Dot product and linear least squares problems

Dot Product

For vectors \( u, v \in \mathbb{R}^n \) we define the **dot product**

\[
\mathbf{u} \cdot \mathbf{v} = u_1v_1 + \cdots + u_nv_n
\]

Note that we can also write this as \( \mathbf{u}^\top \mathbf{v} = [u_1, \ldots, u_n] \left[ \begin{array}{c} v_1 \\ \vdots \\ v_n \end{array} \right] = u_1v_1 + \cdots + u_nv_n. \)

The dot product \( \mathbf{u} \cdot \mathbf{u} = u_1^2 + \cdots + u_n^2 \) gives the square of the Euclidean length of the vector, a.k.a. **norm of the vector**:

\[
\|\mathbf{u}\| = (\mathbf{u} \cdot \mathbf{u})^{1/2} = (u_1^2 + \cdots + u_n^2)^{1/2}
\]

**Theorem 1 (Cauchy-Schwarz inequality).** For \( a, b \in \mathbb{R}^n \) we have

\[
|a \cdot b| \leq \|a\| \|b\|
\]

**Proof.** Step 1 for vectors with \( \|a\| = 1 \) and \( \|b\| = 1 \): Then

\[
0 \leq (b - a) \cdot (b - a) = b \cdot b - 2a \cdot b + a \cdot a = 2a \cdot b \leq a \cdot a + b \cdot b = 1 + 1
\]

Hence \( a \cdot b \leq 1 \). Using \( (b + a) \) instead of \( (b - a) \) gives \(-2a \cdot b \leq 2\), i.e., \( a \cdot b \geq -1 \).

Step 2 for general vectors \( a, b \): If \( a = \mathbf{0} \) or \( b = \mathbf{0} \) we see that (1) obviously holds. If both vectors are different from \( \mathbf{0} \) let \( u := a/\|a\| \) and \( v := b/\|b\| \), then \( u \cdot v = \frac{a \cdot b}{\|a\| \|b\|} \). Since \( \|u\| = 1 \) and \( \|v\| = 1 \) we get from step 1 that \( |u \cdot v| \leq 1 \).

Consider a triangle with the three points \( \mathbf{0}, a, b \). Then the vector from \( a \) to \( b \) is given by \( c = b - a \), and the lengths of the three sides of the triangle are \( \|a\|, \|b\|, \|c\| \).

Let \( \theta \) denote the **angle between the vectors** \( a \) and \( b \). Then the **law of cosines** tells us that

\[
\|c\|^2 = \|a\|^2 + \|b\|^2 - 2\|a\|\|b\| \cos \theta
\]

Multiplying out \( \|c\|^2 = (b - a) \cdot (b - a) \) gives

\[
\|c\|^2 = \|a\|^2 + \|b\|^2 - 2a \cdot b
\]

By comparing the last two equations we obtain

\[
a \cdot b = \|a\| \|b\| \cos \theta
\]

If \( a \) and \( b \) are different from \( \mathbf{0} \) we have

\[
\cos \theta = \frac{a \cdot b}{\|a\| \|b\|}
\]

This tells us **how to compute the angle** \( \theta \) **between two vectors**:

\[
q := \frac{a \cdot b}{\|a\| \|b\|}, \quad \theta := \cos^{-1} q
\]

Because of (1) we have \(-1 \leq q \leq 1\), hence the inverse cosine function gives \( 0 \leq \theta \leq \pi \):

\[
q = 1 \iff \theta = 0, \quad \text{i.e., } a, b \text{ point in the same direction}
\]

\[
q = 0 \iff \theta = \frac{\pi}{2}, \quad \text{i.e., } a, b \text{ are orthogonal}
\]

\[
q = -1 \iff \theta = \pi, \quad \text{i.e. } a, b \text{ point in opposite directions}
\]
Example 1. Find the angle between the vectors \( a = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \) and \( b = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} \): We get

\[
q := \frac{a \cdot b}{\|a\| \|b\|} = \frac{1}{\sqrt{2} \sqrt{2}} = \frac{1}{2}, \quad \theta := \cos^{-1} q = \cos^{-1} \frac{1}{2} = \frac{\pi}{3}
\]

We say vectors \( a, b \in \mathbb{R}^n \) are orthogonal or \( a \perp b \)

\[ a \perp b \iff a \cdot b = 0 \]

We say a vector \( a \in \mathbb{R}^n \) is orthogonal on a subspace \( V \) of \( \mathbb{R}^n \) or \( a \perp V \)

\[ a \perp V \iff a \cdot v = 0 \quad \text{for all } v \in V \]

**Orthogonal projection onto a line**

Consider a 1-dimensional subspace \( V = \text{span} \{ v \} \) of \( \mathbb{R}^n \) given by a vector \( v \in \mathbb{R}^n \). This is a line through the origin.

For a given point \( b \in \mathbb{R}^n \) we want to find the point \( u \in V \) which is closest to the point \( b \):

\[
\text{Find } u \in V \text{ such that } \|u - b\| \text{ is minimal}
\]

The point \( u \) must have the following properties:

- \( u \in V \), i.e., \( u = cv \) with some unknown \( c \in \mathbb{R} \)
- \( u - b \perp V \), i.e., \( v \cdot (u - b) = 0 \)

By plugging \( u = cv \) into the second property we get by multiplying out

\[ v \cdot (cv - b) = 0 \iff c(v \cdot v) - (v \cdot b) = 0 \iff c = \frac{v \cdot b}{v \cdot v} \]

Therefore the point \( u \) on the line given by \( v \) which is closest to the point \( b \) is given by

\[ u = \frac{v \cdot b}{v \cdot v} v \]

We say that \( u \) is the **orthogonal projection of the point \( b \) onto the line given by \( v \)** and use the notation

\[ \text{pr}_v b = \frac{v \cdot b}{v \cdot v} v \]
Orthogonal projection onto a 2-dimensional subspace $V$

Consider a 2-dimensional subspace $V = \text{span}\{v^{(1)}, v^{(2)}\}$ of $\mathbb{R}^n$ given by two linearly independent vectors $v^{(1)}, v^{(2)} \in \mathbb{R}^n$. This is a plane through the origin.

For a given point $b \in \mathbb{R}^n$ we want to find the point $u \in V$ which is closest to the point $b$:

\[
\text{Find } u \in V \text{ such that } \|u - b\| \text{ is minimal}
\]

The point $u$ must have the following properties:

- $u \in V$, i.e., $u = c_1v^{(1)} + c_2v^{(2)}$ with some unknowns $c_1, c_2 \in \mathbb{R}$
- $u - b \perp V$, i.e., $v^{(1)} \cdot (u - b) = 0$ and $v^{(2)} \cdot (u - b) = 0$

By plugging $u = c_1v^{(1)} + c_2v^{(2)}$ into the second property we obtain a linear system of two equations for two unknowns which we can then solve.

We can use the $k \times 2$ matrix $A = [v^{(1)}, v^{(2)}]$ to express the two properties:

- $u \in V$, i.e., $u = Ac$ with an unknown vector $c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \in \mathbb{R}^2$
- $u - b \perp V$, i.e., $v^{(1)}^\top(u - b) = 0$ and $v^{(2)}^\top(u - b) = 0$, i.e., $\begin{bmatrix} v^{(1)}^\top \\ v^{(2)}^\top \end{bmatrix}(u - b) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ or
  \[
  A^\top(u - b) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}
  \]

By plugging $u = Ac$ into the second property we obtain

\[
A^\top (Ac - b) = 0
\]
\[
A^\top Ac = A^\top b
\]

These are the so-called normal equations (since they express that $u - b$ is orthogonal or normal on the subspace $V$).

This is how to find the point $u \in V$ which is closest to $b$:

- find the matrix $M := A^\top A \in \mathbb{R}^{2 \times 2}$ and the vector $g := A^\top b \in \mathbb{R}^2$
- solve the $2 \times 2$ linear system $Mc = g$ for $c \in \mathbb{R}^2$
- let $u := Ac$

**Example 2.** Consider the plane $V = \text{span}\left\{ \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$. Find the point $u \in V$ which is closest to $b = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$.

Let $A = \begin{bmatrix} -1 & 2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$. Then

\[
M = A^\top A = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}, \quad g = \begin{bmatrix} 5 \\ 4 \end{bmatrix}
\]

Solving the linear system $\begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 4 \end{bmatrix}$ gives $c = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$.

Hence the closest point is $u = Ac = \begin{bmatrix} 1 \/ 2 \\ 1 \/ 2 \\ 1 \/ 2 \end{bmatrix}$. We can check that this is correct by finding the difference vector $r = u - b$ and checking $v^{(1)} \cdot r = 0$ and $v^{(2)} \cdot r = 0$: We have

\[
r = Ac - b = \begin{bmatrix} 1 \/ 2 \\ 1 \/ 2 \\ -3 \/ 2 \end{bmatrix} \quad v^{(1)} \cdot r = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \/ 2 \\ 1 \/ 2 \\ -3 \/ 2 \end{bmatrix} = 0, \quad v^{(2)} \cdot r = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \/ 2 \\ 1 \/ 2 \\ -3 \/ 2 \end{bmatrix} = 0.
\]
Orthogonal projection onto a $k$-dimensional subspace $V$ a.k.a. “least squares problem”

Consider a $k$-dimensional subspace $V = \text{span}\{v^{(1)}, \ldots, v^{(k)}\}$ of $\mathbb{R}^n$ given by $k$ linearly independent vectors $v^{(1)}, \ldots, v^{(k)} \in \mathbb{R}^n$. I.e., the vectors $v^{(1)}, \ldots, v^{(k)}$ form a basis for the subspace $V$.

For a given point $b \in \mathbb{R}^n$ we want to find the point $u \in V$ which is closest to the point $b$:

Find $u \in V$ such that $\|u - b\|$ is minimal \hfill (2)

Let us define the $n \times k$ matrix $A = [v^{(1)}, \ldots, v^{(k)}]$.

The point $u$ must have the following properties:

• $u \in V$, i.e., $u = c_1 v^{(1)} + \cdots + c_k v^{(k)} = Ac$ with some unknown vector $c \in \mathbb{R}^k$

• $u - b \perp V$, i.e., $v^{(1)} \cdot (u - b) = 0, \ldots, v^{(k)} \cdot (u - b) = 0$, i.e.,

$$
\begin{bmatrix}
  v^{(1)\top} \\
  \vdots \\
  v^{(k)\top}
\end{bmatrix}
$$

$$(u - b) = \begin{bmatrix}
  0 \\
  \vdots \\
  0
\end{bmatrix}
$$

or

$$
A^\top (u - b) = \begin{bmatrix}
  0 \\
  \vdots \\
  0
\end{bmatrix}
$$

By plugging $u = Ac$ into the second property we obtain

$$
A^\top (Ac - b) = 0 \hfill \text{(3)}
$$

$$
A^\top Ac = A^\top b \hfill \text{(4)}
$$

These are the so-called normal equations (since they express that $u - b$ is “orthogonal” or “normal” on the subspace $V$.)

This is how to find the point $u \in V$ which is closest to $b$: 

![Diagram of orthogonal projection]
• find the matrix $M := A^\top A \in \mathbb{R}^{k \times k}$ and the vector $g := A^\top b \in \mathbb{R}^k$

• solve the $k \times k$ linear system $Mc = g$ for $c \in \mathbb{R}^k$

• let $u := Ac$

Note that the normal equations (4) always have a unique solution:

**Theorem 2.** Assume the matrix $A \in \mathbb{R}^{n \times k}$ with $k \leq n$ has linearly independent columns (i.e., $\text{rank} A = k$). Then the matrix $M = A^\top A$ is nonsingular.

**Proof.** We have to show that $Mc = \vec{0}$ implies $c = \vec{0}$.

Assume we have $c \in \mathbb{R}^k$ such that $Mc = \vec{0}$. Then we can multiply from the left with $c^\top$ and get with $y := Ac$

$$0 = c^\top Mc = c^\top A^\top Ac = y^\top y = \|y\|^2$$

as $y^\top = c^\top A^\top$. Since $\|y\| = 0$ we have $y = Ac = \vec{0}$. This means that we have a linear combination of the columns of $A$ which gives the zero vector. Since by assumption the columns of $A$ are linearly independent we must have $c = \vec{0}$.

So solving the normal equations gives us a unique $c \in \mathbb{R}^n$. We then get by $u = Ac$ a point on the subspace $V$. We now want to formally prove that this point $u \in V$ is really the unique answer to our minimization problem (2).

**Theorem 3.** Assume the matrix $A \in \mathbb{R}^{n \times k}$ with $k \leq n$ has linearly independent columns (i.e., $\text{rank} A = k$). Then for any given $b \in \mathbb{R}^n$ the minimization problem

$$\text{find } c \in \mathbb{R}^k \text{ such that } \|Ac - b\| \text{ is minimal} \tag{5}$$

has a unique solution which is obtained by solving the normal equations $A^\top Ac = A^\top b$.

**Proof.** Let $c \in \mathbb{R}^k$ be the unique solution of the normal equations. Consider now $\tilde{c} = c + d$ where $d \in \mathbb{R}^k$ is nonzero. We then have

$$\|A\tilde{c} - b\|^2 = \|Ac - b + Ad\|^2 = \left((Ac - b) + Ad\right) \cdot \left((Ac - b) + Ad\right)$$

$$= (Ac - b) \cdot (Ac - b) + 2(Ad) \cdot (Ac - b) + (Ad) \cdot (Ad)$$

We have for the middle term $(Ad) \cdot (Ac - b) = (Ad)^\top (Ac - b) = d^\top A^\top (Ac - b) = 0$ by the normal equations (3). Hence

$$\|A\tilde{c} - b\|^2 = \|Ac - b\|^2 + \|Ad\|^2$$

Since $d \neq \vec{0}$ we have $Ad \neq \vec{0}$ since the columns of $A$ are linearly independent, and hence $\|Ad\| > 0$. This means that for any vector $\tilde{c}$ different from $c$ we get $\|A\tilde{c} - b\| > \|Ac - b\|$, i.e., the vector $c$ from the normal equations is the unique solution of (5).

**Least squares problem with orthogonal basis**

For a least squares problem we are given $n$ linearly independent vectors $a^{(1)}, \ldots, a^{(n)} \in \mathbb{R}^m$ which form a basis for the subspace $V = \text{span}\{a^{(1)}, \ldots, a^{(n)}\}$. For a given right hand side vector $b \in \mathbb{R}^n$ we want to find $u \in V$ such that $\|u - b\|$ is minimal. We can write $u = c_1 a^{(1)} + \cdots + c_n a^{(n)} = Ac$ with the matrix $A = [a^{(1)}, \ldots, a^{(n)}] \in \mathbb{R}^{m \times n}$. Hence we want to find $c \in \mathbb{R}^n$ such that $\|Ac - b\|$ is minimal.

Solving this problem is much simpler if we have an **orthogonal basis for the subspace** $V$: Assume we have vectors $p^{(1)}, \ldots, p^{(n)}$ such that

• span $\{p^{(1)}, \ldots, p^{(n)}\} = V$

• the vectors are orthogonal on each other: $p^{(i)} \cdot p^{(j)} = 0$ for $i \neq j$
We can then write \( u = d_1 p^{(1)} + \cdots + d_n p^{(n)} = P d \) with the matrix \( P = [p^{(1)}, \ldots, p^{(n)}] \in \mathbb{R}^{m \times n} \). Hence we want to find \( d \in \mathbb{R}^n \) such that \( \| P d - b \| \) is minimal. The normal equations for this problem give
\[
(P^\top P) d = P^\top b
\]
where the matrix
\[
P^\top P = \begin{bmatrix}
p^{(1)^\top} \\
\vdots \\
p^{(n)^\top}
\end{bmatrix}
\begin{bmatrix}
p^{(1)} \\
\vdots \\
p^{(n)}
\end{bmatrix}
= \begin{bmatrix}
p^{(1)} \cdot p^{(1)} & \cdots & p^{(1)} \cdot p^{(n)} \\
\vdots & \ddots & \vdots \\
p^{(n)} \cdot p^{(1)} & \cdots & p^{(n)} \cdot p^{(n)}
\end{bmatrix}
= \begin{bmatrix}
p^{(1)} \cdot p^{(1)} & \cdots & 0 \\
0 & \ddots & \vdots \\
0 & \cdots & p^{(n)} \cdot p^{(n)}
\end{bmatrix}
\]
is now diagonal since \( p^{(i)} \cdot p^{(j)} = 0 \) for \( i \neq j \). Therefore the normal equations (6) are actually decoupled
\[
\begin{align*}
(p^{(1)} \cdot p^{(1)}) d_1 &= p^{(1)} \cdot b \\
(\vdots) \\
(p^{(n)} \cdot p^{(n)}) d_n &= p^{(n)} \cdot b
\end{align*}
\]
and have the solution
\[
d_i = \frac{p^{(i)} \cdot b}{p^{(i)} \cdot p^{(i)}} \quad \text{for } i = 1, \ldots, n
\]

**Gram-Schmidt orthogonalization**

We still need a method to construct from a given basis \( a^{(1)}, \ldots, a^{(n)} \) an orthogonal basis \( p^{(1)}, \ldots, p^{(n)} \).

Given \( n \) linearly independent vectors \( a^{(1)}, \ldots, a^{(n)} \in \mathbb{R}^m \) we want to find vectors \( p^{(1)}, \ldots, p^{(n)} \) such that

- \( \text{span} \{ p^{(1)}, \ldots, p^{(n)} \} = \text{span} \{ a^{(1)}, \ldots, a^{(n)} \} \)
- the vectors are orthogonal on each other: \( p^{(i)} \cdot p^{(j)} = 0 \) for \( i \neq j \)

**Step 1:** \( p^{(1)} := a^{(1)} \)

**Step 2:** \( p^{(2)} := a^{(2)} - s_{12} p^{(1)} \) where we choose \( s_{12} \) such that \( p^{(1)} \cdot p^{(2)} = 0 \):
\[
p^{(1)} \cdot a^{(2)} - s_{12} p^{(1)} \cdot p^{(1)} = 0 \quad \iff \quad s_{12} = \frac{p^{(1)} \cdot a^{(2)}}{p^{(1)} \cdot p^{(1)}}
\]

**Step 3:** \( p^{(3)} := a^{(3)} - s_{13} p^{(1)} - s_{23} p^{(2)} \) where we choose \( s_{13}, s_{23} \) such that
\[
\begin{align*}
p^{(1)} \cdot p^{(3)} &= 0, \text{ i.e., } p^{(1)} \cdot a^{(3)} - s_{13} p^{(1)} \cdot p^{(1)} - s_{23} p^{(2)} \cdot p^{(1)} = 0, \text{ hence } s_{13} = \frac{p^{(1)} \cdot a^{(3)}}{p^{(1)} \cdot p^{(1)}} \\
p^{(2)} \cdot p^{(3)} &= 0, \text{ i.e., } p^{(2)} \cdot a^{(3)} - s_{13} p^{(2)} \cdot p^{(1)} - s_{23} p^{(2)} \cdot p^{(2)} = 0, \text{ hence } s_{23} = \frac{p^{(2)} \cdot a^{(3)}}{p^{(2)} \cdot p^{(2)}}
\end{align*}
\]

**Step \( n \):** \( p^{(n)} := a^{(n)} - s_{1n} p^{(1)} - \cdots - s_{n-1,n} p^{(n-1)} \) where we choose \( s_{1n}, \ldots, s_{n-1,n} \) such that \( p^{(j)} \cdot p^{(n)} = 0 \) for \( j = 1, \ldots, n - 1 \) which yields
\[
s_{jn} = \frac{p^{(j)} \cdot p^{(n)}}{p^{(j)} \cdot p^{(j)}} \quad \text{for } j = 1, \ldots, n - 1
\]
Example: We are given the vectors $a^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $a^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$, $a^{(3)} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$. Use Gram-Schmidt orthogonalization to find an orthogonal basis $p^{(1)}, p^{(2)}, p^{(3)}$ for the subspace $V = \text{span} \{a^{(1)}, a^{(2)}, a^{(3)}\}$.

Step 1: $p^{(1)} := a^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

Step 2: $p^{(2)} := a^{(2)} - \frac{p^{(1)} \cdot a^{(2)}}{p^{(1)} \cdot p^{(1)}} p^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} - \frac{6}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{3}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$

Step 3: $p^{(3)} := a^{(3)} - \frac{p^{(1)} \cdot a^{(3)}}{p^{(1)} \cdot p^{(1)}} p^{(1)} - \frac{p^{(2)} \cdot a^{(3)}}{p^{(2)} \cdot p^{(2)}} p^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix} - \frac{14}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{15}{5} \begin{bmatrix} -\frac{3}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$

Note that we have

$$a^{(1)} = p^{(1)}$$
$$a^{(2)} = p^{(2)} + \frac{6}{4} p^{(1)}$$
$$a^{(3)} = p^{(3)} + \frac{14}{4} p^{(1)} + \frac{15}{5} p^{(2)}$$

which we can write as

$$\begin{bmatrix} a^{(1)}, a^{(2)}, a^{(3)} \end{bmatrix} = \begin{bmatrix} p^{(1)}, p^{(2)}, p^{(3)} \end{bmatrix} \begin{bmatrix} 1 & 6 & 14 \\ \frac{4}{3} & \frac{15}{5} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{vmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{vmatrix} = A$$

$$\begin{vmatrix} 1 & -1.5 & 1 \\ 1 & -0.5 & -1 \\ 1 & 0.5 & -1 \\ 1 & 1.5 & 1 \end{vmatrix} = P$$

$$\begin{vmatrix} 1 & 1.5 & 3.5 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{vmatrix} = S$$

In the general case we have

$$a^{(1)} = p^{(1)}$$
$$a^{(2)} = p^{(2)} + s_{12} p^{(1)}$$
$$a^{(3)} = p^{(3)} + s_{13} p^{(1)} + s_{13} p^{(2)}$$
$$\vdots$$
$$a^{(n)} = p^{(n)} + s_{1n} p^{(1)} + \cdots + s_{n-1,n} p^{(n-1)}$$

which we can write as

$$\begin{bmatrix} a^{(1)}, a^{(2)}, \ldots, a^{(n)} \end{bmatrix} = \begin{bmatrix} p^{(1)}, p^{(2)}, \ldots, p^{(n)} \end{bmatrix} \begin{bmatrix} 1 & s_{12} & \cdots & s_{1n} \\ 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & s_{n-1,n} \\ 0 & \cdots & 0 & 1 \end{bmatrix}$$

Therefore we obtain a decomposition $A = PS$ where
\* \( P \in \mathbb{R}^{m \times n} \) has orthogonal columns
\* \( S \in \mathbb{R}^{n \times n} \) is upper triangular, with 1 on the diagonal.

Note that the vectors \( p^{(1)}, \ldots, p^{(n)} \) are different from \( \mathbf{0} \): Assume, e.g., that \( p^{(3)} = a^{(3)} - s_{13} p^{(1)} - s_{23} p^{(2)} = \mathbf{0} \), then \( a^{(3)} = s_{13} p^{(1)} + s_{23} p^{(2)} \) is in span \( \{ p^{(1)}, p^{(2)} \} = \text{span} \{ a^{(1)}, a^{(2)} \} \).

This is a contradiction to the assumption that \( a^{(1)}, a^{(2)}, a^{(3)} \) are linearly independent.

**Solving the least squares problem** \( \| A c - b \| = \min \) using orthogonization

We are given \( A \in \mathbb{R}^{m \times n} \) with linearly independent columns, \( b \in \mathbb{R}^n \). We want to find \( c \in \mathbb{R}^n \) such that \( \| A c - b \| = \min \).

From the Gram-Schmidt method we get \( A = PS \), hence we want to find \( c \) such that
\[
\| P \underbrace{Sc}_{d} - b \| = \min
\]

This gives the following method for solving the least squares problem:

\* use Gram-Schmidt to find decomposition \( A = PS \)
\* solve \( \| Pd - b \| = \min \): \( d_i := \frac{p^{(i)} \cdot b}{p^{(i)} \cdot p^{(i)}} \) for \( i = 1, \ldots, n \)
\* solve \( Sc = d \) by back substitution

**Example:** Solve the least squares problem \( \| A c - b \| = \min \) for \( A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix}, b = \begin{bmatrix} 0 \\ 1 \\ 4 \\ 7 \end{bmatrix} \).

\* Gram-Schmidt gives \( A = \begin{bmatrix} 1 & -1.5 & 1 \\ 1 & -0.5 & -1 \\ 1 & 0.5 & -1 \\ 1 & 1.5 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1.5 & 3.5 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \) (see above)

\* \( d_1 = \frac{p^{(1)} \cdot b}{p^{(1)} \cdot p^{(1)}} = \frac{12}{4} = 3 \), \( d_2 = \frac{p^{(2)} \cdot b}{p^{(2)} \cdot p^{(2)}} = \frac{12}{5} = 2.4 \), \( d_3 = \frac{p^{(3)} \cdot b}{p^{(3)} \cdot p^{(3)}} = \frac{2}{4} = 0.5 \)

\* solving \( \begin{bmatrix} 1 & 1.5 & 3.5 \\ 0 & 1 & 3 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2.4 \\ 0.5 \end{bmatrix} \) by back substitution gives \( c_3 = 0.5, c_2 = 0.9, c_1 = -1.1 \)

Hence the solution of our least squares problem is the vector \( c = \begin{bmatrix} -1.1 \\ 0.9 \\ 0.5 \end{bmatrix} \).

**Note:** If you want to solve a least squares problem by hand with pencil and paper, it is usually easier to use the normal equations. But for numerical computation on a computer using orthogonalization is usually more efficient and more accurate.

**Finding an orthonormal basis** \( q^{(1)}, \ldots, q^{(n)} \): the QR decomposition

The Gram-Schmidt method gives an orthogonal basis \( p^{(1)}, \ldots, p^{(n)} \) for \( V = \text{span} \{ a^{(1)}, \ldots, a^{(n)} \} \).

Often it is convenient to have a so-called orthonormal basis \( q^{(1)}, \ldots, q^{(n)} \) where the basis vectors have length 1: Define
\[
q^{(j)} = \frac{1}{\| p^{(j)} \|} p^{(j)} \quad \text{for} \ j = 1, \ldots, n
\]
then we have
• \( \text{span} \{q^{(1)}, \ldots, q^{(n)}\} = V \)
• \( q^{(j)} \cdot q^{(k)} = \begin{cases} 1 & \text{for } j = k \\ 0 & \text{otherwise} \end{cases} \)

This means that the matrix \( Q = [q^{(1)}, \ldots, q^{(n)}] \) satisfies \( Q^T Q = I \) where \( I \) is the \( n \times n \) identity matrix.

Since \( p^{(j)} = \|p^{(j)}\|q^{(j)} \) we have

\[
a^{(1)} = \begin{bmatrix} p^{(1)} \end{bmatrix}_{r_{11}} \qquad \begin{bmatrix} q^{(1)} \end{bmatrix}_{r_{11}}
\]
\[
a^{(2)} = \begin{bmatrix} p^{(2)} \end{bmatrix}_{r_{22}} + s_{12} \begin{bmatrix} p^{(1)} \end{bmatrix}_{r_{12}} \begin{bmatrix} q^{(2)} \end{bmatrix}_{r_{22}}
\]
\[\vdots\]
\[
a^{(n)} = \begin{bmatrix} p^{(n)} \end{bmatrix}_{r_{nn}} + s_{1n} \begin{bmatrix} p^{(1)} \end{bmatrix}_{r_{1n}} + \cdots + s_{n-1,n} \begin{bmatrix} p^{(n-1)} \end{bmatrix}_{r_{n-1,n}} \begin{bmatrix} q^{(n)} \end{bmatrix}_{r_{nn}}
\]

which we can write as

\[
\begin{bmatrix} a^{(1)}, a^{(2)}, \ldots, a^{(n)} \end{bmatrix} = \begin{bmatrix} q^{(1)}, q^{(2)}, \ldots, q^{(n)} \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & r_{nn} \end{bmatrix}
\]

\( A = QR \)

where the \( n \times n \) matrix \( R \) is given by

\[
\begin{bmatrix} \text{row 1 of } R \\ \vdots \\ \text{row not of } R \end{bmatrix} = \begin{bmatrix} \|p^{(1)}\| (\text{row 1 of } R) \\ \vdots \\ \|p^{(n)}\| (\text{row not of } R) \end{bmatrix}
\]

We obtain the so-called \textbf{QR decomposition} \( A = QR \) where

• the matrix \( Q \in \mathbb{R}^{m \times n} \) has orthonormal columns, range \( Q = \text{range} A \)
• the matrix \( R \in \mathbb{R}^{n \times n} \) is upper triangular, with nonzero diagonal elements

\textbf{Example:} In our example we have \( p^{(1)} \cdot p^{(1)} = 4, p^{(2)} \cdot p^{(2)} = 5, p^{(3)} \cdot p^{(3)} = 4, \) hence

\[
q^{(1)} = \frac{1}{2} p^{(1)} = \begin{bmatrix} .5 \\ .5 \\ .5 \end{bmatrix}, \quad q^{(2)} = \frac{1}{\sqrt{5}} p^{(2)} = \frac{1}{\sqrt{5}} \begin{bmatrix} -1.5 \\ -.5 \\ .5 \end{bmatrix}, \quad q^{(3)} = \frac{1}{2} p^{(3)} = \begin{bmatrix} .5 \\ -.5 \\ -.5 \end{bmatrix}
\]

and we obtain the QR decomposition

\[
\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} = \begin{bmatrix} .5 & -1.5/\sqrt{5} & .5 \\ .5 & -.5/\sqrt{5} & -.5 \\ .5 & .5/\sqrt{5} & -.5 \\ .5 & 1.5/\sqrt{5} & .5 \end{bmatrix} \begin{bmatrix} 2 & 3 & 7 \\ 0 & \sqrt{5} & 3\sqrt{5} \\ 0 & 0 & 2 \end{bmatrix}
\]
In Matlab we can find the QR decomposition using\( [Q, R] = \text{qr}(A, 0) \)

This works for \textbf{symbolic matrices}

\[
>> A = \text{sym}([1 1 1 1; 0 1 2 3; 0 1 4 9])';
>> [Q, R] = \text{qr}(A, 0)
\]

\[
Q = \\
\begin{bmatrix}
\frac{1}{2}, \frac{-3+5^{(1/2)}}{10}, \frac{1}{2} \\
\frac{1}{2}, \frac{-5^{(1/2)}}{10}, -\frac{1}{2} \\
\frac{1}{2}, \frac{5^{(1/2)}}{10}, -\frac{1}{2} \\
\frac{1}{2}, \frac{3+5^{(1/2)}}{10}, \frac{1}{2}
\end{bmatrix}
\]

\[
R = \\
\begin{bmatrix}
2, 3, 7 \\
0, 5^{(1/2)}, 3+5^{(1/2)} \\
0, 0, 2
\end{bmatrix}
\]

and it works for \textbf{numerical matrices}

\[
>> A = [1 1 1 1; 0 1 2 3; 0 1 4 9]';
>> [Q, R] = \text{qr}(A, 0)
\]

\[
Q = \\
\begin{bmatrix}
-0.5000, 0.6708, 0.5000, 0.2236 \\
-0.5000, 0.2236, -0.5000, -0.5000 \\
-0.5000, -0.2236, -0.5000, 0.6708 \\
-0.5000, -0.6708, 0.5000, -0.2236
\end{bmatrix}
\]

\[
R = \\
\begin{bmatrix}
-2.0000, -3.0000, -7.0000 \\
0, -2.2361, -6.7082 \\
0, 0, 2.0000 \\
0, 0, 0
\end{bmatrix}
\]

Note that for numerical matrices Matlab returned the basis \(-q^{(1)}, -q^{(2)}, q^{(3)}\) (which is also an orthonormal basis) and hence rows 1 and 2 of the matrix \(R\) is \((-1)\) times our previous matrix \(R\).

If we want to find an orthonormal basis for range\(A\) and an orthonormal basis for the orthogonal complement \((\text{range } A)^\perp = \ker A^\top\) we can use the command \( [Q_h, R_h] = \text{qr}(A) \): It returns matrices \( \hat{Q} \in \mathbb{R}^{m \times m} \) and \( \hat{R} \in \mathbb{R}^{m \times n} \) with

\[
\hat{Q} = \begin{bmatrix}
\text{basis for range } A, \text{ basis for (range } A)^\perp \\
q^{(1)}, \ldots, q^{(n)}, q^{(n+1)}, \ldots, q^{(m)}
\end{bmatrix}, \quad \hat{R} = \begin{bmatrix}
R \\
0 \cdots 0 \\
\vdots \vdots \\
0 \cdots 0
\end{bmatrix}
\]

\[
\text{m-n rows of zeros}
\]

\[
>> A = [1 1 1 1; 0 1 2 3; 0 1 4 9]';
>> [Q_h, R_h] = \text{qr}(A)
\]

\[
Q_h = \\
\begin{bmatrix}
-0.5000, 0.6708, 0.5000, 0.2236 \\
-0.5000, 0.2236, -0.5000, -0.6708 \\
-0.5000, -0.2236, -0.5000, 0.6708 \\
-0.5000, -0.6708, 0.5000, -0.2236
\end{bmatrix}
\]

\[
R_h = \\
\begin{bmatrix}
-2.0000, -3.0000, -7.0000 \\
0, -2.2361, -6.7082 \\
0, 0, 2.0000 \\
0, 0, 0
\end{bmatrix}
\]

But in most cases we only need an orthonormal basis for range\(A\) and we should use \([Q, R] = \text{qr}(A, 0)\) (which Matlab calls the “economy size” decomposition).
Solving the least squares problem \( \| Ac - b \| = \min \) using the QR decomposition

If we use an orthonormal basis \( q^{(1)}, \ldots, q^{(n)} \) for \( \text{span}\{ a^{(1)}, \ldots, a^{(n)} \} \) we have \( Q^T Q = I \). The solution of \( \| Qd - b \| = \min \) is therefore given by the normal equations \( Q^T Q d = Q^T b \), i.e., we obtain \( d = Q^T b \).

This gives the following method for solving the least squares problem:

- find the QR decomposition \( A = QR \)
- let \( d = Q^T b \)
- solve \( R c = d \) by back substitution

In MATLAB we can do this as follows:

```matlab
[Q, R] = qr(A, 0);
d = Q'*b;
c = R\d;
```

In our example we have

```matlab
g = [1 1 1 1; 0 1 2 3; 0 1 4 9]'; b = [0;1;4;7];
g = qr(A, 0);
g = d = Q'*b;
g = c = R\d
c =
-0.1000
 0.9000
 0.5000
```

This works for both numerical and symbolic matrices.

**For a numerical matrix** \( A \) we can use the shortcut \( c = A\backslash y \) which actually uses the QR decomposition to find the solution of \( \| Ac - b \| = \min \)

```matlab
g = [1 1 1 1; 0 1 2 3; 0 1 4 9]'; b = [0;1;4;7];
g = c = A\b
c =
-0.1000
 0.9000
 0.5000
```

**Warning:** This shortcut does not work for symbolic matrices:

```matlab
g = sym([1 1 1 1; 0 1 2 3; 0 1 4 9])'; b = sym([0;1;4;7]);
g = c = A\b
Warning: The system is inconsistent. Solution does not exist.
c =
Inf
Inf
Inf