: The distance from a point $\mathbf{y}$ in $\mathbb{R}^n$ to a subspace W is defined as the distance from $\mathbf{y}$ to the nearest point in W.

Theorem: *(the best approximation theorem)*

Let W be a subspace of $\mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$, and $\check{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$.
Then $\check{\mathbf{y}}$ is the point in W closest to $\mathbf{y}$ in the sense that

$$\|\mathbf{y} - \check{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \qquad \forall\, \mathbf{v} \neq \check{\mathbf{y}}$$

Consequence:

The nearest point to $\mathbf{y}$ in W is its projection $\check{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$,
and the distance from $\mathbf{y}$ to W is given by $\|\check{\mathbf{y}} - \mathbf{y}\|$.

Example: Find the distance from $\mathbf{y}$ to $W = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ where

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}$$

Solution: Distance from $\mathbf{y}$ to W is $\|\mathbf{y} - \check{\mathbf{y}}\|$, where $\check{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$.

Vectors $\mathbf{v}_1, \mathbf{v}_2$ form an orthogonal basis for W, and

$$\mathbf{y} \cdot \mathbf{v}_1 = -5 + 10 + 10 = 15 \qquad \mathbf{y} \cdot \mathbf{v}_2 = -1 - 10 - 10 = -21$$

$$\mathbf{v}_1 \cdot \mathbf{v}_1 = 25 + 4 + 1 = 30 \qquad \mathbf{v}_2 \cdot \mathbf{v}_2 = 1 + 4 + 1 = 6$$

Then

$$\check{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 + \frac{\mathbf{y} \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 = \frac{15}{30}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{21}{6}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$\mathbf{v}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad \check{\mathbf{y}} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

Then

$$\mathbf{y} - \check{\mathbf{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

and the distance from $\mathbf{y}$ to W is

$$\|\mathbf{y} - \check{\mathbf{y}}\| = \sqrt{0^2 + 3^2 + 6^2} = \sqrt{45} = 3\sqrt{5}$$

So, $\check{\mathbf{y}}$ is the best approximation for $\mathbf{y}$ within $\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$; any other vector in $\mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2\}$ will have a greater distance from $\mathbf{y}$.

Definition: A set $\{\mathbf{u}_1, \dots \mathbf{u}_p\}$ is called an *orthonormal set* if it is an orthogonal set of unit vectors.

If V is the subspace spanned by such a set, then $\{\mathbf{u}_1, \dots \mathbf{u}_p\}$ is an *orthonormal basis* for V, since this set is linearly independent.

The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1, \dots \mathbf{e}_n\}$ for $\mathbb{R}^n$.

Any non-empty subset of $\{\mathbf{e}_1, \dots \mathbf{e}_n\}$ is orthonormal too, and forms an orthonormal basis for the corresponding subspace.

Theorem: If $\mathbf{U} = [\mathbf{u}_1 \; \mathbf{u}_2 \; \ldots \; \mathbf{u}_p]$, where $\{\mathbf{u}_1, \ldots \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then $\forall \, \mathbf{y} \in \mathbb{R}^n$:

- $\mathrm{proj}_W \, \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1) \, \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2) \, \mathbf{u}_2 + \ldots + (\mathbf{y} \cdot \mathbf{u}_p) \, \mathbf{u}_p$
- $\mathrm{proj}_W \, \mathbf{y} = \mathbf{U}\mathbf{U}^{\mathsf{T}}\mathbf{y}$

Proof: By the definition of projection

$$\check{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \, \mathbf{u}_1 + \ldots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \, \mathbf{u}_p$$

and taking into account that the basis is orthonormal

$$\mathbf{u}_1 \cdot \mathbf{u}_1 = 1, \quad \mathbf{u}_2 \cdot \mathbf{u}_2 = 1, \quad \ldots \quad \mathbf{u}_p \cdot \mathbf{u}_p = 1$$

we immediately obtain

$$\mathrm{proj}_W \, \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1) \, \mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2) \, \mathbf{u}_2 + \ldots + (\mathbf{y} \cdot \mathbf{u}_p) \, \mathbf{u}_p.$$

Proof (continuing):

From $\quad \mathrm{proj}_W\, \mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\,\mathbf{u}_2 + \ldots + (\mathbf{y}\cdot\mathbf{u}_p)\,\mathbf{u}_p$
we see that $\mathrm{proj}_W\, \mathbf{y}$ is a linear combination of the columns of $\mathbf{U}$
with the coefficients $(\mathbf{y}\cdot\mathbf{u}_1),\ (\mathbf{y}\cdot\mathbf{u}_2),\ \ldots\ (\mathbf{y}\cdot\mathbf{u}_p)$.

Denoting $\mathbf{x} = \begin{bmatrix} \mathbf{y}\cdot\mathbf{u}_1 \\ \mathbf{y}\cdot\mathbf{u}_2 \\ \vdots \\ \mathbf{y}\cdot\mathbf{u}_p \end{bmatrix}$ we can write $\mathrm{proj}_W\, \mathbf{y} = \mathbf{U}\mathbf{x}$.

In turn, the elements of $\mathbf{x}$ can be written as

$$\mathbf{u}_1^\mathsf{T}\mathbf{y}, \quad \mathbf{u}_2^\mathsf{T}\mathbf{y}, \quad \ldots \quad \mathbf{u}_p^\mathsf{T}\mathbf{y}$$

which are the entries of $\mathbf{U}^\mathsf{T}\mathbf{y}$.

Thus $\mathbf{x} = \mathbf{U}^\mathsf{T}\mathbf{y}$ and so $\mathrm{proj}_W\, \mathbf{y} = \mathbf{U}\mathbf{U}^\mathsf{T}\mathbf{y}$

Consider a matrix formed from orthonormal vectors (as columns).

Theorem: $\mathbf{U}$ has orthonormal columns if and only if $\mathbf{U}^{\mathsf{T}}\mathbf{U} = \mathbf{I}$.

Theorem: If $\mathbf{U}$ has orthonormal columns, then $\forall\, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

(i) $\|\mathbf{U}\mathbf{x}\| = \|\mathbf{x}\|$

(ii) $(\mathbf{U}\mathbf{x}) \cdot (\mathbf{U}\mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$

(iii) $(\mathbf{U}\mathbf{x}) \cdot (\mathbf{U}\mathbf{y}) = 0$ if and only if $\mathbf{x} \cdot \mathbf{y} = 0$

These properties imply that the linear transformation $\mathbf{x} \mapsto \mathbf{U}\mathbf{x}$ preserves length, scalar product, and orthogonality.

Definition: An *orthogonal matrix* $\mathbf{U}$ is a square invertible matrix such that $\mathbf{U}^{-1} = \mathbf{U}^{\mathsf{T}}$.

Equivalent definitions:

- $\mathbf{U}$ has orthonormal columns and orthonormal rows:

$$\sum_{k=1}^{n} u_{ki}u_{kj} = \delta_{ij} \qquad \text{and} \qquad \sum_{k=1}^{n} u_{ik}u_{jk} = \delta_{ij}$$

- Given two different orthonormal bases in $\mathbb{R}^n$, the change of basis matrices between such bases are orthogonal matrices.

Note: For an orthogonal matrix $\mathbf{U}$: $|\det \mathbf{U}| = 1$

(warning: but $|\det \mathbf{A}| = 1$ does not mean that $\mathbf{A}$ is orthogonal)

Example 1: The most trivial example: unitary matrix in $\mathbb{R}^n$

Example 2: Matrix of a rotation transformation in $\mathbb{R}^2$:

$$\mathbf{R}_\varphi = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$$

Example 3: Permutation matrix, for example cyclic permutation:

$$\mathbf{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Theorem: Given a basis $\mathbf{x}_1, \dots \mathbf{x}_p$ for a subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\,\mathbf{v}_2 \\
&\dots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\,\mathbf{v}_2 - \dots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\,\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \dots \mathbf{v}_p\}$ is an orthogonal basis for W.

In addition, $\mathrm{Span}\{\mathbf{v}_1, \dots \mathbf{v}_k\} = \mathrm{Span}\{\mathbf{x}_1, \dots \mathbf{x}_k\}$ for $1 \leqslant k \leqslant p$.

Orthonormal basis is then obtained by normalising $\mathbf{v}_i$ to unit vectors.

Gram-Schmidt process is an algorithm for producing an orthogonal or orthonormal basis for any non-zero subspace of $\mathbb{R}^n$.

Example: Consider a linearly independent set

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

which is a basis for a subspace W in $\mathbb{R}^4$.

We aim to construct an orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ for W.

Solution:

Step 1: Let $\mathbf{v}_1 = \mathbf{x}_1$ and $W_1 = \mathrm{Span}\{\mathbf{x}_1\} = \mathrm{Span}\{\mathbf{v}_1\}$.

Step 2: Vector $\mathbf{v}_2$ is then produced by subtracting from $\mathbf{x}_2$ its projection onto the subspace W_1. That is,

$$\mathbf{v}_2 = \mathbf{x}_2 - \mathrm{proj}_{W_1}(\mathbf{x}_2) = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1.$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

For the ease of further calculations, renormalise $\mathbf{v}_2$ into $\mathbf{v}_2'$:

$$\mathbf{v}_2' = 4\mathbf{v}_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{and then} \quad W_2 = \mathrm{Span}\{\mathbf{v}_1, \mathbf{v}_2'\}$$

Step 3: Get $\mathbf{v}_3$ by subtracting its W_2-projection from $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$:

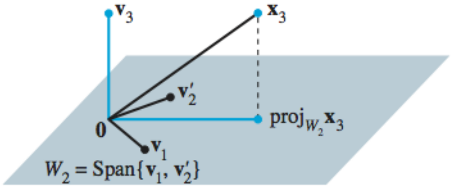
$$\operatorname{proj}_{W_2}(\mathbf{x}_3) = \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\, \mathbf{v}_1 + \frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\, \mathbf{v}_2'$$

$$= \frac{2}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} + \frac{2}{12} \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix}.$$

Then $\mathbf{v}_3 = \mathbf{x}_3 - \operatorname{proj}_{W_2}(\mathbf{x}_3)$ is

$$\mathbf{v}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 \\ 2/3 \\ 2/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}; \qquad \mathbf{v}_3' = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

Illustration to step 3:



Thus $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3\}$ is an orthogonal set in W and it is basis for W.

The same is true for $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ or $\{\mathbf{v}_1, \mathbf{v}_2', \mathbf{v}_3'\}$ or $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3'\}$

Example: Construct an orthonormal basis for

$$\operatorname{Span}\left\{ \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \right\}$$

Solution:

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} - \frac{\overbrace{3 \cdot 1 + 6 \cdot 2 + 0 \cdot 2}^{15}}{\underbrace{3^2 + 6^2 + 0^2}_{45}} \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}.$$

The corresponding orthonormal basis is

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}, \qquad \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

But what happens if we swap the order of the vectors?

$$\mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} \right\}$$

Then:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \qquad \mathbf{v}_2 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} - \frac{\overbrace{3 \cdot 1 + 6 \cdot 2 + 0 \cdot 2}^{15}}{\underbrace{1^2 + 2^2 + 2^2}_{9}} \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 8/3 \\ -10/3 \end{bmatrix}.$$

The corresponding orthonormal basis is (please check!)

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|} = \begin{bmatrix} 1/3 \\ 2/3 \\ 2/3 \end{bmatrix}, \qquad \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|} = \begin{bmatrix} 2/3\sqrt{5} \\ 4/3\sqrt{5} \\ -\sqrt{5}/3 \end{bmatrix}$$

- Scalar product $\mathbf{u} \cdot \mathbf{v} \equiv \mathbf{u}^\mathsf{T} \mathbf{v} = \sum_i \mathbf{u}_i \mathbf{v}_i$
- Norm: $\|\mathbf{v}\| = \sqrt{\mathbf{v} \cdot \mathbf{v}}$ and distance: $\operatorname{dist}(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$
- Orthogonal / orthonormal vectors, sets, bases
- Orthogonal complements $\dim W^\perp + \dim W = n$
- Orthogonal projections and decompositions
- Gram-Schmidt process: $\mathbf{v}_1 = \mathbf{x}_1$, then

$$\mathbf{v}_i = \mathbf{x}_i + \sum_{j=1}^{i-1} \left(-\frac{\mathbf{x}_i \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j} \, \mathbf{v}_j \right) \qquad i = 2, \ldots p$$

- Orthogonal and orthonormal matrices

Questions?

This week: quick test 7

(matrix subspaces)

Next week: quick test 8

(orthogonality)