A set of vectors \( \{u_1, \ldots, u_k\} \) in \( \mathbb{R}^n \) is

i) **orthogonal** if \( u_i^T u_j = 0 \) for all \( i \neq j \)

ii) **normalized** if \( \|u_i\| = 1 \) for all \( i \)

"orthonormal" if both are true.

**Example:** \( S = \text{span} \left( \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix} \right) = \text{span} \left( \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right) \).

**Theorem:** Every subspace has an orthonormal basis.

(we'll see how to construct one!)

In other words, given any \( A = [a_1, \ldots, a_k] \) and \( S = \mathbb{R}(A) \),

\( \text{dim. indep.} \)

there are orthonormal vectors \( \{q_1, \ldots, q_k\} \) such that if \( Q = [q_1, \ldots, q_k] \) then \( \mathbb{R}(A) = \mathbb{R}(Q) \). A matrix like \( Q \) (orthonormal columns)

is called an **orthogonal matrix**. (really should be called an orthonormal matrix, but that's not the convention.)
Orthogonal matrices: \( Q = [q_1, \ldots, q_n] \in \mathbb{R}^{m \times n} \) is orthogonal if

\[
q_i \cdot q_j = \begin{cases} 
0 & \text{if } i \neq j \\
1 & \text{if } i = j
\end{cases}
\]

equivalently, \( Q \) is orthogonal if:

\[
Q^T Q = \begin{bmatrix} q_1^T q_1 & \cdots & q_n^T q_n \\
\vdots & \ddots & \vdots \\
q_n^T q_n & \cdots & q_n^T q_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\
0 & 1 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & 1 \end{bmatrix} = I
\]

Properties

[Remember: \( P, Q \) need not be square!]

i) if \( P, Q \) orthogonal, then \( PQ \) is orthogonal.

\[
\text{proof: } (PQ)^T(PQ) = Q^T P^T P Q = Q^T Q = I
\]

ii) if \( Q \) is orthogonal, then 2-norm is preserved: \( \|Qx\| = \|x\| \).

\[
\text{proof: } \|Qx\|^2 = (Qx)^T(Qx) = x^T Q^T Q x = x^T x = \|x\|^2
\]

More properties:

i) in general, \( QQ^T \neq I \). But if \( Q \) is square and orthogonal,

\( Q^T Q = QQ^T = I \), i.e. \( Q^T \) is also orthogonal, and \( Q^{-1} = Q^T \).

ii) if \( Q \in \mathbb{R}^{m \times n} \) is orthogonal, there exists \( Q_2 \in \mathbb{R}^{m \times (m-n)} \) also orthogonal such that \([Q_1, Q_2] \) is orthogonal (and square).

In fact, \( R(Q_2) = R(Q_1)^{-1} \)
projections: \[ w = \text{proj}_u(v) \] (projection of \( v \) onto \( u \)).

(see picture) - it's a decomposition of \( v \) into \( w + w' \)
where \( w \) is aligned with \( u \) and \( w' \) is orthogonal to \( u \).

\[ v \text{ has length } \|v\|. \text{ Therefore, } \|w\| = \|v\| \cos \theta \]

\[ \Rightarrow \|w\| = \frac{\|u\| \cdot \|v\| \cos \theta}{\|u\|} = \left( \frac{u^T v}{\|u\|^2} \right) \]

and the direction (normalized) of \( w \) should be \( \frac{u}{\|u\|} \).

Therefore, \[ w = \frac{\|w\|}{\|u\|} \cdot \frac{u}{\|u\|} = \frac{u^T v}{\|u\|^2} u \]

so: \[ \text{proj}_u(v) = \frac{u^T v}{\|u\|^2} u \]. We also have \( w' = v - \text{proj}_u(v) \).

Gram-Schmidt orthogonalization given \( \{a_1, a_2, \ldots, a_n\} \),

\[ a_1' = a_1 \]
\[ a_2' = a_2 - \text{proj}_{a_1'}(a_2) \]
\[ a_3' = a_3 - \text{proj}_{a_1'}(a_3) - \text{proj}_{a_2'}(a_3) \]
\[ \vdots \]
\[ a_n' = a_n - \sum_{i=1}^{n-1} \text{proj}_{a_i'}(a_n) \]

then normalize: \( u_i = \frac{a_i'}{\|a_i'\|} \) for all \( i \).

\[ \Rightarrow \{u_1, \ldots, u_n\} \text{ is an orthonormal basis for span } \{a_1, \ldots, a_n\} \]
What if the \( \{a_i, \ldots, a_n\} \) are not linearly independent? In this case, some \( a_i \) is a linear combination of the previous \( \{a_1, \ldots, a_{i-1}\} \), so \( a_i = a_i - \sum_{j=1}^{i-1} \text{proj}_{a_j}(a_i) = 0 \). Simply move on and ignore this \( a_i \). Note result:

\[
\{a_1, \ldots, a_n\} \xrightarrow{\text{Gram-Schmidt}} \{u_1, \ldots, u_r\}
\]

any set of vectors

orthonormal basis for \( \mathbb{R}(A) \),
where \( r = \text{rank}(A) \).

**Note:** \( \{u_1, \ldots, u_r\} \) is not unique! (Could rearrange \( a_i \)'s and get a different result). There are many possible orthonormal bases in general.

**in Matlab:**

- \( \text{orth}(A) \): produces a matrix that is orthogonal and for which the range equals the range of \( A \).
- \( \text{null}(A) \): produces a matrix that is orthogonal whose columns are an orthonormal basis for \( \mathbb{N}(A) \).
Gram-Schmidt examples

\[ A = \begin{bmatrix} -1 & 1 & 2 \\ 1 & -1 & 2 \end{bmatrix} \quad ; \quad a_1 = \begin{pmatrix} -1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 \end{pmatrix}, \quad a_3 = \begin{pmatrix} 2 \end{pmatrix} \]

\[ a'_1 = \begin{pmatrix} -1 \end{pmatrix}, \]
\[ a' = \begin{pmatrix} 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} -1 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix} - \begin{pmatrix} -1 \end{pmatrix} = \begin{pmatrix} 2 \end{pmatrix} \]

\[ a''_3 = \begin{pmatrix} 2 \end{pmatrix} - \frac{0}{2} \begin{pmatrix} 1 \end{pmatrix} = \begin{pmatrix} 2 \end{pmatrix} \]

so \[ a'_1 = \begin{pmatrix} -1 \end{pmatrix}, \quad a'_2 = \begin{pmatrix} 0 \end{pmatrix}, \quad a''_3 = \begin{pmatrix} 2 \end{pmatrix} \]

\[ \Rightarrow u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad u_2 = \begin{pmatrix} \frac{1}{\sqrt{10}} \end{pmatrix} \quad \Rightarrow \quad U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{10}} & \frac{1}{\sqrt{2}} \end{bmatrix} \]

\[ A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 3 & 1 \end{bmatrix} \quad ; \quad a_1 = \begin{pmatrix} 1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 \end{pmatrix} \]

\[ a'_1 = \begin{pmatrix} 1 \end{pmatrix}, \]
\[ a''_2 = \begin{pmatrix} 2 \end{pmatrix} - \frac{1}{3} \begin{pmatrix} 1 \end{pmatrix} = \begin{pmatrix} 2 \end{pmatrix} - \begin{pmatrix} \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{4}{3} \end{pmatrix} \]

\[ \Rightarrow u_1 = \begin{pmatrix} \frac{1}{\sqrt{3}} \end{pmatrix}, \quad u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \end{pmatrix} \quad \Rightarrow \quad U = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \]
if S is any subspace, we can find orthonormal \( U \in \mathbb{R}^{n \times r} \), such that \( R(U_1) = S \). Then, find \( U_2 \) such that \([U_1, U_2]\) is orthonormal and square. [how? one way is to apply G.S. to \([U_1, I]\) ...].

Now take any vector \( x \in \mathbb{R}^n \). Since \( U = [U_1, U_2] \) is orthonormal and square, \( U^T U = U U^T = I \). So:

\[
X = U U^T x
= [U_1, U_2][U_1, U_2]^T x
= U_1 U_1^T x + U_2 U_2^T x.
= \left( u_1 u_1^T x + \cdots + u_r u_r^T x \right) + \left( u_{r+1} u_{r+1}^T x + \cdots + u_n u_n^T x \right)
= \left( \sum_{i=1}^{r} \text{proj}_{u_i} x \right) + \left( \sum_{i=r+1}^{n} \text{proj}_{u_i} x \right)
= \text{proj}_{\text{span}(u_1, \ldots, u_r)} x + \text{proj}_{\text{span}(u_{r+1}, \ldots, u_n)} x.
\]

Every \( x \in \mathbb{R}^n \) can be written as \( x_1 + x_2 \) where \( x_1 \in S \), \( x_2 \in S^\perp \) in a unique way. If \( R(U_1) = S \) and \( R(U_2) = S^\perp \), \([U_1, U_2]\) orthonormal, then \( x_1 = U_1 U_1^T x = \text{proj}_S x \), \( x_2 = U_2 U_2^T x = \text{proj}_{S^\perp} x \).
Solving \( \| b - A x \| \) is the same as finding \( \hat{b} \in \text{R}(A) \) such that \( \| b - \hat{b} \| \) is as small as possible.

\[ b = \hat{b} + \hat{r} \]

where \( \hat{b}^T \hat{r} = 0 \), \( \hat{b} \in \text{R}(A) \), \( \hat{r} \in \text{R}(A)^{\perp} \).

If \( R(U_1) = R(A) \), \( U_1 \) orthogonal,

\[ \begin{aligned} \hat{b} &= U_1 U_1^T b \\ \hat{r} &= U_2 U_2^T b = (I - U_1 U_1^T) b. \end{aligned} \]

Note: \( \hat{b} = \text{proj}_{\text{R}(A)} b \), \( \hat{r} = \text{proj}_{\text{R}(A)^{\perp}} b \). Also, \( \hat{b}^T \hat{r} = b^T U_1 U_1^T U_2 U_2^T b = 0 \).

\( \hat{b} \) and \( \hat{r} \) are always unique! There may be multiple \( x \)'s such that \( \hat{b} = A x \), but there is only one \( \hat{b} \)!

We already saw that LS solutions satisfy \( A^T A \hat{x} = A^T b \).

If columns of \( A \) are independent, \( \hat{x} = (A^T A)^{-1} A^T b \).

Therefore, \( \hat{b} = A \hat{x} = A (A^T A)^{-1} A^T b \).

So:

\[ \begin{aligned} \hat{b} &= \text{proj}_{\text{R}(A)} b = A (A^T A)^{-1} A^T b \\ \hat{r} &= \text{proj}_{\text{R}(A)^{\perp}} b = (I - A (A^T A)^{-1} A^T) b. \end{aligned} \]

Can easily check that \( \hat{b} + \hat{r} = b \) and \( \hat{b}^T \hat{r} = 0 \).
in Matlab: if $A$ has full column rank, "$A \backslash b$" (backslash) is the same as $(A^T A)^{-1} A^T b$.

A check: $A = \text{rand}(10, 5)$;
$U_1 = \text{orth}(A)$;

(should be zero!

$U_1 U_1^T = A \ast \text{inv}(A^T A) \ast A$