TOPIC 1

Cholesky, QR and Householder

2 A key concept: orthogonality

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 $\mathbf{u}$ and $\mathbf{v}$ in $V$ are orthogonal if their inner product satisfies $\langle \mathbf{u}, \mathbf{v} \rangle = 0$.

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
\[
\| \mathbf{u} \|^2 = \| \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} \|^2 + \| \mathbf{u} - \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} \|^2 \geq \| \langle \mathbf{u}, \mathbf{q} \rangle \mathbf{q} \|^2 = \langle \mathbf{u}, \mathbf{q} \rangle^2 .
\]
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 \( \mathbf{u} \)'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 \( \|x\|^2 = \langle x, x \rangle \) and \( \langle \beta x_1 + x_2, x_3 \rangle = \beta \langle x_1, x_3 \rangle + \langle x_2, x_3 \rangle \) (which follow directly from the definition of the inner product), we have

\[
\left\| \sum_{j=1}^m \alpha_j u_j \right\|^2 = \left\langle \sum_{j=1}^m \alpha_j u_j, \sum_{j=1}^m \alpha_j u_j \right\rangle \\
= \sum_{i=1}^m \alpha_i \left\langle u_i, \sum_{j=1}^m \alpha_j u_j \right\rangle \\
= \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j \left\langle u_i, u_j \right\rangle \\
= \sum_{i=1}^m \alpha_i^2
\]

where we used orthonormality in the rightmost equation, that is, \( \left\langle u_i, u_j \right\rangle \) is 1 if \( i = j \) and 0 otherwise.

For 2, suppose \( \sum_{i=1}^m \beta_i u_i = 0 \), 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 \( u_i \)'s are linearly independent. \( \square \)

Given a basis \( \{u_1, \ldots, u_m\} \) of \( \mathcal{U} \), we know that: for any \( w \in \mathcal{U}, w = \sum_{i=1}^m \alpha_i u_i \) 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 Basis Expansion):** Let \( q_1, \ldots, q_m \) be an orthonormal basis of \( \mathcal{U} \) and let \( u \in \mathcal{U} \). Then

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

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

\[
\langle u, q_j \rangle = \left\langle \sum_{i=1}^m \alpha_i q_i, q_j \right\rangle = \sum_{i=1}^m \alpha_i \langle q_i, q_j \rangle = \alpha_j.
\]

\( \square \)

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 $U \subseteq V$ and a vector $v \notin U$. We want to find the vector $v^*$ in $U$ that is closest to $v$ in 2-norm, that is, we want to solve

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

**Example:** Consider the two-dimensional case with a one-dimensional subspace, say $U = \text{span}(u_1)$ with $\|u_1\| = 1$. The geometrical intuition is in the following figure. The solution $v^*$ has the property that the difference $v - v^*$ makes a right angle with $u_1$, that is, it is orthogonal to it.

Letting $v^* = \alpha^* u_1$, the geometrical condition above translates into

$$0 = \langle u_1, v - v^* \rangle = \langle u_1, v - \alpha^* u_1 \rangle = \langle u_1, v \rangle - \alpha^* \langle u_1, u_1 \rangle = \langle u_1, v \rangle - \alpha^*$$

so

$$v^* = \langle u_1, v \rangle u_1.$$ 

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

$$\|v - \alpha u_1\|^2 = \|v^* + v^* - \alpha u_1\|^2$$

$$= \|v^* + (\alpha^* - \alpha) u_1\|^2$$

$$= \|v - v^*\|^2 + \|\alpha^* - \alpha\| u_1\|^2$$

$$\geq \|v - v^*\|^2.$$ 

That confirms the optimality of $v^*$. The argument in this example carries through in higher dimension, as we show next. $\triangleright$
**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 is Closest in Subspace):** 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 \( u \in U \),
\[
(*) \quad \langle v - (P_{\{q_i\}_{i=1}^m}) v, u \rangle = 0
\]
and
\[
(**) \quad \|v - (P_{\{q_i\}_{i=1}^m}) v\| \leq \|v - u\|.
\]
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
\[
(***) \quad P_U v = (P_{\{q_i\}_{i=1}^m}) v = (P_{\{q'_i\}_{i=1}^m}) v,
\]
where the first equality is a definition. We refer to \( P_U v \) as the orthogonal projection of \( v \) on \( U \).
Proof: By definition, it is immediate that \( (P_{\{q_i\}_{i=1}^m}) \mathbf{v} \in \text{span}(\{q_i\}_{i=1}^m) = U' \). We first prove (**). We can write any \( \mathbf{u} \in U' \) as \( \sum_{j=1}^m \alpha_j q_j \) for some \( \alpha_j \)'s. Then

\[
\left\langle \mathbf{v} - \sum_{j=1}^m \langle \mathbf{v}, q_j \rangle q_j, \sum_{j=1}^m \alpha'_j q_j \right\rangle = \sum_{j=1}^m \langle \mathbf{v}, q_j \rangle \alpha'_j - \sum_{j=1}^m \alpha'_j \langle \mathbf{v}, q_j \rangle = 0
\]

where we used the orthonormality of the \( q_j \)'s in the leftmost equality.

To prove (**), note that for any \( \mathbf{u} \in U' \) the vector \( \mathbf{u}' = (P_{\{q_i\}_{i=1}^m}) \mathbf{v} - \mathbf{u} \) is also in \( U' \). 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 \( \{q'_i\}_{i=1}^m \) for \( \{q_i\}_{i=1}^m \).
The **Orthogonal Projection is Closest in Subspace Theorem** implies that any \( \mathbf{v} \in V \) can be decomposed into its orthogonal projection onto \( U \) and a vector orthogonal to it.

**Definition (Orthogonal complement):** Let \( U \subseteq V \) be a linear subspace. The orthogonal complement of \( U \), denoted \( U^\perp \), is defined as
\[
U^\perp = \{ \mathbf{w} \in V : \langle \mathbf{w}, \mathbf{u} \rangle = 0, \forall \mathbf{u} \in U \}.
\]

**Exercise:** Establish that \( U^\perp \) is a linear subspace. \( \triangleright \)

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

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

The map \( P_U \) is linear, that is, \( P_U(\alpha \mathbf{x} + \mathbf{y}) = \alpha P_U \mathbf{x} + P_U \mathbf{y} \) for all \( \alpha \in \mathbb{R} \) and \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \). Indeed,
\[
P_U(\alpha \mathbf{x} + \mathbf{y}) = \sum_{j=1}^{m} \langle \alpha \mathbf{x} + \mathbf{y}, \mathbf{q}_j \rangle \mathbf{q}_j = \sum_{j=1}^{m} \{ \alpha \langle \mathbf{x}, \mathbf{q}_j \rangle + \langle \mathbf{y}, \mathbf{q}_j \rangle \} \mathbf{q}_j = \alpha P_U \mathbf{x} + P_U \mathbf{y}.
\]

Therefore it can be encoded as an \( n \times n \) matrix \( P \). Let
\[
Q = \begin{pmatrix}
\mathbf{q}_1 & \cdots & \mathbf{q}_m
\end{pmatrix}
\]
and note that computing
\[
Q^T \mathbf{v} = \begin{pmatrix}
\langle \mathbf{v}, \mathbf{q}_1 \rangle \\
\vdots \\
\langle \mathbf{v}, \mathbf{q}_m \rangle
\end{pmatrix}
\]
lists the coefficients in the expansion of \( P_U \mathbf{v} \) over the basis \( \mathbf{q}_1, \ldots, \mathbf{q}_m \).
Hence we see that

\[ P = QQ^T. \]

This is not to be confused with

\[
Q^T Q = \begin{pmatrix}
\langle q_1, q_1 \rangle & \cdots & \langle q_1, q_m \rangle \\
\langle q_2, q_1 \rangle & \cdots & \langle q_2, q_m \rangle \\
\vdots & \ddots & \vdots \\
\langle q_m, q_1 \rangle & \cdots & \langle q_m, q_m \rangle
\end{pmatrix} = I_{m \times m}
\]

where \( I_{m \times m} \) denotes the \( m \times m \) identity matrix.

**Exercise:** Let \( U' \subseteq V \) be a linear subspace and let \( v \in U' \). Show that \( P_{U'} v = v \). □

### 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 exist 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 Decomposition Lemma**, \( \mathbf{v}_2 \) 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 \( \text{span}(\mathbf{q}_1, \mathbf{b}_2) = \text{span}(\mathbf{a}_1, \mathbf{a}_2) \). 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 Lemma, \( \{ \mathbf{q}_i \}_{i=1}^{j-1} \) forms an orthonormal basis for \( U_{j-1} \), so we can compute the orthogonal projection of \( \mathbf{a}_j \) on \( U_{j-1} \) as

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

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

And we set

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

Here we used that \( \| \mathbf{v}_j \| > 0 \): indeed otherwise \( \mathbf{a}_j \) would be equal to its projection \( \mathcal{P}_{U_{j-1}} \mathbf{a}_j \in \text{span}(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}) \) which would contradict linear independence of the \( \mathbf{a}_k \)'s. By the Orthogonal Decomposition Lemma, \( \mathbf{q}_j \) 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 \( \mathbf{q}_1, \ldots, \mathbf{q}_j \)

\[
\mathbf{a}_j = \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i + \left( \mathbf{a}_j - \sum_{i=1}^{j-1} r_{ij} \mathbf{q}_i \right) \mathbf{q}_j.
\]

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 Lemma 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 [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
```
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 $w_i$'s. ⬤
Lemma (Dimension of the Null Space): Let $B \in \mathbb{R}^{n \times m}$ be a matrix of rank $k$. Then $\dim(\text{null}(B)) = n - k$. Put differently,

$$\dim(\text{col}(B)) + \dim(\text{null}(B)) = n.$$  

Proof idea: Take an orthonormal basis of the column space of $B$ and complete it into an orthonormal basis of $\mathbb{R}^n$. The added vectors are in the null space and a dimension argument gives the claim.

Proof: By the previous exercise, any orthonormal basis $w_1, \ldots, w_b$ of $\text{col}(B)$, with $b = k$ by assumption, can be completed into an orthonormal basis of $\mathbb{R}^n$

$$\{w_1, \ldots, w_b, \tilde{w}_{b+1}, \ldots, \tilde{w}_n\}.$$  

By definition the vectors $\tilde{w}_{b+1}, \ldots, \tilde{w}_n$ are orthogonal to the columns of $B$, hence they are in $\text{null}(B)$. Because they are orthonormal, they are linearly independent. Hence, the dimension of $\text{null}(B)$ is at least $n - b = n - k$. ☐