Orthogonality plays a key role in linear algebra for data science thanks to its computational properties and its connection to the least-squares problem.

**Definition (Orthogonality):** Vectors \( u \) and \( v \) in \( V \) are orthogonal if their inner product satisfies \( \langle u, v \rangle = 0 \).
Orthogonality has important implications. The following classical result will be useful below.

**Lemma (Pythagoras):** Let \( \mathbf{u}, \mathbf{v} \in V \) be orthogonal. Then \( \| \mathbf{u} + \mathbf{v} \|^2 = \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2 \).

**Proof:** Using \( \| \mathbf{w} \|^2 = \langle \mathbf{w}, \mathbf{w} \rangle \), we get
\[
\| \mathbf{u} + \mathbf{v} \|^2 = \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\
= \langle \mathbf{u}, \mathbf{u} \rangle + 2 \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\
= \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2.
\]
\[\square\]

Here is an application of *Pythagoras*.

**Lemma (Cauchy-Schwarz):** For any \( \mathbf{u}, \mathbf{v} \in V \), \( |\langle \mathbf{u}, \mathbf{v} \rangle| \leq \| \mathbf{u} \| \| \mathbf{v} \| \).

**Proof:** Let \( \mathbf{q} = \frac{\mathbf{v}}{\| \mathbf{v} \|} \) be the unit vector in the direction of \( \mathbf{v} \). We want to show \( |\langle \mathbf{u}, \mathbf{q} \rangle| \leq \| \mathbf{u} \| \). Decompose \( \mathbf{u} \) into its projection onto \( \mathbf{q} \) and what’s left:
\[
\mathbf{u} = \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} + \{ \mathbf{u} - \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} \}.
\]
The two terms on the right-hand side are orthogonal, so *Pythagoras* gives
\[
\| \mathbf{u} \|^2 = |\langle \mathbf{u}, \mathbf{q} \rangle|^2 + \| \mathbf{u} - \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} \|^2 \geq |\langle \mathbf{u}, \mathbf{q} \rangle|^2 = \langle \mathbf{u}, \mathbf{q} \rangle^2.
\]
Taking a square root gives the claim.\[\square\]

### 2.1 Basis expansion

To begin to see the power of orthogonality, consider the following. A list of vectors \( \{ \mathbf{u}_1, \ldots, \mathbf{u}_m \} \) is orthonormal if the \( \mathbf{u}_i \)'s are pairwise orthogonal and each has norm \( 1 \), that is, for all \( i \) and all \( j \neq i \), \( \langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0 \) and \( \| \mathbf{u}_i \| = 1 \).

**Lemma (Properties of Orthonormal Lists):** Let \( \{ \mathbf{u}_1, \ldots, \mathbf{u}_m \} \) be an orthonormal list of vectors. Then:

1. \( \| \sum_{j=1}^m \alpha_j \mathbf{u}_j \|^2 = \sum_{j=1}^m \alpha_j^2 \) for any \( \alpha_j \in \mathbb{R} \), \( j \in [m] \), and
2. \( \{ \mathbf{u}_1, \ldots, \mathbf{u}_m \} \) are linearly independent.
Proof: For 1, using that \(\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle\) and \(\beta \mathbf{x}_1 + \mathbf{x}_2, \mathbf{x}_3 = \beta \langle \mathbf{x}_1, \mathbf{x}_3 \rangle + \langle \mathbf{x}_2, \mathbf{x}_3 \rangle\) (which follow directly from the definition of the inner product), we have

\[
\left\| \sum_{j=1}^{m} \alpha_j \mathbf{u}_j \right\|^2 = \left\langle \sum_{i=1}^{m} \alpha_i \mathbf{u}_i, \sum_{j=1}^{m} \alpha_j \mathbf{u}_j \right\rangle
\]

\[
= \sum_{i=1}^{m} \alpha_i \left\langle \mathbf{u}_i, \sum_{j=1}^{m} \alpha_j \mathbf{u}_j \right\rangle
\]

\[
= \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j \langle \mathbf{u}_i, \mathbf{u}_j \rangle
\]

\[
= \sum_{i=1}^{m} \alpha_i^2
\]

where we used orthonormality in the rightmost equation, that is, \(\langle \mathbf{u}_i, \mathbf{u}_j \rangle\) is 1 if \(i = j\) and 0 otherwise.

For 2, suppose \(\sum_{i=1}^{m} \beta_i \mathbf{u}_i = \mathbf{0}\), then we must have by 1 that \(\sum_{i=1}^{m} \beta_i^2 = 0\). That implies \(\beta_i = 0\) for all \(i\). Hence the \(\mathbf{u}_i\)'s are linearly independent. 

Given a basis \(\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}\) of \(\mathcal{U}\), we know that: for any \(\mathbf{w} \in \mathcal{U}\), \(\mathbf{w} = \sum_{i=1}^{m} \alpha_i \mathbf{u}_i\) for some \(\alpha_i\)'s. It is not immediately obvious in general how to find the \(\alpha_i\)'s. In the orthonormal case, however, it is straightforward.

**Theorem (Orthonormal Expansion):** Let \(\mathbf{q}_1, \ldots, \mathbf{q}_m\) be an orthonormal basis of \(\mathcal{U}\) and let \(\mathbf{u} \in \mathcal{U}\). Then

\[
\mathbf{u} = \sum_{j=1}^{m} \langle \mathbf{u}, \mathbf{q}_j \rangle \mathbf{q}_j.
\]

*Proof:* Because \(\mathbf{u} \in \mathcal{U}\), \(\mathbf{u} = \sum_{i=1}^{m} \alpha_i \mathbf{q}_i\) for some \(\alpha_i\). Take the inner product with \(\mathbf{q}_j\) and use orthonormality:

\[
\langle \mathbf{u}, \mathbf{q}_j \rangle = \left\langle \sum_{i=1}^{m} \alpha_i \mathbf{q}_i, \mathbf{q}_j \right\rangle = \sum_{i=1}^{m} \alpha_i \langle \mathbf{q}_i, \mathbf{q}_j \rangle = \alpha_j.
\]

So we’ve shown that working with orthonormal bases is desirable. What if we don’t have one? We review the Gram-Schmidt algorithm (https://en.wikipedia.org/wiki/Gram–Schmidt_process) below, which will imply that every linear subspace has an orthonormal basis.

But, first, we define the orthogonal projection.
2.2 Orthogonal projection

Let's consider the following problem. We have a linear subspace $\mathcal{U} \subseteq \mathcal{V}$ and a vector $\mathbf{v} \notin \mathcal{U}$. We want to find the vector $\mathbf{v}^*$ in $\mathcal{U}$ that is closest to $\mathbf{v}$ in 2-norm, that is, we want to solve

$$\min_{\mathbf{v} \in \mathcal{U}} \| \mathbf{v} - \mathbf{v}^* \|.$$

**Example:** Consider the two-dimensional case with a one-dimensional subspace, say $\mathcal{U} = \text{span}(\mathbf{u}_1)$ with $\|\mathbf{u}_1\| = 1$. The geometrical intuition is in the following figure. The solution $\mathbf{v}^*$ has the property that the difference $\mathbf{v} - \mathbf{v}^*$ makes a right angle with $\mathbf{u}_1$, that is, it is orthogonal to it.

![Diagram](https://commons.wikimedia.org/wiki/File:Linalg_projection_4.png)

Letting $\mathbf{v}^* = \alpha^* \mathbf{u}_1$, the geometrical condition above translates into

$$0 = \langle \mathbf{u}_1, \mathbf{v} - \mathbf{v}^* \rangle = \langle \mathbf{u}_1, \mathbf{v} - \mathbf{v}^* \rangle = \langle \mathbf{u}_1, \mathbf{v} \rangle - \alpha^* \langle \mathbf{u}_1, \mathbf{u}_1 \rangle = \langle \mathbf{u}_1, \mathbf{v} \rangle - \alpha^*$$

so

$$\mathbf{v}^* = \langle \mathbf{u}_1, \mathbf{v} \rangle \mathbf{u}_1.$$

By Pythagoras, we then have for any $\alpha \in \mathbb{R}$

$$\| \mathbf{v} - \alpha \mathbf{u}_1 \|^2 = \| \mathbf{v} - \mathbf{v}^* + \mathbf{v}^* - \alpha \mathbf{u}_1 \|^2$$

$$= \| \mathbf{v} - \mathbf{v}^* + (\alpha^* - \alpha) \mathbf{u}_1 \|^2$$

$$= \| \mathbf{v} - \mathbf{v}^* \|^2 + \| (\alpha^* - \alpha) \mathbf{u}_1 \|^2$$

$$\geq \| \mathbf{v} - \mathbf{v}^* \|^2.$$

That confirms the optimality of $\mathbf{v}^*$. The argument in this example carries through in higher dimension, as we show next. \(\triangleleft\)
Definition (Orthogonal Projection on an Orthonormal List): Let \( q_1, \ldots, q_m \) be an orthonormal list. The orthogonal projection of \( v \in V \) on \( \{q_i\}_{i=1}^m \) is defined as

\[
(P_{\{q_i\}_{i=1}^m}) v = \sum_{j=1}^m \langle v, q_j \rangle q_j.
\]

Definition and Theorem (Orthogonal Projection): Let \( U \subseteq V \) be a linear subspace with orthonormal basis \( q_1, \ldots, q_m \) and let \( v \in V \). Then \( (P_{\{q_i\}_{i=1}^m}) v \in U \) and, for any \( u \in U \),

\[
\langle v - (P_{\{q_i\}_{i=1}^m}) v, u \rangle = 0
\]

and

\[
\| v - (P_{\{q_i\}_{i=1}^m}) v \| \leq \| v - u \|.
\]

Furthermore, if \( u \in U \) and the inequality above is an equality, then \( u = (P_{\{q_i\}_{i=1}^m}) v \). Hence, for any orthonormal basis \( q'_1, \ldots, q'_m \) of \( U \), it holds that

\[
(P_U) v = (P_{\{q_i\}_{i=1}^m}) v = (P_{\{q'_i\}_{i=1}^m}) v,
\]

where the first equality is a definition. We refer to \( P_U \) as the orthogonal projection of \( v \) on \( U \).
Proof: By definition, it is immediate that $(P_{\{q_i\}_{i=1}}^m) \mathbf{v} \in \text{span}(\{q_i\}_{i=1}^m) = \mathcal{U}$. We first prove (*). We can write any $\mathbf{u} \in \mathcal{U}$ as $\sum_{j=1}^m \alpha_j \mathbf{q}_j$ for some $\alpha_j$'s. Then

$$
\left\langle \mathbf{v} - \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle \mathbf{q}_j, \sum_{j=1}^m \alpha_j' \mathbf{q}_j \right\rangle = \sum_{j=1}^m \langle \mathbf{v}, \mathbf{q}_j \rangle \alpha_j' - \sum_{j=1}^m \alpha_j' \langle \mathbf{v}, \mathbf{q}_j \rangle = 0
$$

where we used the orthonormality of the $\mathbf{q}_j$'s in the leftmost equality.

To prove (**), note that for any $\mathbf{u} \in \mathcal{U}$ the vector $\mathbf{u}' = (P_{\{q_i\}_{i=1}}^m) \mathbf{v} - \mathbf{u}$ is also in $\mathcal{U}$. Hence by (*) and Pythagoras,

$$
\| \mathbf{v} - \mathbf{u} \|^2 = \| \mathbf{v} - (P_{\{q_i\}_{i=1}}^m) \mathbf{v} + (P_{\{q_i\}_{i=1}}^m) \mathbf{v} - \mathbf{u} \|^2 \\
= \| \mathbf{v} - (P_{\{q_i\}_{i=1}}^m) \mathbf{v} \|^2 + \|(P_{\{q_i\}_{i=1}}^m) \mathbf{v} - \mathbf{u} \|^2 \\
\geq \| \mathbf{v} - (P_{\{q_i\}_{i=1}}^m) \mathbf{v} \|^2.
$$

Furthermore, equality holds only if $\|(P_{\{q_i\}_{i=1}}^m) \mathbf{v} - \mathbf{u} \|^2 = 0$ which holds only if $\mathbf{u} = (P_{\{q_i\}_{i=1}}^m) \mathbf{v}$ by the point-separating property of the 2-norm. The rightmost equality in (*** ) follows from swapping $\{q_1'\}_{i=1}^m$ for $\{q_i\}_{i=1}^m$.

The Orthogonal Projection Theorem implies that any $\mathbf{v} \in \mathcal{V}$ can be decomposed into its orthonormal projection onto $\mathcal{U}$ and a vector orthogonal to it.
**Definition (Orthogonal Complement):** Let $\mathcal{U} \subseteq \mathcal{V}$ be a linear subspace. The orthogonal complement of $\mathcal{U}$, denoted $\mathcal{U}^\perp$, is defined as

$$\mathcal{U}^\perp = \{ w \in \mathcal{V} : \langle w, u \rangle = 0, \forall u \in \mathcal{U} \}.$$ 

\[\triangleleft\]

*Exercise:* Establish that $\mathcal{U}^\perp$ is a linear subspace. $\triangleleft$

**Lemma (Orthogonal Decomposition):** Let $\mathcal{U} \subseteq \mathcal{V}$ be a linear subspace and let $v \in \mathcal{V}$. Then $v$ can be decomposed as $(v - P_{\mathcal{U}}v) + P_{\mathcal{U}}v$ where $(v - P_{\mathcal{U}}v) \in \mathcal{U}^\perp$ and $P_{\mathcal{U}}v \in \mathcal{U}$.

*Proof:* Immediate consequence of the previous theorem. $\square$

The map $P_{\mathcal{U}}$ is linear, that is, $P_{\mathcal{U}}(\alpha x + y) = \alpha P_{\mathcal{U}}x + P_{\mathcal{U}}y$ for all $\alpha \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$. Indeed,

$$P_{\mathcal{U}}(\alpha x + y) = \sum_{j=1}^{m} \langle \alpha x + y, q_j \rangle q_j = \sum_{j=1}^{m} \{ \alpha \langle x, q_j \rangle + \langle y, q_j \rangle \} q_j = \alpha P_{\mathcal{U}}x + P_{\mathcal{U}}y.$$ 

Therefore it can be encoded as an $n \times n$ matrix $P$. Let

$$Q = \begin{pmatrix} q_1 & \ldots & q_m \end{pmatrix}$$

and note that computing

$$Q^Tv = \begin{pmatrix} \langle v, q_1 \rangle \\ \vdots \\ \langle v, q_m \rangle \end{pmatrix}$$

lists the coefficients in the expansion of $P_{\mathcal{U}}v$ over the basis $q_1, \ldots, q_m$.

Hence we see that

$$P = QQ^T.$$ 

This is not to be confused with

$$Q^TQ = \begin{pmatrix} \langle q_1, q_1 \rangle & \ldots & \langle q_1, q_m \rangle \\ \langle q_2, q_1 \rangle & \ldots & \langle q_2, q_m \rangle \\ \vdots & \ddots & \vdots \\ \langle q_m, q_1 \rangle & \ldots & \langle q_m, q_m \rangle \end{pmatrix} = I_{m \times m}$$

where $I_{m \times m}$ denotes the $m \times m$ identity matrix.
**Exercise:** Let \( \mathcal{U} \subseteq V \) be a linear subspace and let \( \mathbf{v} \in \mathcal{U} \). Show that \( \mathbf{v} = \mathbf{v} \). \( \triangleright \)

### 2.3 Gram-Schmidt

We have some business left over: constructing orthonormal bases. Let \( \mathbf{a}_1, \ldots, \mathbf{a}_m \) be linearly independent. We use the Gram-Schmidt algorithm to obtain an orthonormal basis of \( \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_m) \). The process takes advantage of the properties of the orthogonal projection derived above. In essence we add the vectors \( \mathbf{a}_i \) one by one, but only after taking out their orthogonal projection on the previously included vectors. The outcome spans the same subspace and the Orthogonal Decomposition Lemma ensures orthogonality.

**Theorem (Gram-Schmidt):** Let \( \mathbf{a}_1, \ldots, \mathbf{a}_m \) be linearly independent. Then there exists an orthonormal basis \( \mathbf{q}_1, \ldots, \mathbf{q}_m \) of \( \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_m) \).

**Proof idea:** Suppose first that \( m = 1 \). In that case, all that needs to be done is to divide \( \mathbf{a}_1 \) by its norm to obtain a unit vector whose span is the same as \( \mathbf{a}_1 \), that is, we set \( \mathbf{q}_1 = \frac{\mathbf{a}_1}{\| \mathbf{a}_1 \|} \).

Suppose now that \( m = 2 \). We first let \( \mathbf{q}_1 = \frac{\mathbf{a}_1}{\| \mathbf{a}_1 \|} \) as in the previous case. Then we subtract from \( \mathbf{a}_2 \) its projection on \( \mathbf{q}_1 \), that is, we set \( \mathbf{v}_2 = \mathbf{a}_2 - \langle \mathbf{q}_1, \mathbf{a}_2 \rangle \mathbf{q}_1 \). By the Orthogonal Projection Theorem, \( \mathbf{v}_2 \) is orthogonal to \( \mathbf{q}_1 \). Moreover, because \( \mathbf{a}_2 \) is a linear combination of \( \mathbf{q}_1 \) and \( \mathbf{v}_2 \), we have \( \text{span}(\mathbf{q}_1, \mathbf{v}_2) = \text{span}(\mathbf{a}_1, \mathbf{a}_2) \). It remains to divide by the norm of the resulting vector: \( \mathbf{q}_2 = \frac{\mathbf{v}_2}{\| \mathbf{v}_2 \|} \).

For general \( m \), we proceed similarly but project onto the subspace spanned by the previously added vectors at each step.
Proof: The inductive step is the following. Assume that we have constructed orthonormal vectors \( \mathbf{q}_1, \ldots, \mathbf{q}_{j-1} \) such that

\[
U_{j-1} := \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}) = \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}).
\]

By the Properties of Orthonormal Lists, \( \{\mathbf{q}\}_{i=1}^{j-1} \) forms an orthonormal basis for \( U_{j-1} \), so we can compute the orthogonal projection of \( \mathbf{a}_j \) on \( U_{j-1} \) as

\[
P_{U_{j-1}} \mathbf{a}_j = \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i,
\]

where we defined \( r_{ij} = \langle \mathbf{q}_i, \mathbf{a}_j \rangle \). And we set

\[
\mathbf{v}_j = \mathbf{a}_j - P_{U_{j-1}} \mathbf{a}_j \quad \text{and} \quad \mathbf{q}_j = \frac{\mathbf{v}_j}{\|\mathbf{v}_j\|}.
\]

Here we used that \( \|\mathbf{v}_j\| > 0 \): indeed otherwise \( \mathbf{a}_j \) would be equal to its projection \( P_{U_{j-1}} \mathbf{a}_j \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}) \) which would contradict linear independence of the \( \mathbf{a}_k \)'s.
By the Orthogonal Projection Theorem, \( \mathbf{q}_j \) is orthogonal to \( \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_{j-1}) \) and, unrolling the calculations above, \( \mathbf{a}_j \) can be re-written as the following linear combination of \( \mathbf{q}_1, \ldots, \mathbf{q}_j \):

\[
\mathbf{a}_j = P_{U_{j-1}} \mathbf{a}_j + \mathbf{v}_j \\
= P_{U_{j-1}} \mathbf{a}_j + \| \mathbf{v}_j \| \mathbf{q}_j \\
= P_{U_{j-1}} \mathbf{a}_j + \| \mathbf{a}_j - P_{U_{j-1}} \mathbf{a}_j \| \mathbf{q}_j \\
= \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i + \left\| \mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i \right\| \mathbf{q}_j \\
= \sum_{i=1}^{j} r_{ij} \mathbf{q}_i + r_{jj} \mathbf{q}_j,
\]

where we defined \( r_{jj} = \left\| \mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i \right\| = \| \mathbf{v}_j \| \).

Hence \( \mathbf{q}_1, \ldots, \mathbf{q}_j \) forms an orthonormal list with \( \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_j) \subseteq \text{span}(\mathbf{q}_1, \ldots, \mathbf{q}_j) \). The opposite inclusion holds by construction. Moreover, because \( \mathbf{q}_1, \ldots, \mathbf{q}_j \) are orthonormal, they are linearly independent by the Properties of Orthonormal Lists so must form a basis of their span. So induction goes through. \( \square \)

**NUMERICAL CORNER** We implement the Gram-Schmidt algorithm in Julia. For reasons that will become clear in a future notebook, we output the \( \mathbf{q}_j \)'s and \( r_{ij} \)'s, each in matrix form.

```
In [1]:  # Julia version: 1.5.1
using LinearAlgebra
```
In [2]: function mmids_gramschmidt(A)
    n,m = size(A)
    Q = zeros(Float64,n,m)
    R = zeros(Float64,m,m)
    for j = 1:m
        v = A[:,j]
        for i = 1:j-1
            R[i,j] = dot(Q[i,:],A[:,j])
            v -= R[i,j]*Q[i,:]
        end
        R[j,j] = norm(v)
        Q[:,j] = v/R[j,j]
    end
    return Q,R
end

Out[2]: mmids_gramschmidt (generic function with 1 method)

Let's try a simple example.

In [3]: w1, w2 = [1,0,1], [0,1,1]
A = hcat(w1,w2)

Out[3]: 3×2 Array{Int64,2}:
    1  0
    0  1
    1  1

In [4]: Q,R = mmids_gramschmidt(A);

In [5]: Q
Out[5]: 3×2 Array{Float64,2}:
    0.707107  -0.408248
    0.0        0.816497
    0.707107   0.408248

In [6]: R
Out[6]: 2×2 Array{Float64,2}:
    1.41421   0.707107
    0.0      1.22474

Exercise: Let $B \in \mathbb{R}^{n \times m}$ be a matrix. Show that there exist orthonormal bases of $\text{col}(B)$ and $\text{null}(B)$. 

Exercise: Let $\mathcal{W}$ be a linear subspace of $\mathbb{R}^d$ and let $w_1, \ldots, w_k$ be an orthonormal basis of $\mathcal{W}$. Show that there exists an orthonormal basis of $\mathbb{R}^d$ that includes the $w_j$'s.
Lemma (Nullity): Let $B \in \mathbb{R}^{n \times m}$ be a matrix of rank $k$. Then $\text{dim(null}(B)) = n - k$. Put differently, $\text{dim}(\text{col}(B)) + \text{dim}(\text{null}(B)) = n$.

For a proof, see Wikipedia (https://en.wikipedia.org/wiki/Rank%E2%80%93nullity_theorem#First_proof).