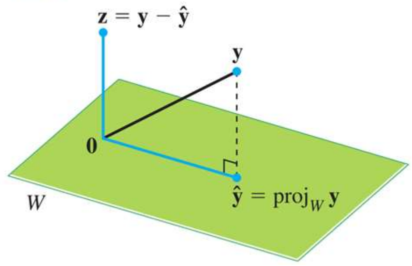
# Orthogonal projections

- Let $W$ be a sub vector space of $\mathbb{R}^n$, let $\mathbf{y} \in \mathbb{R}^n$ be a point. A natural question is to find the distance between $\mathbf{y}$ and $W$, i.e. to look for a point $\mathbf{y}_0 \in W$ such that

$$\| \mathbf{y} - \mathbf{y}_0 \| = \min\{\| \mathbf{y} - \mathbf{w} \| \mid \mathbf{w} \in W\}$$

and to calculate the minimum.

- Intuitively, $\mathbf{y}_0$ will be the projection of $\mathbf{y}$




- Proposition: Let $W$ be a subspace of $\mathbb{R}^n$, let $W^\perp$ be its orthogonal complement, then $W \cap W^\perp = \{0\}$ and $W + W^\perp = \mathbb{R}^n$.

Proof For the first assertion, let $\vec{w} \in W \cap W^\perp$, then

$$\langle \underset{W}{\underset{\cap}{\vec{w}}}, \underset{W^\perp}{\underset{\cap}{\vec{w}}} \rangle = 0$$

$$\Rightarrow \|\vec{w}\|^2 = 0 \qquad \Rightarrow \vec{w} = 0.$$

For the second assertion, we show firstly that

$$\dim(W) + \dim(W^\perp) = n.$$

Take a basis $\{\vec w_1, \cdots, \vec w_p\}$ of $W$, let $A = [\vec w_1 \cdots \vec w_p]$. then

$$W = \text{Span}\{\vec w_1, \cdots, \vec w_p\} = \text{Col}(A)$$

and
$$W^\perp = \{\vec v \in \mathbb{R}^n \mid \langle \vec v, \vec w \rangle = 0,\ \forall \vec w \in W\}$$
$$= \{\vec v \in \mathbb{R}^n \mid \langle \vec v, \vec w_i \rangle = 0,\ i = 1, \cdots, p\}$$
$$= \text{Nul}(A).$$

By a previous theorem, $\dim(\text{Col}(A)) + \dim \text{Nul}(A) = n$.

Hence $\dim(W) + \dim(W^\perp) = n$.

---

To show $W + W^\perp = \mathbb{R}^n$, i.e. the union of a basis $\{\vec w_1, \cdots, \vec w_p\}$ for $W$ and a basis $\{\vec v_1, \cdots, \vec v_{n-p}\}$ of $W^\perp$ forms a basis of $\mathbb{R}^n$, it is enough to show that they are linearly independent. Suppose that

$$\underbrace{\lambda_1 \vec w_1 + \cdots + \lambda_p \vec w_p}_{w_0 \in W} + \underbrace{\mu_1 \vec v_1 + \cdots + \mu_{n-p} \vec v_p}_{v_0 \in W^\perp} = 0$$

for some $\lambda_1, \cdots, \lambda_p, \mu_1, \cdots, \mu_{n-p} \in \mathbb{R}$. Then for any $\vec w \in W$,

$$0 = \langle \vec w, \vec w_0 + \vec v_0 \rangle = \langle \vec w, \vec w_0 \rangle + \langle \vec w, \underset{\in W^\perp}{\vec v_0} \rangle = \langle \vec w, \vec w_0 \rangle$$

$$\Rightarrow \vec w_0 \in W^\perp \Rightarrow \vec w_0 \in W^\perp \cap W = \{0\} \Rightarrow \vec w_0 = 0$$

$$\Rightarrow \quad \lambda_1 \vec{w}_1 + \cdots + \lambda_p \vec{w}_p = \vec{w}_0 = 0$$

$$\Rightarrow \quad \lambda_1 = \cdots = \lambda_p = 0.$$

Similarly, for any $\vec{v} \in W^\perp$,

$$0 = \langle \vec{v}, \vec{w}_0 + \vec{v}_0 \rangle = \langle \vec{v}, \vec{w}_0 \rangle + \langle \vec{v}, \vec{v}_0 \rangle = \langle \vec{v}, \vec{v}_0 \rangle$$

$$\langle \vec{v}, \vec{w}_0 \rangle = 0$$

$$\Rightarrow \quad \vec{v}_0 \in (W^\perp)^\perp = W \quad \Rightarrow \quad \vec{v}_0 \in W^\perp \cap W = \{0\}$$

$$\Rightarrow \quad \vec{v}_0 = 0$$

$$\Rightarrow \quad \mu_1 \vec{v}_1 + \cdots + \mu_{n-p} \vec{v}_{n-p} = 0$$

$$\Rightarrow \quad \mu_1 = \cdots = \mu_{n-p} = 0.$$

- Theorem

**The Orthogonal Decomposition Theorem**

Let $W$ be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \tag{1}$$

where $\hat{\mathbf{y}}$ is in $W$ and $\mathbf{z}$ is in $W^\perp$. In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of $W$, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p} \mathbf{u}_p \tag{2}$$

and $\mathbf{z} = \mathbf{y} - \hat{\mathbf{y}}$.

Proof: By the previous proposition, let $\vec{u}_1, \cdots, \vec{u}_p$ be a basis of $W$, let $\vec{v}_1, \cdots, \vec{v}_{n-p}$ be a basis of $W^\perp$, then $\{\vec{u}_1, \cdots, \vec{u}_p, \vec{v}_1, \cdots, \vec{v}_{n-p}\}$ forms a basis of $\mathbb{R}^n$.

Hence any $y \in \mathbb{R}^n$ can be written uniquely as

$$y = \underbrace{\lambda_1 \vec{u}_1 + \cdots + \lambda_p \vec{u}_p}_{\hat{y} \in W} + \underbrace{\mu_1 \vec{v}_1 + \cdots + \mu_{n-p} \vec{v}_{n-p}}_{z \in W^\perp}.$$

If $\vec{u}_1, \cdots, \vec{u}_p$ is an orthogonal basis, then

$$\langle \vec{y}, \vec{u}_i \rangle = \lambda_1 \underbrace{\langle \vec{u}_1, \vec{u}_i \rangle}_{=0} + \cdots + \lambda_i \underbrace{\langle \vec{u}_i, \vec{u}_i \rangle}_{\neq 0} + \cdots + \lambda_p \underbrace{\langle \vec{u}_p, \vec{u}_i \rangle}_{=0} + \underbrace{\langle \underset{W^\perp \ni}{z}, \vec{u}_i \rangle}_{=0} = \lambda_i \langle \vec{u}_i, \vec{u}_i \rangle$$

$$\Rightarrow \quad \lambda_i = \frac{\langle \vec{y}, \vec{u}_i \rangle}{\langle \vec{u}_i, \vec{u}_i \rangle}$$

So

$$\hat{y} = \lambda_1 \vec{u}_1 + \cdots + \lambda_p \vec{u}_p = \frac{\langle \vec{y}, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \cdot \vec{u}_1 + \cdots + \frac{\langle \vec{y}, \vec{u}_p \rangle}{\langle \vec{u}_p, \vec{u}_p \rangle} \cdot \vec{u}_p,$$

and $z = \vec{y} - \hat{y}$. $\square$

**EXAMPLE 2** Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write $\mathbf{y}$ as the sum of a vector in $W$ and a vector orthogonal to $W$.

Solution: One verifies that

$$\langle \vec{u}_1, \vec{u}_2 \rangle = 2 \times (-2) + 5 \times 1 + (-1) \times 1 = -4 + 5 - 1 = 0,$$

i.e. $\vec{u}_1, \vec{u}_2$ are orthogonal. With the projection formula

$$\text{proj}_W(\vec{y}) = \frac{\langle y, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \vec{u}_1 + \frac{\langle y, \vec{u}_2 \rangle}{\langle \vec{u}_2, \vec{u}_2 \rangle} \vec{u}_2,$$

we calculate:

$$\langle \vec{u}_1, \vec{u}_1 \rangle = 2 \times 2 + 5 \times 5 + (-1) \times (-1) = 4 + 25 + 1 = 30,$$

$$\langle \vec{y}, \vec{u}_1 \rangle = 1 \times 2 + 2 \times 5 + 3 \times (-1) = 2 + 10 - 3 = 9.$$

$$\langle \vec{u}_2, \vec{u}_2 \rangle = (-2)^2 + 1^2 + 1^2 = 4 + 1 + 1 = 6$$

$$\langle \vec{y}, \vec{u}_2 \rangle = 1 \times (-2) + 2 \times 1 + 3 \times 1 = -2 + 2 + 3 = 3.$$

$$\Rightarrow \text{proj}_W(\vec{y}) = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

$$= \frac{1}{10} \begin{bmatrix} 6 \\ 15 \\ -3 \end{bmatrix} + \begin{bmatrix} -1 \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -0.4 \\ 2 \\ 0.2 \end{bmatrix},$$

and $\vec{y} - \text{proj}_W(\vec{y}) = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -0.4 \\ 2 \\ 0.2 \end{bmatrix} = \begin{bmatrix} 0.6 \\ 0 \\ 2.8 \end{bmatrix}.$ $\square$

- Geometric interpretation of the formula:

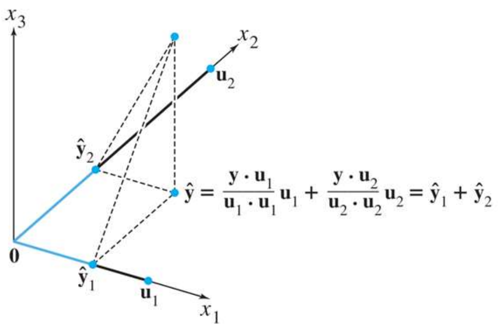
Let $\vec{u}_1, \cdots, \vec{u}_p$ be an orthogonal basis of W, let $L_i$ be the line spanned by $\vec{u}_i$. Recall that

$$\frac{\langle \vec{y}, \vec{u}_i \rangle}{\langle \vec{u}_i, \vec{u}_i \rangle} \vec{u}_i = \mathrm{Proj}_{L_i}(\vec{y}).$$

hence

$$\hat{y} = \frac{\langle \vec{y}, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \cdot \vec{u}_1 + \cdots + \frac{\langle \vec{y}, \vec{u}_p \rangle}{\langle \vec{u}_p, \vec{u}_p \rangle} \cdot \vec{u}_p,$$

$$= \mathrm{Proj}_{L_1}(\vec{y}) + \cdots + \mathrm{Proj}_{L_p}(\vec{y}).$$




- Theorem

**The Best Approximation Theorem**

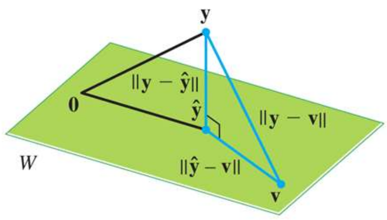
Let $W$ be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto $W$. Then $\hat{\mathbf{y}}$ is the closest point in $W$ to $\mathbf{y}$, in the sense that

$$\|\mathbf{y} - \hat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\| \qquad (3)$$

for all $\mathbf{v}$ in $W$ distinct from $\hat{\mathbf{y}}$.

Proof: From the picture, it is clear that $\vec{y}, \hat{y}, \vec{v}$ forms a right triangle, hence

$$\|\vec{y} - \hat{y}\| < \|\vec{y} - \vec{v}\|.$$



**EXAMPLE 4** 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 $W$. Find the distance from $\mathbf{y}$ to $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, where

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

Solution One verifies

$$\langle \vec{u}_1, \vec{u}_2 \rangle = 5 \times 1 + (-2) \times 2 + 1 \times (-1) = 5 - 4 - 1 = 0,$$

hence $\vec{u}_1$ is orthogonal to $\vec{u}_2$, they form an orthogonal basis of $W$.

With the projection formula

$$\text{Proj}_W(\vec{y}) = \frac{\langle \vec{y}, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle} \vec{u}_1 + \frac{\langle \vec{y}, \vec{u}_2 \rangle}{\langle \vec{u}_2, \vec{u}_2 \rangle} \vec{u}_2,$$

we calculate

$$\langle \vec{u}_1, \vec{u}_1 \rangle = 5^2 + (-2)^2 + 1^2 = 25 + 4 + 1 = 30$$

$$\langle \vec{y}, \vec{u}_1 \rangle = (-1) \times 5 + (-5) \times (-2) + 10 \times 1 = -5 + 10 + 10 = 15.$$

$$\langle \vec{u}_2, \vec{u}_2 \rangle = 1^2 + 2^2 + (-1)^2 = 1 + 4 + 1 = 6$$

$$\langle \vec{y}, \vec{u}_2 \rangle = (-1) \times 1 + (-5) \times 2 + 10 \times (-1) = -1 - 10 - 10 = -21$$

$$\Rightarrow \text{Proj}_W(\vec{y}) = \frac{15}{30}\begin{pmatrix}5\\-2\\1\end{pmatrix} + \frac{-21}{6}\begin{pmatrix}1\\2\\-1\end{pmatrix}$$

$$= \begin{pmatrix}\frac{5}{2}\\-1\\\frac{1}{2}\end{pmatrix} - \begin{pmatrix}\frac{7}{2}\\7\\-\frac{7}{2}\end{pmatrix} = \begin{pmatrix}-1\\-8\\-3\end{pmatrix}$$

$$\Rightarrow \vec{y} - \text{Proj}_W(\vec{y}) = \begin{pmatrix}-1\\-5\\10\end{pmatrix} - \begin{pmatrix}-1\\-8\\-3\end{pmatrix} = \begin{pmatrix}0\\3\\13\end{pmatrix}$$

So the distance between $\vec{y}$ and $W$ is

$$\|\vec{y} - \text{Proj}_W(\vec{y})\| = \sqrt{0+3^2+13^2} = \sqrt{9+169} = \sqrt{178}.$$

- Theorem

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace $W$ of $\mathbb{R}^n$, then

$$\text{proj}_W\,\mathbf{y} = (\mathbf{y}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y}\cdot\mathbf{u}_p)\mathbf{u}_p \quad (4)$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\,]$, then

$$\text{proj}_W\,\mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \quad (5)$$

Proof The first assertion follows from the projection formula, as $\langle \vec{u}_i, \vec{u}_i \rangle = \|\vec{u}_i\|^2 = 1$, $i = 1, \ldots, p$.

For the second formula, notice that

$$U^t\vec{y} = \begin{pmatrix}\vec{u}_1^{\,t}\\\vdots\\\vec{u}_p^{\,t}\end{pmatrix}\vec{y} = \begin{pmatrix}\vec{u}_1^{\,t}\cdot\vec{y}\\\vdots\\\vec{u}_p^{\,t}\cdot\vec{y}\end{pmatrix} = \begin{pmatrix}\langle\vec{u}_1,\vec{y}\rangle\\\vdots\\\langle\vec{u}_p,\vec{y}\rangle\end{pmatrix}.$$

So $\quad U \cdot U^t \vec{y} = [\vec{u}_1 \cdots \vec{u}_p] \begin{pmatrix} \langle \vec{u}_1, \vec{y} \rangle \\ \vdots \\ \langle \vec{u}_p, \vec{y} \rangle \end{pmatrix}$

$$= \langle \vec{u}_1, \vec{y} \rangle \vec{u}_1 + \cdots + \langle \vec{u}_p, \vec{y} \rangle \vec{u}_p .$$

$$= \operatorname{proj}_W(\vec{y})$$

# The Gram-Schmidt process

- With the projection formula, the question is reduced to finding an orthogonal or orthonormal basis of $W$, this is given by the Gram-Schmidt process.

**The Gram–Schmidt Process**

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

$$\mathbf{v}_1 = \mathbf{x}_1$$
$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$
$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$
$$\vdots$$
$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$

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

$$\operatorname{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \operatorname{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \qquad \text{for } 1 \le k \le p \tag{1}$$

**Proof** Let $W_k = \text{Span}\{\vec{x}_1, \cdots, \vec{x}_k\}$, $k = 1, \cdots, p$. Observe that in the formula

$$\vec{v}_k = \vec{x}_k - \underbrace{\left[ \frac{\langle \vec{x}_k, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 + \cdots + \frac{\langle \vec{x}_k, \vec{v}_{k-1} \rangle}{\langle \vec{v}_{k-1}, \vec{v}_{k-1} \rangle} \vec{v}_{k-1} \right]}_{\text{Proj}_{\text{Span}\{\vec{v}_1, \cdots, \vec{v}_{k-1}\}} (\vec{x}_k)}.$$

Hence $\vec{v}_k$ must be orthogonal to $\text{Span}\{\vec{v}_1, \cdots, \vec{v}_{k-1}\}$. In particular, $\vec{v}_1, \cdots, \vec{v}_{k-1}, \vec{v}_k$ form an orthogonal basis of $\text{Span}\{\vec{v}_1, \cdots, \vec{v}_{k-1}, \vec{v}_k\}$.

It remains to show that

(*) $\text{Span}\{\vec{v}_1, \cdots, \vec{v}_k\} = \text{Span}\{\vec{x}_1, \cdots, \vec{x}_k\}$.

Once this is proven, $\vec{v}_1, \cdots, \vec{v}_p$ forms an orthogonal basis of $\text{Span}\{\vec{x}_1, \cdots, \vec{x}_p\} = W$ as claimed.

To show (*), notice that the transformation between $\{\vec{v}_1, \cdots, \vec{v}_k\}$ and $\{\vec{x}_1, \cdots, \vec{x}_k\}$ can be rewritten as:

$$\vec{x}_1 = \vec{v}_1$$

$$\vec{x}_2 = \vec{v}_2 + \frac{\langle \vec{x}_2, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1$$

$$\vec{x}_3 = \vec{v}_3 + \frac{\langle \vec{x}_3, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 + \frac{\langle \vec{x}_3, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\vec{x}_p = \vec{v}_p + \frac{\langle \vec{x}_p, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 + \cdots + \frac{\langle \vec{x}_p, \vec{v}_{p-1} \rangle}{\langle \vec{v}_{p-1}, \vec{v}_{p-1} \rangle} \vec{v}_{p-1}$$

In terms of matrix,

$$[\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_k] = [\vec{v}_1 \ \vec{v}_2 \ \cdots \ \vec{v}_k] \underbrace{\begin{pmatrix} 1 & \frac{\langle \vec{x}_1, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} & \cdots & \frac{\langle \vec{x}_k, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \\ 0 & 1 & \ddots & \vdots \\ 0 & 0 & \ddots & \frac{\langle \vec{x}_k, \vec{v}_{k-1} \rangle}{\langle \vec{v}_{k-1}, \vec{v}_{k-1} \rangle} \\ \vdots & \vdots & & \ddots \\ 0 & 0 & \cdots \cdots & 1 \end{pmatrix}}_{\text{upper triangular with 1 at the diagonal. hence invertible.}}$$

$$\Rightarrow \operatorname{Span}\{\vec{v}_1, \cdots, \vec{v}_k\} = \operatorname{Span}\{\vec{x}_1, \cdots, \vec{x}_k\}.$$

$\boxed{S}$

- Once we get an orthogonal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ of $W$, we can normalize them by setting $\mathbf{u}_i = \mathbf{v}_i / \| \mathbf{v}_i \|$, then $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ form an orthonormal basis of $W$.
- **The QR Factorization**

  If $A$ is an $m \times n$ matrix with linearly independent columns, then $A$ can be factored as $A = QR$, where $Q$ is an $m \times n$ matrix whose columns form an orthonormal basis for Col $A$ and $R$ is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

- In particular, any $n \times n$ invertible matrix $A$ can be factorized as $A = QR$ with $Q$ being an $n \times n$ orthogonal matrix and $R$ being an $n \times n$ invertible upper triangular matrix with positive entries on the diagonal.

Proof This is a reformulation of the Gram-Schmidt process. Let $A = [\vec{x}_1 \cdots \vec{x}_n]$, by assumption $\{\vec{x}_1, \cdots, \vec{x}_n\}$ forms a basis of $\mathrm{Col}(A)$. Applying the Gram-Schmidt process to the basis $\{\vec{x}_1, \cdots, \vec{x}_n\}$, we get an orthogonal basis $\{\vec{v}_1, \cdots, \vec{v}_n\}$ such that

$$(*) \quad [\vec{x}_1\ \vec{x}_2\ \cdots\ \vec{x}_n] = [\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_n] \underbrace{\begin{bmatrix} 1 & \frac{\langle \vec{x}_1, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} & \cdots & \frac{\langle \vec{x}_n, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \\ 0 & 1 & & \vdots \\ 0 & 0 & \ddots & \frac{\langle \vec{x}_n, \vec{v}_{n-1} \rangle}{\langle \vec{v}_{n-1}, \vec{v}_{n-1} \rangle} \\ \vdots & \vdots & & \\ 0 & 0 & \cdots & 1 \end{bmatrix}}_{B}.$$

Normalize the basis $\{\vec{v}_1, \cdots, \vec{v}_n\}$ by setting

$$\vec{u}_i = \frac{\vec{v}_i}{\|\vec{v}_i\|}, \quad i = 1, \cdots, n.$$

we get an orthonormal basis $\{\vec{u}_1, \cdots, \vec{u}_n\}$. The equation (*) can be rewritten as

$$\underbrace{[\vec{x}_1 \ \vec{x}_2 \ \cdots \ \vec{x}_n]}_{A} = \underbrace{[\vec{u}_1 \ \vec{u}_2 \ \cdots \ \vec{u}_n]}_{Q} \underbrace{\begin{bmatrix} \|v_1\| & & & \\ & \|v_2\| & & B \\ & & \ddots & \\ & & & \|v_n\| \end{bmatrix}}_{R}$$

upper triangular with 1's at the diagonal

$\Rightarrow$ get the factorization $A = QR$. $\boxed{5}$

**EXAMPLE 4** Find a QR factorization of $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

Solution: Let $A = [\vec{x}_1 \ \vec{x}_2 \ \vec{x}_3]$, applying Gram-Schmidt process, we get orthogonal sets

$$\vec{v}_1 = \vec{x}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

$$\vec{v}_2 = \vec{x}_2 - \frac{\langle x_2, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} - \frac{3}{4} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

$$\vec{v}_3 = \vec{x}_3 - \frac{\langle \vec{x}_3, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle} \vec{v}_1 - \frac{\langle \vec{x}_3, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle} \vec{v}_2$$

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

$$= \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{pmatrix} - \begin{pmatrix} -\frac{1}{2} \\ 1/6 \\ 1/6 \\ 1/6 \end{pmatrix} = \begin{pmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 0 \\ -2 \\ 1 \\ 1 \end{pmatrix}.$$

Normalise them by setting $\vec{u}_i = \frac{\vec{v}_i}{\|\vec{v}_i\|}$, we get

$$\vec{u}_1 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{u}_2 = \frac{1}{\sqrt{12}} \begin{pmatrix} -3 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{u}_3 = \frac{1}{\sqrt{6}} \begin{pmatrix} 0 \\ -2 \\ 1 \\ 1 \end{pmatrix}.$$

Let $Q = [\vec{u}_1 \ \vec{u}_2 \ \vec{u}_3] = \begin{pmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{pmatrix}$, its column vectors are orthonormal, this is the matrix $Q$ that we are looking for.

To find the matrix $R$ such that $A = QR$. Notice that since the columns of $Q$ are orthonormal, we have

$$Q^t Q = \begin{pmatrix} \vec{u}_1^t \\ \vec{u}_2^t \\ \vec{u}_3^t \end{pmatrix} [\vec{u}_1 \ \vec{u}_2 \ \vec{u}_3] = I_3.$$

$$\Rightarrow Q^t A = Q^t Q R = R$$

$$\Rightarrow \quad R = Q^t A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{-3}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{12}} \\ 0 & \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{pmatrix} \begin{pmatrix} 1 & & \\ 1 & 1 & \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{3}{\sqrt{12}} & \frac{2}{\sqrt{12}} \\ 0 & 0 & \frac{2}{\sqrt{6}} \end{pmatrix}.$$