

Inner Products and Orthogonality

Last modified: 2026-10-27 20:53:54 (PDT)

Inner products

Definition 0.1 (Inner product). An **inner product** [1] on $\mathbb{R}^p$ is a function that assigns to each ordered pair of vectors $x, y \in \mathbb{R}^p$ a real number $\langle x, y \rangle$, and that has the following four properties for all $x, y, z \in \mathbb{R}^p$ and all real numbers a, b :

- **positivity:** $\langle x, x \rangle \geq 0$;
- **definiteness:** $\langle x, x \rangle = 0$ if and only if $x = 0_{p \times 1}$;
- **linearity in the first slot:** $\langle a x + b y, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$;
- **symmetry:** $\langle x, y \rangle = \langle y, x \rangle$.

i Remark (The definition in other sources)

Remark 0.1 (The definition in other sources). Axler [2], Definition 6.2, p. 183 states the same definition for any vector space over the real or complex numbers. Axler splits linearity in the first slot into additivity and homogeneity, and his last property is *conjugate* symmetry, $\langle u, v \rangle = \overline{\langle v, u \rangle}$, which for real numbers is the same as symmetry. Symmetry and linearity in the first slot together give linearity in the second slot.

Example 0.1 (The dot product is an inner product). The dot product (1) is an inner product on $\mathbb{R}^p$, called the **Euclidean** or **standard** inner product. Each of the four properties in Definition 0.1 holds:

- **positivity:** $x \cdot x = \sum_{i=1}^p x_i^2$ is a sum of squares, so it is at least 0;
- **definiteness:** a sum of squares is 0 exactly when every term is 0, so $x \cdot x = 0$ exactly when every $x_i = 0$, that is, when $x = 0_{p \times 1}$;
- **linearity in the first slot:** this property is the first-slot half of [Theorem 5 in Direct Sums and Orthogonal Complements](#);
- **symmetry:** this property is [Theorem 1 in Vectors](#).

For instance, with $\tilde{x} = (1, -2, 2)$,

$$\begin{aligned}\tilde{x} \cdot \tilde{x} &= 1 \cdot 1 + (-2) \cdot (-2) + 2 \cdot 2 && \text{(definition of the dot product)} \\ &= 1 + 4 + 4 && \text{(multiply)} \\ &= 9 && \text{(add)}\end{aligned}$$

which is positive, as positivity and definiteness require of a nonzero vector.

Definition 0.2 (Inner product space). A **real inner product space** [1] is a pair $(V, \langle \cdot, \cdot \rangle)$ of a real vector space V and an inner product $\langle \cdot, \cdot \rangle$ on it, a function with the four properties of Definition 0.1. We say that V is **equipped with** that inner product: the inner product is part of the data of the space, so the same set of vectors with a different inner product is a different inner product space. Definition 0.1 states the properties for $V = \mathbb{R}^p$ (Definition 2 in Vectors), and the same four properties define an inner product on other real vector spaces, such as spaces of matrices or functions.

Example 0.2 (Real coordinate space as an inner product space). The Euclidean space $\mathbb{R}^p$ equipped with the standard dot product (1), that is, the pair $(\mathbb{R}^p, \langle \cdot, \cdot \rangle)$ with $\langle \tilde{x}, \tilde{y} \rangle = \tilde{x} \cdot \tilde{y}$, is an inner product space, because the dot product satisfies the four inner product axioms (Example 0.1). For $p = 2$, any vector $\tilde{x} = (x_1, x_2) \in \mathbb{R}^2$ has squared norm $\langle \tilde{x}, \tilde{x} \rangle = x_1^2 + x_2^2 \geq 0$, which is zero if and only if

$$\begin{aligned}x_1 &= x_2 \\ &= 0.\end{aligned}$$

Equipping the same vector space $\mathbb{R}^p$ with different inner products yields different inner product spaces with different geometric notions of angle and distance on the same set of vectors.

Example 0.3 (A weighted inner product). Let $\tilde{c} = (c_1, \dots, c_p)$ be a vector of positive numbers, and define

$$\langle \tilde{x}, \tilde{y} \rangle_{\tilde{c}} \stackrel{\text{def}}{=} \sum_{i=1}^p c_i x_i y_i.$$

This function is also an inner product on $\mathbb{R}^p$ [2], Example 6.3(b), p. 184:

- **positivity:** $\langle \tilde{x}, \tilde{x} \rangle_{\tilde{c}} = \sum_{i=1}^p c_i x_i^2$, and each term $c_i x_i^2$ is a positive number times a square, so the sum is at least 0;
- **definiteness:** a sum of nonnegative terms is 0 exactly when every term is 0, and $c_i x_i^2 = 0$ exactly when $x_i = 0$, because $c_i > 0$; so $\langle \tilde{x}, \tilde{x} \rangle_{\tilde{c}} = 0$ exactly when $\tilde{x} = \mathbf{0}_{p \times 1}$;
- **linearity in the first slot:** see the derivation after this list;

- **symmetry:** $c_i x_i y_i = c_i y_i x_i$ for each i , so $\langle \tilde{x}, \tilde{y} \rangle_{\tilde{c}} = \langle \tilde{y}, \tilde{x} \rangle_{\tilde{c}}$.

For linearity in the first slot, take $\tilde{x}, \tilde{y}, \tilde{z} \in \mathbb{R}^p$ and real numbers a, b :

$$\begin{aligned}
\langle a \tilde{x} + b \tilde{y}, \tilde{z} \rangle_{\tilde{c}} &= \sum_{i=1}^p c_i (a \tilde{x}_i + b \tilde{y}_i) \tilde{z}_i && \text{(definition of } \langle \cdot, \cdot \rangle_{\tilde{c}} \text{)} \\
&= \sum_{i=1}^p c_i (a x_i + b y_i) z_i && \text{(Definition 6 in Matrices, Definition 7 in Vectors)} \\
&= \sum_{i=1}^p (a c_i x_i z_i + b c_i y_i z_i) && \text{(distribute, and commute the factors in each product)} \\
&= \sum_{i=1}^p a c_i x_i z_i + \sum_{i=1}^p b c_i y_i z_i && \text{(split the finite sum)} \\
&= a \sum_{i=1}^p c_i x_i z_i + b \sum_{i=1}^p c_i y_i z_i && \text{(factor } a \text{ and } b \text{ out of the sums)} \\
&= a \langle \tilde{x}, \tilde{z} \rangle_{\tilde{c}} + b \langle \tilde{y}, \tilde{z} \rangle_{\tilde{c}} && \text{(definition of } \langle \cdot, \cdot \rangle_{\tilde{c}} \text{)}
\end{aligned}$$

Different inner products can give different numbers for the same pair of vectors. With $p = 2$, $\tilde{c} = (2, 1)$, $\tilde{x} = (1, 3)$ and $\tilde{y} = (4, -1)$:

$$\begin{aligned}
\langle \tilde{x}, \tilde{y} \rangle_{\tilde{c}} &= 2 \cdot 1 \cdot 4 + 1 \cdot 3 \cdot (-1) && \text{(definition of } \langle \cdot, \cdot \rangle_{\tilde{c}} \text{)} \\
&= 8 - 3 && \text{(multiply)} \\
&= 5 && \text{(subtract)}
\end{aligned}$$

while the dot product of the same vectors is

$$\begin{aligned}
\tilde{x} \cdot \tilde{y} &= 1 \cdot 4 + 3 \cdot (-1) && \text{(definition of the dot product)} \\
&= 4 - 3 && \text{(multiply)} \\
&= 1 && \text{(subtract)}
\end{aligned}$$

Example 0.4 (Functions that are not inner products). Each of these functions on $\mathbb{R}^2$ has some of the properties in Definition 0.1, but not all four:

- $f(\tilde{x}, \tilde{y}) \stackrel{\text{def}}{=} x_1 y_1 - x_2 y_2$ is symmetric, because $x_i y_i = y_i x_i$, and linear in the first slot, because it is a linear combination of x_1 and x_2 with coefficients y_1 and $-y_2$. It fails positivity: for $\tilde{x} = (0, 1)$,

$$\begin{aligned}
f(\tilde{x}, \tilde{x}) &= 0 \cdot 0 - 1 \cdot 1 \\
&= -1 \\
&< 0.
\end{aligned}$$

It also fails definiteness: for $\tilde{x} = (1, 1)$,

$$\begin{aligned}
f(\tilde{x}, \tilde{x}) &= 1 \cdot 1 - 1 \cdot 1 \\
&= 0
\end{aligned}$$

although $\tilde{x} \neq \tilde{0}_{2 \times 1}$.

- $g(\tilde{x}, \tilde{y}) \stackrel{\text{def}}{=} x_1 y_1$ is symmetric and linear in the first slot, for the same reasons as f , and has positivity, because $g(\tilde{x}, \tilde{x}) = x_1^2 \geq 0$. It fails definiteness: for $\tilde{x} = (0, 1)$,

$$\begin{aligned}
g(\tilde{x}, \tilde{x}) &= 0 \cdot 0 \\
&= 0
\end{aligned}$$

although $\tilde{x} \neq \tilde{0}_{2 \times 1}$.

- $h(\tilde{x}, \tilde{y}) \stackrel{\text{def}}{=} |x_1 y_1| + |x_2 y_2|$, where $|\cdot|$ is the [absolute value](#), is symmetric, because $|x_i y_i| = |y_i x_i|$, and has positivity and definiteness, because $h(\tilde{x}, \tilde{x}) = x_1^2 + x_2^2$ is a sum of squares. It is not linear in the first slot: take

$$\begin{aligned}
\tilde{x} &= \tilde{y} \\
&= (1, 0),
\end{aligned}$$

$a = -1$ and $b = 0$; then

$$\begin{aligned}
h(a\tilde{x}, \tilde{y}) &= |(-1) \cdot 1| + |0 \cdot 0| \\
&= 1,
\end{aligned}$$

while

$$\begin{aligned}
a h(\tilde{x}, \tilde{y}) &= (-1) \cdot (|1 \cdot 1| + |0 \cdot 0|) \\
&= -1.
\end{aligned}$$

Definition 0.3 (Orthogonality in an inner product space). In an inner product space (Definition 0.2) with inner product $\langle \cdot, \cdot \rangle$, two vectors $\tilde{x}$ and $\tilde{y}$ are **orthogonal with respect to $\langle \cdot, \cdot \rangle$** , written $\tilde{x} \perp \tilde{y}$, if

$$\langle \tilde{x}, \tilde{y} \rangle = 0$$

[3], Definition 7.3, p. 181.

Example 0.5 (Orthogonality depends on the inner product). Take the weighted inner product of Example 0.3 with $\tilde{c} = (2, 1)$, and let $\tilde{x} = (1, 2)$ and $\tilde{y} = (1, -1)$. Then

$$\begin{aligned} \langle \tilde{x}, \tilde{y} \rangle_{\tilde{c}} &= 2 \cdot 1 \cdot 1 + 1 \cdot 2 \cdot (-1) \quad (\text{definition of } \langle \cdot, \cdot \rangle_{\tilde{c}}) \\ &= 2 - 2 \quad (\text{multiply}) \\ &= 0, \quad (\text{subtract}) \end{aligned}$$

so $\tilde{x}$ and $\tilde{y}$ are orthogonal with respect to $\langle \cdot, \cdot \rangle_{\tilde{c}}$. They are not orthogonal with respect to the dot product, since

$$\begin{aligned} \tilde{x} \cdot \tilde{y} &= 1 \cdot 1 + 2 \cdot (-1) \\ &= -1 \\ &\neq 0. \end{aligned}$$

i Remark (Orthogonality for the dot product)

Remark 0.2 (Orthogonality for the dot product). With the dot product as the inner product (Example 0.1), the condition of Definition 0.3 is orthogonality as in [Definition 19 in Vectors](#).

Cauchy-Schwarz and the triangle inequality

i Source

This section is adapted from Zhou [4], used under the MIT License (see the license text in [Section 2 in Subspaces and Rank](#)). The source leaves the proof of the triangle inequality to class; it is written out here. These parts of the source are not part of this section:

- its table of names for the angle between two vectors (aligned, acute, orthogonal, obtuse, opposed)
- the triangle inequality for distances
- the identity relating the average, the standard deviation and the root mean square of a vector

Theorem 0.1 (Basic properties of the Euclidean norm). *For any $\tilde{x} \in \mathbb{R}^p$ and any number c :*

1. $\|\tilde{x}\| \geq 0$, and $\|\tilde{x}\| = 0$ if and only if $\tilde{x} = 0$;
2. $\|c\tilde{x}\| = |c| \|\tilde{x}\|$.

i Proof

Proof. Part 1. $\|\tilde{x}\| = \sqrt{x_1^2 + \cdots + x_p^2}$ (Equation 2 in Vectors) is the square root of a sum of squares, so it is at least 0. It is 0 exactly when $x_1^2 + \cdots + x_p^2 = 0$, and a sum of squares is 0 exactly when every term is 0, that is, when every $x_i = 0$: if some $x_i \neq 0$, its square is positive and the sum is positive, and if every $x_i = 0$, the sum is 0.

Part 2.

$$\begin{aligned}
 \|c\tilde{x}\| &= \sqrt{\sum_{i=1}^p (cx_i)^2} && \text{(Equation 2 in Vectors, and Definition 6 in Matrices)} \\
 &= \sqrt{\sum_{i=1}^p c^2 x_i^2} && ((cx_i)^2 = c^2 x_i^2) \\
 &= \sqrt{c^2 \sum_{i=1}^p x_i^2} && \text{(factor } c^2 \text{ out of the sum)} \\
 &= \sqrt{c^2} \sqrt{\sum_{i=1}^p x_i^2} && \text{(the square root of a product of nonnegative numbers)} \\
 &= |c| \|\tilde{x}\|. && (\sqrt{c^2} = |c|, \text{ and Equation 2 in Vectors})
 \end{aligned}$$

□

Example 0.6 (Scaling a vector scales its length). For $\tilde{x} = (3, 4)$, $\|\tilde{x}\| = 5$ (Example 15 in Vectors). Then $-2\tilde{x} = (-6, -8)$ has norm

$$\begin{aligned}
 \sqrt{36 + 64} &= 10 \\
 &= |-2| \cdot 5,
 \end{aligned}$$

as part 2 says. Dividing by the norm gives a vector of length 1:

$$\begin{aligned}
 \left\| \frac{1}{5} \tilde{x} \right\| &= \frac{1}{5} \cdot 5 \\
 &= 1,
 \end{aligned}$$

and $\frac{1}{5} \tilde{x} = (0.6, 0.8)$ is the unit vector in Example 15 in Vectors.

Theorem 0.2 (The squared norm of a sum). *For any $\tilde{x}, \tilde{y} \in \mathbb{R}^p$,*

$$\|\tilde{x} + \tilde{y}\|^2 = \|\tilde{x}\|^2 + 2(\tilde{x} \cdot \tilde{y}) + \|\tilde{y}\|^2.$$

In particular, if $\tilde{x} \perp \tilde{y}$ (Definition 19 in Vectors), then $\|\tilde{x} + \tilde{y}\|^2 = \|\tilde{x}\|^2 + \|\tilde{y}\|^2$ (the Pythagorean theorem) [5].

i Proof

Proof.

$$\begin{aligned} \|\tilde{x} + \tilde{y}\|^2 &= (\tilde{x} + \tilde{y}) \cdot (\tilde{x} + \tilde{y}) && \text{(square both sides of Equation 2 in Vectors)} \\ &= \tilde{x} \cdot (\tilde{x} + \tilde{y}) + \tilde{y} \cdot (\tilde{x} + \tilde{y}) && \text{(Theorem 5 in Direct Sums and Orthogonal Complements, first slot, c)} \\ &= \tilde{x} \cdot \tilde{x} + \tilde{x} \cdot \tilde{y} + \tilde{y} \cdot \tilde{x} + \tilde{y} \cdot \tilde{y} && \text{(Theorem 5 in Direct Sums and Orthogonal Complements, second slot)} \\ &= \tilde{x} \cdot \tilde{x} + \tilde{x} \cdot \tilde{y} + \tilde{x} \cdot \tilde{y} + \tilde{y} \cdot \tilde{y} && \text{(Theorem 1 in Vectors)} \\ &= \tilde{x} \cdot \tilde{x} + 2(\tilde{x} \cdot \tilde{y}) + \tilde{y} \cdot \tilde{y} && \text{(combine like terms)} \\ &= \|\tilde{x}\|^2 + 2(\tilde{x} \cdot \tilde{y}) + \|\tilde{y}\|^2. && \text{(Equation 2 in Vectors, squared)} \end{aligned}$$

If $\tilde{x} \perp \tilde{y}$, then $\tilde{x} \cdot \tilde{y} = 0$, and the middle term drops out. □

Example 0.7 (Checking the expansion). For $\tilde{x} = (3, 0)$ and $\tilde{y} = (1, 4)$: $\tilde{x} + \tilde{y} = (4, 4)$, so

$$\begin{aligned} \|\tilde{x} + \tilde{y}\|^2 &= 16 + 16 \\ &= 32, \end{aligned}$$

and

$$\begin{aligned} \|\tilde{x}\|^2 + 2(\tilde{x} \cdot \tilde{y}) + \|\tilde{y}\|^2 &= 9 + 2 \cdot 3 + 17 \\ &= 32. \end{aligned}$$

For the orthogonal vectors $(3, 0)$ and $(0, 4)$,

$$\begin{aligned} \|(3, 4)\|^2 &= 25 \\ &= 9 + 16, \end{aligned}$$

the 3-4-5 right triangle.

Theorem 0.3 (Cauchy-Schwarz inequality). *For any $\tilde{x}, \tilde{y} \in \mathbb{R}^p$,*

$$|\tilde{x} \cdot \tilde{y}| \leq \|\tilde{x}\| \|\tilde{y}\|,$$

with equality exactly when $\tilde{x}$ and $\tilde{y}$ are linearly dependent (Definition 1 in Subspaces and

Rank) [6].

i Proof

Proof. If $y = 0$. Then

$$\begin{aligned} \tilde{x} \cdot \tilde{0} &= \sum_i x_i \cdot 0 \\ &= 0 \end{aligned}$$

(1) and $\|\tilde{0}\| = 0$ (Theorem 0.1, part 1), so both sides are 0 and equality holds; and $\tilde{x}, \tilde{y}$ are linearly dependent, because $0\tilde{x} + 1\tilde{y} = \tilde{0}$.

If $y \neq 0$. Then $\|y\| > 0$ (Theorem 0.1, part 1). Let $t \stackrel{\text{def}}{=} (\tilde{x} \cdot \tilde{y})/\|y\|^2$. Then

$$\begin{aligned} \|\tilde{x} - t\tilde{y}\|^2 &= \|\tilde{x} + (-1)(t\tilde{y})\|^2 && (\tilde{x} - \tilde{v} = \tilde{x} + (-1)\tilde{v}) \\ &= \|\tilde{x} + (-t)\tilde{y}\|^2 && ((-1)(t\tilde{y}) = (-t)\tilde{y}, \text{ entrywise}) \\ &= \|\tilde{x}\|^2 + 2(\tilde{x} \cdot (-t)\tilde{y}) + \|(-t)\tilde{y}\|^2 && (\text{Theorem 0.2}) \\ &= \|\tilde{x}\|^2 + 2(-t)(\tilde{x} \cdot \tilde{y}) + \|(-t)\tilde{y}\|^2 && (\text{Theorem 5 in Direct Sums and Orthogonal Complements, sp}) \\ &= \|\tilde{x}\|^2 + 2(-t)(\tilde{x} \cdot \tilde{y}) + (|-t|\|y\|)^2 && (\text{Theorem 0.1, part 2}) \\ &= \|\tilde{x}\|^2 + 2(-t)(\tilde{x} \cdot \tilde{y}) + |-t|^2\|y\|^2 && ((ab)^2 = a^2b^2) \\ &= \|\tilde{x}\|^2 - 2t(\tilde{x} \cdot \tilde{y}) + t^2\|y\|^2 && (2(-t) = -2t \text{ and } |-t|^2 = t^2) \\ &= \|\tilde{x}\|^2 - 2\frac{(\tilde{x} \cdot \tilde{y})^2}{\|y\|^2} + \frac{(\tilde{x} \cdot \tilde{y})^2}{\|y\|^4}\|y\|^2 && (\text{substitute } t) \\ &= \|\tilde{x}\|^2 - 2\frac{(\tilde{x} \cdot \tilde{y})^2}{\|y\|^2} + \frac{(\tilde{x} \cdot \tilde{y})^2}{\|y\|^2} && (\text{cancel one factor } \|y\|^2) \\ &= \|\tilde{x}\|^2 - \frac{(\tilde{x} \cdot \tilde{y})^2}{\|y\|^2}. && (\text{combine the last two terms}) \end{aligned} \tag{1}$$

The left side is at least 0 (Theorem 0.1, part 1), so

$$\begin{aligned} 0 &\leq \|\tilde{x}\|^2 - \frac{(\tilde{x} \cdot \tilde{y})^2}{\|y\|^2} && (\text{Equation 1}) \\ \frac{(\tilde{x} \cdot \tilde{y})^2}{\|y\|^2} &\leq \|\tilde{x}\|^2 && (\text{add } (\tilde{x} \cdot \tilde{y})^2/\|y\|^2 \text{ to both sides}) \\ (\tilde{x} \cdot \tilde{y})^2 &\leq \|\tilde{x}\|^2\|y\|^2, && (\text{multiply both sides by } \|y\|^2 > 0) \end{aligned}$$

and taking nonnegative square roots of both sides, which preserves the inequality, gives $|\tilde{x} \cdot \tilde{y}| \leq \|\tilde{x}\| \|y\|$.

When equality holds ($y \neq 0$). Both sides of the inequality are nonnegative, so equality holds exactly when $(x \cdot y)^2 = \|x\|^2 \|y\|^2$, that is (dividing by $\|y\|^2 > 0$ and rearranging), when $\|x\|^2 - (x \cdot y)^2 / \|y\|^2 = 0$. By Equation 1, that quantity is $\|x - t y\|^2$, so equality holds exactly when $x - t y = 0$ (Theorem 0.1, part 1), that is, when $x = t y$.

- If $x = t y$, then $1 x + (-t) y = 0$, so the two vectors are linearly dependent.
- Conversely, suppose $a x + b y = 0$ with a, b not both 0. Then $a \neq 0$: otherwise $b \neq 0$ and $b y = 0$, so

$$\begin{aligned} y &= \frac{1}{b} (b y) \\ &= 0, \end{aligned}$$

contrary to assumption. Subtracting $b y$ from both sides gives $a x = -b y$, and multiplying by $\frac{1}{a}$ gives $x = s y$ with $s = -b/a$. Then

$$\begin{aligned} |x \cdot y| &= |(s y) \cdot y| && \text{(substitute } x = s y) \\ &= |s (y \cdot y)| && \text{(Theorem 5 in Direct Sums and Orthogonal Complements, special case } b = 0, \text{ first part)} \\ &= |s| |y \cdot y| && (|ab| = |a| |b|) \\ &= |s| \|y\|^2 && (y \cdot y = \|y\|^2 \geq 0, \text{ Equation 2 in Vectors}) \\ &= (|s| \|y\|) \|y\| && \text{(regroup)} \\ &= \|s y\| \|y\| && \text{(Theorem 0.1, part 2)} \\ &= \|x\| \|y\|, && (s y = x) \end{aligned}$$

so equality holds. □

Example 0.8 (A strict case and an equality case).

- For $x = (1, 2)$ and $y = (3, 4)$:

$$\begin{aligned} x \cdot y &= 3 + 8 \\ &= 11, \end{aligned}$$

while $\|x\| \|y\| = \sqrt{5} \cdot 5 \approx 11.18$. The inequality is strict, and the vectors are linearly independent: $(3, 4)$ is not a multiple of $(1, 2)$, since $3 \cdot 2 \neq 4$.

- For $x = (1, 2)$ and

$$\begin{aligned} y &= (2, 4) \\ &= 2 \tilde{x} : \end{aligned}$$

$$\begin{aligned} \tilde{x} \cdot \tilde{y} &= 2 + 8 \\ &= 10 \end{aligned}$$

and

$$\begin{aligned} \|\tilde{x}\| \|\tilde{y}\| &= \sqrt{5} \sqrt{20} \\ &= \sqrt{100} \\ &= 10, \end{aligned}$$

so equality holds.

Theorem 0.4 (Triangle inequality). *For any $\tilde{x}, \tilde{y} \in \mathbb{R}^p$,*

$$\|\tilde{x} + \tilde{y}\| \leq \|\tilde{x}\| + \|\tilde{y}\|.$$

i Proof

Proof.

$$\begin{aligned} \|\tilde{x} + \tilde{y}\|^2 &= \|\tilde{x}\|^2 + 2(\tilde{x} \cdot \tilde{y}) + \|\tilde{y}\|^2 && \text{(Theorem 0.2)} \\ &\leq \|\tilde{x}\|^2 + 2|\tilde{x} \cdot \tilde{y}| + \|\tilde{y}\|^2 && \text{(a number is at most its absolute value)} \\ &\leq \|\tilde{x}\|^2 + 2\|\tilde{x}\| \|\tilde{y}\| + \|\tilde{y}\|^2 && \text{(Theorem 0.3)} \\ &= (\|\tilde{x}\| + \|\tilde{y}\|)^2. && \text{(expand the square)} \end{aligned}$$

Both $\|\tilde{x} + \tilde{y}\|$ and $\|\tilde{x}\| + \|\tilde{y}\|$ are nonnegative (Theorem 0.1), and for nonnegative numbers $a^2 \leq b^2$ implies $a \leq b$. $\square$

Example 0.9 (A strict case and an equality case).

- For $\tilde{x} = (3, 0)$ and $\tilde{y} = (0, 4)$:

$$\begin{aligned} \|\tilde{x} + \tilde{y}\| &= \|(3, 4)\| \\ &= 5, \end{aligned}$$

while

$$\begin{aligned}\|\tilde{x}\| + \|\tilde{y}\| &= 3 + 4 \\ &= 7.\end{aligned}$$

- For $\tilde{x} = (1, 0)$ and $\tilde{y} = (2, 0)$, which point the same way:

$$\begin{aligned}\|(3, 0)\| &= 3 \\ &= 1 + 2,\end{aligned}$$

so equality holds.

Definition 0.4 (Norm (normed vector space, normed space)). A **norm** [7] on a real vector space V is a function $\|\cdot\| : V \rightarrow [0, \infty)$ that assigns to each vector $\tilde{x} \in V$ a nonnegative real number $\|\tilde{x}\|$, and satisfies the following three axioms for all $\tilde{x}, \tilde{y} \in V$ and all scalars $c \in \mathbb{R}$:

1. **Positivity and definiteness**: $\|\tilde{x}\| \geq 0$, and $\|\tilde{x}\| = 0$ if and only if $\tilde{x} = \tilde{0}$;
2. **Absolute homogeneity**: $\|c\tilde{x}\| = |c| \|\tilde{x}\|$;
3. **Subadditivity (triangle inequality)**: $\|\tilde{x} + \tilde{y}\| \leq \|\tilde{x}\| + \|\tilde{y}\|$ [8].

A vector space equipped with a norm is called a **normed vector space** (or **normed space**).

Example 0.10 (Euclidean space as a normed vector space). The Euclidean space $\mathbb{R}^p$ equipped with the Euclidean norm (Definition 20 in Vectors) is a normed vector space:

- Positivity and absolute homogeneity hold by Theorem 0.1.
- The triangle inequality holds by Theorem 0.4.

Every inner product space $(\mathbb{R}^p, \langle \cdot, \cdot \rangle)$ (Definition 0.2) naturally defines an induced norm $\|\tilde{x}\| \stackrel{\text{def}}{=} \sqrt{\langle \tilde{x}, \tilde{x} \rangle}$, which satisfies all three norm axioms.

Definition 0.5 (Vector p -norm (ℓ_p norm, ℓ_1 norm, taxicab norm, Manhattan norm, ℓ_2 norm, infinity norm)). For a vector $\tilde{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$ and any real number $k \geq 1$, the **vector p -norm** [9] (or ℓ_p **norm**, traditionally named with parameter p but denoted here with exponent k to avoid clashing with the ambient vector dimension p) is:

$$\|\tilde{x}\|_k \stackrel{\text{def}}{=} \left(\sum_{i=1}^p |x_i|^k \right)^{1/k}.$$

The cases $k = 1$ and $k = 2$ are named the ℓ_1 **norm** (also the **taxicab norm** or **Manhattan**

norm) and the ℓ_2 **norm**.

Example 0.11 (Calculating ℓ_1 , ℓ_2 , and ℓ_∞ norms). For $\tilde{x} = (3, -4) \in \mathbb{R}^2$:

- ℓ_1 norm:

$$\begin{aligned}\|\tilde{x}\|_1 &= |3| + |-4| \\ &= 3 + 4 \\ &= 7.\end{aligned}$$

- ℓ_2 norm:

$$\begin{aligned}\|\tilde{x}\|_2 &= \sqrt{3^2 + (-4)^2} \\ &= \sqrt{9 + 16} \\ &= \sqrt{25} \\ &= 5.\end{aligned}$$

- ℓ_∞ norm:

$$\begin{aligned}\|\tilde{x}\|_\infty &= \max(|3|, |-4|) \\ &= 4.\end{aligned}$$

Notice that $4 \leq 5 \leq 7$. For any vector $\tilde{x} \in \mathbb{R}^p$, the norms satisfy the chain of inequalities:

$$\|\tilde{x}\|_\infty \leq \|\tilde{x}\|_2 \leq \|\tilde{x}\|_1 \leq \sqrt{p} \|\tilde{x}\|_2 \leq p \|\tilde{x}\|_\infty.$$

i Remark (Unit balls and regularization in machine learning)

Remark 0.3 (Unit balls and regularization in machine learning). The unit ball in $\mathbb{R}^2$ under the ℓ_p norm is the set $\{\tilde{x} \in \mathbb{R}^2 : \|\tilde{x}\|_k \leq 1\}$:

- For $k = 1$, the unit ball is a diamond (cross-polytope) with sharp vertices along the coordinate axes $(\pm 1, 0)$ and $(0, \pm 1)$.
- For $k = 2$, the unit ball is the smooth round Euclidean unit disk.
- For $k = \infty$, the unit ball is an axis-aligned square with vertices $(\pm 1, \pm 1)$.

In regression and machine learning, adding an ℓ_1 penalty (Lasso regularization) to an objective function encourages sparse solutions: the level curves of the loss function tend to make contact with the sharp corners of the ℓ_1 unit ball on the coordinate axes, setting irrelevant feature weights exactly to zero. In contrast, the smooth spherical ℓ_2 penalty

(Ridge regularization) shrinks all weights toward zero without setting them exactly to zero.

Definition 0.6 (Infinity norm (ℓ_∞ norm, maximum norm, Chebyshev norm)). For a vector $\tilde{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$, the **infinity norm** [10] (also the ℓ_∞ **norm**, **maximum norm**, or **Chebyshev norm**) is defined separately from the p -norms (Definition 0.5) by

$$\|\tilde{x}\|_\infty \stackrel{\text{def}}{=} \max_{1 \leq i \leq p} |x_i|.$$

Theorem 0.5 (The ℓ_1 , ℓ_2 and infinity norms, and the norm axioms). Let $\tilde{x} = (x_1, \dots, x_p) \in \mathbb{R}^p$.

1. $\|\tilde{x}\|_1 = \sum_{i=1}^p |x_i|$, and $\|\tilde{x}\|_2 = \sqrt{\sum_{i=1}^p x_i^2}$, which is the Euclidean norm (Definition 20 in Vectors).
2. $\|\tilde{x}\|_\infty = \lim_{k \rightarrow \infty} \|\tilde{x}\|_k$.
3. For every real number $k \geq 1$, and for $k = \infty$, $\|\cdot\|_k$ is a norm on $\mathbb{R}^p$ (Definition 0.4).

i Proof

Proof. **Part 1.** Setting $k = 1$ in Definition 0.5 gives $(\sum_{i=1}^p |x_i|)^{1/1} = \sum_{i=1}^p |x_i|$. Setting $k = 2$ gives $(\sum_{i=1}^p |x_i|^2)^{1/2}$, and $|x_i|^2 = x_i^2$ for every real number x_i .

Part 2. If $\tilde{x} = 0$, every $\|\tilde{x}\|_k$ is 0, and so is $\|\tilde{x}\|_\infty$. Otherwise, let $M \stackrel{\text{def}}{=} \max_i |x_i| > 0$, and pick j with $|x_j| = M$. Every term of $\sum_{i=1}^p |x_i|^k$ is at most M^k , and the j th term equals M^k , so

$$M^k \leq \sum_{i=1}^p |x_i|^k \leq p M^k.$$

Taking the $1/k$ power, which preserves order on $[0, \infty)$, gives

$$M \leq \|\tilde{x}\|_k \leq p^{1/k} M.$$

As $k \rightarrow \infty$, $p^{1/k} = \exp\left\{\frac{\ln p}{k}\right\} \rightarrow 1$, so $\|\tilde{x}\|_k$ is squeezed between M and a quantity that tends to M . Thus $\|\tilde{x}\|_k \rightarrow M = \|\tilde{x}\|_\infty$.

Part 3, finite k .

- *Positivity and definiteness.* Each term $|x_i|^k$ is at least 0, so $\|\tilde{x}\|_k \geq 0$. The sum $\sum_{i=1}^p |x_i|^k$ is 0 exactly when every term is 0, that is, when $\tilde{x} = \tilde{0}$.
- *Absolute homogeneity.* For a number c ,

$$\begin{aligned}
\|c \tilde{x}\|_k &= \left(\sum_{i=1}^p |c x_i|^k \right)^{1/k} && \text{(definition)} \\
&= \left(|c|^k \sum_{i=1}^p |x_i|^k \right)^{1/k} && (|c x_i| = |c| |x_i|, \text{ and factor out } |c|^k) \\
&= |c| \|\tilde{x}\|_k. && \text{(take the } 1/k \text{ power of each factor)}
\end{aligned}$$

- *Triangle inequality.* Let $y \in \mathbb{R}^p$. If $\tilde{x} = 0$ or $\tilde{y} = 0$, both sides of $\|\tilde{x} + \tilde{y}\|_k \leq \|\tilde{x}\|_k + \|\tilde{y}\|_k$ are equal. Otherwise, let $a \stackrel{\text{def}}{=} \|\tilde{x}\|_k > 0$, $b \stackrel{\text{def}}{=} \|\tilde{y}\|_k > 0$, $\lambda \stackrel{\text{def}}{=} \frac{a}{a+b}$, $\tilde{u} \stackrel{\text{def}}{=} \tilde{x}/a$ and $\tilde{v} \stackrel{\text{def}}{=} \tilde{y}/b$. By absolute homogeneity, $\|\tilde{u}\|_k = \|\tilde{v}\|_k = 1$, and

$$\frac{\tilde{x} + \tilde{y}}{a + b} = \lambda \tilde{u} + (1 - \lambda) \tilde{v}.$$

The function $s \mapsto s^k$ is non-decreasing and convex on $[0, \infty)$ (its second derivative $k(k-1)s^{k-2}$ is nonnegative). So for each coordinate i ,

$$\begin{aligned}
|\lambda u_i + (1 - \lambda) v_i|^k &\leq (\lambda |u_i| + (1 - \lambda) |v_i|)^k && \text{(triangle inequality for numbers; } s^k \text{ is non-decreasing)} \\
&\leq \lambda |u_i|^k + (1 - \lambda) |v_i|^k. && (s^k \text{ is convex})
\end{aligned}$$

Summing over i , and using $\sum_{i=1}^p |u_i|^k = \|\tilde{u}\|_k^k = 1$ and likewise for $\tilde{v}$,

$$\|\lambda \tilde{u} + (1 - \lambda) \tilde{v}\|_k^k \leq \lambda + (1 - \lambda) = 1.$$

So $\|\tilde{x} + \tilde{y}\|_k / (a + b) \leq 1$ by absolute homogeneity, that is, $\|\tilde{x} + \tilde{y}\|_k \leq a + b$.

Part 3, $k = \infty$. Every $|x_i|$ is at least 0, and the largest of them is 0 exactly when $\tilde{x} = 0$. For a number c , $\max_i |c x_i| = |c| \max_i |x_i|$. For each i ,

$$\begin{aligned}
|x_i + y_i| &\leq |x_i| + |y_i| && \text{(triangle inequality for numbers)} \\
&\leq \|\tilde{x}\|_\infty + \|\tilde{y}\|_\infty, && \text{(each is at most the maximum)}
\end{aligned}$$

and taking the maximum over i on the left gives $\|\tilde{x} + \tilde{y}\|_\infty \leq \|\tilde{x}\|_\infty + \|\tilde{y}\|_\infty$. □

Example 0.12 (The ℓ_k norm approaches the maximum norm as k grows). Take $\tilde{x} = (3, -4)$ and $\tilde{y} = (-1, 5)$ in $\mathbb{R}^2$. The code computes $\|\tilde{x}\|_k$ for several values of k , including $k = \infty$ as the maximum, and checks the triangle inequality for $\tilde{x}$ and $\tilde{y}$ at each of them.

```

lk_norm <- function(x, k) {
  if (is.infinite(k)) max(abs(x)) else sum(abs(x)^k)^(1 / k)
}

x <- c(3, -4)
y <- c(-1, 5)
ks <- c(1, 2, 4, 8, 16, 32, 64, Inf)
norms <- sapply(ks, function(k) lk_norm(x, k))
triangle <- sapply(
  ks,
  function(k) lk_norm(x + y, k) <= lk_norm(x, k) + lk_norm(y, k)
)
data.frame(k = ks, norm = norms, triangle_holds = triangle)
#>      k      norm triangle_holds
#> 1    1 7.000000             TRUE
#> 2    2 5.000000             TRUE
#> 3    4 4.284572             TRUE
#> 4    8 4.047992             TRUE
#> 5   16 4.002494             TRUE
#> 6   32 4.000013             TRUE
#> 7   64 4.000000             TRUE
#> 8  Inf 4.000000             TRUE

```

By part 1 of Theorem 0.5, the ℓ_1 norm 7 equals $|3| + |-4| = 7$, and the ℓ_2 norm 5 equals $\sqrt{3^2 + (-4)^2} = 5$. By part 2, the norms approach the maximum norm 4: at $k = 64$ the ℓ_k norm is within 0.00000000063 of it. By part 3, the triangle inequality holds at all 8 of the 8 values of k .

Definition 0.7 (Cosine similarity). The **cosine similarity** [11] between two nonzero vectors $\tilde{x}, \tilde{y} \in \mathbb{R}^p$ is their dot product divided by the product of their lengths:

$$\text{cossim}(\tilde{x}, \tilde{y}) \stackrel{\text{def}}{=} \frac{\tilde{x} \cdot \tilde{y}}{\|\tilde{x}\| \|\tilde{y}\|}.$$

Example 0.13 (Cosine similarity of parallel, orthogonal, and arbitrary vectors). For vectors in $\mathbb{R}^2$:

- Parallel vectors $\tilde{x} = (1, 2)$ and $\tilde{y} = (2, 4)$:

$$\begin{aligned}\text{cossim}(\tilde{x}, \tilde{y}) &= \frac{2+8}{\sqrt{5}\sqrt{20}} \\ &= \frac{10}{10} \\ &= 1.\end{aligned}$$

- Orthogonal vectors $\tilde{x} = (1, 2)$ and $\tilde{y} = (-2, 1)$:

$$\begin{aligned}\text{cossim}(\tilde{x}, \tilde{y}) &= \frac{-2+2}{\sqrt{5}\sqrt{5}} \\ &= \frac{0}{5} \\ &= 0.\end{aligned}$$

- Arbitrary vectors $\tilde{x} = (1, 0)$ and $\tilde{y} = (1, 1)$:

$$\begin{aligned}\text{cossim}(\tilde{x}, \tilde{y}) &= \frac{1 \cdot 1 + 0 \cdot 1}{1 \cdot \sqrt{2}} \\ &= \frac{1}{\sqrt{2}}.\end{aligned}$$

When both vectors have unit Euclidean norm ($\|\tilde{x}\| = 1$ and $\|\tilde{y}\| = 1$), cosine similarity equals their inner product: $\text{cossim}(\tilde{x}, \tilde{y}) = \langle \tilde{x}, \tilde{y} \rangle$.

Definition 0.8 (Angle between two vectors). The **angle** between two nonzero vectors, $\tilde{x}, \tilde{y} \in \mathbb{R}^p$, is

$$\theta \stackrel{\text{def}}{=} \arccos(\text{cossim}(\tilde{x}, \tilde{y})).$$

In words, the angle is the [arccosine](#) of the [cosine similarity](#) of the two vectors.

i Remark (Why the angle is well defined)

Remark 0.4 (Why the angle is well defined). The [cosine similarity](#) is defined, because both vectors are nonzero (Theorem 0.1). By Theorem 0.3, it lies in the [closed interval](#) $[-1, 1]$, which is where the [arccosine](#), $\arccos$, is defined. So the angle θ lies in $[0, \pi]$, with [cosine](#) $\cos \theta = \text{cossim}(\tilde{x}, \tilde{y})$.

Corollary 0.1 (A right angle means orthogonal). For nonzero vectors $\tilde{x}, \tilde{y} \in \mathbb{R}^p$ with angle θ (Definition 0.8),

$$\theta = \frac{\pi}{2} \iff \tilde{x} \cdot \tilde{y} = 0,$$

so for nonzero vectors, orthogonal ([Definition 19 in Vectors](#)) means at a right angle.

i Proof

Proof. Since $\|\tilde{x}\| \|\tilde{y}\| > 0$ ([Theorem 0.1](#)), $\text{cossim}(\tilde{x}, \tilde{y}) = 0$ exactly when $\tilde{x} \cdot \tilde{y} = 0$ ([Definition 0.7](#)), so $\cos \theta = 0$ exactly when $\tilde{x} \cdot \tilde{y} = 0$ ([Remark 0.4](#)). On $[0, \pi]$, $\cos$ vanishes only at $\pi/2$. $\square$

Example 0.14 (Some angles in the plane).

- $(1, 0)$ and $(1, 1)$: $\cos \theta = \frac{1}{1 \cdot \sqrt{2}}$, so $\theta = \pi/4$.
- $(1, 2)$ and $(-2, 1)$:

$$\begin{aligned} \cos \theta &= \frac{-2 + 2}{\sqrt{5} \sqrt{5}} \\ &= 0, \end{aligned}$$

so $\theta = \pi/2$.

- $(1, 0)$ and $(-3, 0)$:

$$\begin{aligned} \cos \theta &= \frac{-3}{1 \cdot 3} \\ &= -1, \end{aligned}$$

so $\theta = \pi$.

- $(1, 0)$ and $\tilde{0}$: there is no angle, because the ratio would divide by $\|\tilde{0}\| = 0$.

i Remark (How the two angles relate)

Remark 0.5 (How the two angles relate). For nonzero vectors in the plane, x, y , the [angle between the vectors](#) is the [angle on the unit circle](#) between the points $\tilde{x}/\|\tilde{x}\|, \tilde{y}/\|\tilde{y}\|$. The [cosine similarity](#) of the vectors is the [dot product](#) of these points. Rotating the plane puts the first point at $(1, 0)$, with no change to the distance between the points along the circle or to their dot product. For the angle on the unit circle, θ , the second point is then $(\cos \theta, \pm \sin \theta)$, so the dot product of the points is $\cos \theta$. Applying $\arccos$ to it returns θ , which lies in $[0, \pi]$.

Definition 0.9 (Cosine similarity in an inner product space). In a general real inner product space (Definition 0.2) with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\|\cdot\|$, the **cosine similarity** is

$$\text{cossim}(\underset{\sim}{x}, \underset{\sim}{y}) \stackrel{\text{def}}{=} \frac{\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle}{\|\underset{\sim}{x}\| \|\underset{\sim}{y}\|}.$$

Corollary 0.2 (Cosine similarity of unit vectors is their inner product). *In a real inner product space (Definition 0.2), for all vectors $\underset{\sim}{x}$ and $\underset{\sim}{y}$ with $\|\underset{\sim}{x}\| = \|\underset{\sim}{y}\| = 1$,*

$$\text{cossim}(\underset{\sim}{x}, \underset{\sim}{y}) = \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle.$$

i Proof

Proof. Both norms in the denominator of Definition 0.9 equal 1, so

$$\text{cossim}(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\langle \underset{\sim}{x}, \underset{\sim}{y} \rangle}{1 \cdot 1} = \langle \underset{\sim}{x}, \underset{\sim}{y} \rangle.$$

□

Theorem 0.6 (Pearson correlation is the cosine similarity of the centered vectors). *Let $x, y \in \mathbb{R}^n$ be two vectors that are not constant, with means $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. Write $\underset{\sim}{x}^c = \underset{\sim}{x} - \bar{x} \underset{\sim}{1}$ and $\underset{\sim}{y}^c = \underset{\sim}{y} - \bar{y} \underset{\sim}{1}$ for the **centered** vectors, whose means are zero. Then the **Pearson correlation coefficient** r of the n pairs (x_i, y_i) is the cosine similarity (Definition 0.7) of the centered vectors:*

$$r = \text{cossim}(\underset{\sim}{x}^c, \underset{\sim}{y}^c).$$

i Proof

Proof. Entry i of $\underset{\sim}{x}^c$ is $x_i - \bar{x}$, and entry i of $\underset{\sim}{y}^c$ is $y_i - \bar{y}$. Neither centered vector is zero, because neither $\underset{\sim}{x}$ nor $\underset{\sim}{y}$ is constant, so the cosine similarity is defined. The three pieces of the cosine similarity are

$$\begin{aligned}\tilde{x}^c \cdot \tilde{y}^c &= \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \\ \|\tilde{x}^c\| &= \sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \|\tilde{y}^c\| &= \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2},\end{aligned}$$

by the definitions of the dot product and the Euclidean norm. Dividing the first by the product of the other two gives the formula for r . $\square$

Example 0.15 (Centering changes the cosine similarity). Take $x = (1, 2, 3)$ and $y = (1, 3, 2)$. Their means are $\bar{x} = \bar{y} = 2$, so $\tilde{x}^c = (-1, 0, 1)$ and $\tilde{y}^c = (-1, 1, 0)$. Then

$$\begin{aligned}\tilde{x}^c \cdot \tilde{y}^c &= 1 + 0 + 0 = 1, \\ \|\tilde{x}^c\| &= \|\tilde{y}^c\| = \sqrt{2},\end{aligned}$$

so $r = \text{cossim}(\tilde{x}^c, \tilde{y}^c) = 1/2$.

Without centering, $\tilde{x} \cdot \tilde{y} = 1 + 6 + 6 = 13$ and $\|\tilde{x}\| = \|\tilde{y}\| = \sqrt{14}$, so $\text{cossim}(\tilde{x}, \tilde{y}) = 13/14$. The uncentered value is close to 1 mostly because both vectors have positive entries and the same mean, not because they move together.

Corollary 0.3 (Pearson correlation is the cosine of the angle between the centered vectors). *For the two data vectors $x, y \in \mathbb{R}^n$ of Theorem 0.6, the centered vectors $\tilde{x}^c, \tilde{y}^c$ are nonzero, so they have an angle θ . The Pearson correlation coefficient r satisfies*

$$r = \cos \theta.$$

Proof

Proof. By Theorem 0.6, $r = \text{cossim}(\tilde{x}^c, \tilde{y}^c)$. By Definition 0.8, $\theta = \arccos(\text{cossim}(\tilde{x}^c, \tilde{y}^c))$, so $\cos \theta = \text{cossim}(\tilde{x}^c, \tilde{y}^c)$ (Remark 0.4). $\square$

Example 0.16 (A correlation of one half is an angle of sixty degrees). In Example 0.15, the correlation is $r = 1/2$, for the data vectors $x = (1, 2, 3)$, $y = (1, 3, 2)$. By Corollary 0.3, the angle θ between the centered vectors $\tilde{x}^c = (-1, 0, 1)$, $\tilde{y}^c = (-1, 1, 0)$ satisfies

$$\cos \theta = \frac{1}{2},$$

so the angle is $\theta = \pi/3$.

Orthonormal bases and Gram-Schmidt

i Source

Like Section , this section is adapted from Zhou [4], used under the MIT License (see the license text in [Section 2 in Subspaces and Rank](#)). The source leaves the proof that orthonormal vectors are linearly independent to class; it is written out here.

Theorem 0.7 (The dot product of a vector with a linear combination). *For vectors $\underline{x}, \underline{u}_1, \dots, \underline{u}_k \in \mathbb{R}^p$ and numbers $c_1, \dots, c_k$, with $k \geq 1$,*

$$\underline{x} \cdot \left(\sum_{j=1}^k c_j \underline{u}_j \right) = \sum_{j=1}^k c_j (\underline{x} \cdot \underline{u}_j).$$

i Proof

Proof. Write u_{ji} for entry i of $\underline{u}_j$. Then

$$\begin{aligned} \underline{x} \cdot \left(\sum_{j=1}^k c_j \underline{u}_j \right) &= \sum_{i=1}^p x_i \sum_{j=1}^k c_j u_{ji} \quad (\text{Definition 8 in Vectors, Definition 9 in Vectors}) \\ &= \sum_{i=1}^p \sum_{j=1}^k x_i c_j u_{ji} \quad (\text{distribute each } x_i \text{ over the inner sum}) \\ &= \sum_{j=1}^k \sum_{i=1}^p x_i c_j u_{ji} \quad (\text{swap the order of the finite sums}) \\ &= \sum_{j=1}^k \sum_{i=1}^p c_j x_i u_{ji} \quad (\text{commute the factors in each product}) \\ &= \sum_{j=1}^k c_j \sum_{i=1}^p x_i u_{ji} \quad (\text{factor } c_j \text{ out of the inner sum}) \\ &= \sum_{j=1}^k c_j (\underline{x} \cdot \underline{u}_j). \quad (\text{Definition 8 in Vectors}) \end{aligned}$$

□

Example 0.17 (Dotting with a combination of three vectors). Let $\underline{x} = (1, 2, 3)$ and take the combination $2(1, 0, 0) - (0, 1, 0) + 4(0, 0, 1) = (2, -1, 4)$. Directly,

$$\begin{aligned} \underline{x} \cdot (2, -1, 4) &= 2 - 2 + 12 \\ &= 12. \end{aligned}$$

By Theorem 0.7, $2 \cdot 1 - 1 \cdot 2 + 4 \cdot 3 = 12$, the same number.

Theorem 0.8 (Orthonormal vectors are linearly independent). *If $q_1, \dots, q_k$ are orthonormal (Definition 23 in Vectors), then they are linearly independent (Definition 1 in Subspaces and Rank).*

i Proof

Proof. Suppose $\sum_{j=1}^k c_j \underline{q}_j = \underline{0}$. Take any $i \in \{1, \dots, k\}$. Then

$$\begin{aligned} 0 &= \underline{q}_i \cdot \underline{0} && \text{(every term of the dot product is 0)} \\ &= \underline{q}_i \cdot \left(\sum_{j=1}^k c_j \underline{q}_j \right) && \text{(substitute the supposed equation)} \\ &= \sum_{j=1}^k c_j (\underline{q}_i \cdot \underline{q}_j) && \text{(Theorem 0.7)} \\ &= c_i. && (\underline{q}_i \cdot \underline{q}_j \text{ is 1 if } j = i \text{ and 0 otherwise}) \end{aligned}$$

So every $c_i = 0$. □

Example 0.18 (Three orthonormal vectors in $\mathbb{R}^3$). Let

$$\underline{q}_1 = (0, 0, 1), \quad \underline{q}_2 = \frac{1}{\sqrt{2}}(1, 1, 0), \quad \underline{q}_3 = \frac{1}{\sqrt{2}}(1, -1, 0).$$

Each has norm 1:

$$\begin{aligned} \|\underline{q}_2\|^2 &= \frac{1}{2}(1 + 1 + 0) \\ &= 1, \end{aligned}$$

and likewise for the others. They are mutually orthogonal: $\underline{q}_1 \cdot \underline{q}_2 = 0$ and $\underline{q}_1 \cdot \underline{q}_3 = 0$, because $\underline{q}_1$ has its only nonzero entry where the others have 0, and

$$\begin{aligned} \underline{q}_2 \cdot \underline{q}_3 &= \frac{1}{2}(1 - 1 + 0) \\ &= 0. \end{aligned}$$

So by Theorem 0.8 they are linearly independent, with no need to solve $c_1 \underline{q}_1 + c_2 \underline{q}_2 + c_3 \underline{q}_3 = \underline{0}$.

The converse fails: $(1, 0)$ and $(1, 1)$ are linearly independent, since $c_1(1, 0) + c_2(1, 1) = (c_1 + c_2, c_2)$ is $\underline{0}$ only if $c_2 = 0$ and then $c_1 = 0$, but they are not orthonormal, since

$$(1, 0) \cdot (1, 1) = 1 \neq 0.$$

Definition 0.10 (Orthonormal basis). An **orthonormal basis** [12] of a subspace $\mathcal{S}$ of $\mathbb{R}^p$ is a basis of $\mathcal{S}$ (Definition 6 in Subspaces and Rank) whose vectors are orthonormal (Definition 23 in Vectors).

Example 0.19 (Orthonormal bases of $\mathbb{R}^3$).

- The indicator vectors $\underline{e}_1, \underline{e}_2, \underline{e}_3$ are orthonormal (2) and a basis of $\mathbb{R}^3$ (Example 16 in Subspaces and Rank).
- The vectors $\underline{q}_1, \underline{q}_2, \underline{q}_3$ of Example 0.18 are orthonormal and linearly independent, so they are a basis of their span (Definition 6 in Subspaces and Rank), which therefore has dimension 3 (Definition 9 in Subspaces and Rank) and is all of $\mathbb{R}^3$ (Theorem 11 in Subspaces and Rank, Example 16 in Subspaces and Rank). So they are an orthonormal basis of $\mathbb{R}^3$ too.

Example 0.20 (Bases that are not orthonormal, and orthonormal lists that are not bases).

- $(1, 0), (1, 1)$ is a basis of $\mathbb{R}^2$ (Example 11 in Subspaces and Rank). It is not an orthonormal basis, since $(1, 0) \cdot (1, 1) = 1$.
- $(1, 0, 0), (0, 1, 0)$ is orthonormal but not a basis of $\mathbb{R}^3$: its span is the plane $z = 0$ (Example 6 in Subspaces and Rank), which does not contain $(0, 0, 1)$.

Theorem 0.9 (Coordinates in an orthonormal basis are dot products). If $\underline{q}_1, \dots, \underline{q}_k$ is an orthonormal basis of a subspace $\mathcal{S}$ (Definition 0.10), then every $\underline{x} \in \mathcal{S}$ satisfies

$$\underline{x} = \sum_{i=1}^k (\underline{q}_i \cdot \underline{x}) \underline{q}_i.$$

i Proof

Proof. The basis spans $\mathcal{S}$, so $\underline{x} = \sum_{j=1}^k c_j \underline{q}_j$ for some numbers c_j . For each i ,

$$\begin{aligned}
\tilde{q}_i \cdot \tilde{x} &= \tilde{q}_i \cdot \left(\sum_{j=1}^k c_j \tilde{q}_j \right) && \text{(substitute)} \\
&= \sum_{j=1}^k c_j (\tilde{q}_i \cdot \tilde{q}_j) && \text{(Theorem 0.7)} \\
&= c_i, && \text{(Definition 23 in Vectors)}
\end{aligned}$$

so each coefficient c_i is $\tilde{q}_i \cdot \tilde{x}$. □

Example 0.21 (Expanding a vector in an orthonormal basis). With the orthonormal basis $\tilde{q}_1, \tilde{q}_2, \tilde{q}_3$ of $\mathbb{R}^3$ from Example 0.19 and $\tilde{x} = (1, 2, 3)$: $\tilde{q}_1 \cdot \tilde{x} = 3$,

$$\begin{aligned}
\tilde{q}_2 \cdot \tilde{x} &= \frac{1}{\sqrt{2}} (1 + 2) \\
&= \frac{3}{\sqrt{2}}
\end{aligned}$$

and

$$\begin{aligned}
\tilde{q}_3 \cdot \tilde{x} &= \frac{1}{\sqrt{2}} (1 - 2) \\
&= -\frac{1}{\sqrt{2}}.
\end{aligned}$$

Then

$$\begin{aligned}
\sum_{i=1}^3 (\tilde{q}_i \cdot \tilde{x}) \tilde{q}_i &= 3(0, 0, 1) + \frac{3}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} (1, 1, 0) - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} (1, -1, 0) && \text{(substitute)} \\
&= (0, 0, 3) + \left(\frac{3}{2}, \frac{3}{2}, 0\right) + \left(-\frac{1}{2}, \frac{1}{2}, 0\right) && \text{(multiply out each term)} \\
&= (1, 2, 3), && \text{(add entrywise)}
\end{aligned}$$

which is $\tilde{x}$, as Theorem 0.9 says.

Definition 0.11 (Gram-Schmidt process). The **Gram-Schmidt process** [13] takes vectors $\tilde{a}_1, \dots, \tilde{a}_k \in \mathbb{R}^p$ and, for $i = 1, 2, \dots, k$ in turn:

1. **Orthogonalize:** $\tilde{q}'_i \stackrel{\text{def}}{=} \tilde{a}_i - \sum_{j=1}^{i-1} (\tilde{q}_j \cdot \tilde{a}_i) \tilde{q}_j$ (for $i = 1$ the sum is empty, so $\tilde{q}'_1 = \tilde{a}_1$);
2. **Test:** if $\tilde{q}'_i = 0$, stop;
3. **Normalize:** $\tilde{q}_i \stackrel{\text{def}}{=} \tilde{q}'_i / \|\tilde{q}'_i\|$.

The output is the list $\tilde{q}_1, \tilde{q}_2, \dots$ produced before the process stops, or all k of them if it never stops.

Example 0.22 (Two steps of Gram-Schmidt in $\mathbb{R}^3$). Let $\underline{a}_1 = (1, 1, 0)$ and $\underline{a}_2 = (1, 0, 1)$.

- **Step 1.** $\underline{q}'_1 = (1, 1, 0)$, with norm $\sqrt{2}$, so $\underline{q}_1 = \frac{1}{\sqrt{2}} (1, 1, 0)$.
- **Step 2.**

$$\begin{aligned}\underline{q}_1 \cdot \underline{a}_2 &= \frac{1}{\sqrt{2}} (1 + 0 + 0) \\ &= \frac{1}{\sqrt{2}},\end{aligned}$$

so

$$\begin{aligned}\underline{q}'_2 &= (1, 0, 1) - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} (1, 1, 0) \quad (\text{orthogonalize}) \\ &= (1, 0, 1) - \left(\frac{1}{2}, \frac{1}{2}, 0\right) \quad (\text{multiply out}) \\ &= \left(\frac{1}{2}, -\frac{1}{2}, 1\right), \quad (\text{subtract entrywise})\end{aligned}$$

with norm $\sqrt{\frac{1}{4} + \frac{1}{4} + 1} = \sqrt{\frac{3}{2}}$, so

$$\begin{aligned}\underline{q}_2 &= \sqrt{\frac{2}{3}} \left(\frac{1}{2}, -\frac{1}{2}, 1\right) \\ &= \frac{1}{\sqrt{6}} (1, -1, 2).\end{aligned}$$

As a check,

$$\begin{aligned}\underline{q}_1 \cdot \underline{q}_2 &= \frac{1}{\sqrt{12}} (1 - 1 + 0) \\ &= 0.\end{aligned}$$

Example 0.23 (Gram-Schmidt stops on dependent vectors). Let $\underline{a}_1 = (1, 2)$ and $\underline{a}_2 = (2, 4)$. Step 1 gives $\underline{q}_1 = \frac{1}{\sqrt{5}} (1, 2)$. In step 2,

$$\begin{aligned}\underline{q}_1 \cdot \underline{a}_2 &= \frac{1}{\sqrt{5}} (2 + 8) \\ &= \frac{10}{\sqrt{5}},\end{aligned}$$

and

$$\begin{aligned}\underline{q}'_2 &= (2, 4) - \frac{10}{\sqrt{5}} \cdot \frac{1}{\sqrt{5}} (1, 2) \\ &= (2, 4) - 2(1, 2) \\ &= \underline{0},\end{aligned}$$

so the process stops: $\underline{a}_2$ is a multiple of $\underline{a}_1$.

Theorem 0.10 (What Gram-Schmidt produces). *Run the Gram-Schmidt process (Definition 0.11) on $\underline{a}_1, \dots, \underline{a}_k \in \mathbb{R}^p$.*

1. *If it has produced $\underline{q}_1, \dots, \underline{q}_i$ without stopping, then $\underline{q}_1, \dots, \underline{q}_i$ are orthonormal and $\text{span}\{\underline{q}_1, \dots, \underline{q}_i\} = \text{span}\{\underline{a}_1, \dots, \underline{a}_i\}$.*
2. *If it reaches step i (that is, it did not stop at steps $1, \dots, i-1$), it stops there exactly when $\underline{a}_i \in \text{span}\{\underline{a}_1, \dots, \underline{a}_{i-1}\}$.*
3. *It completes all k steps exactly when $\underline{a}_1, \dots, \underline{a}_k$ are linearly independent; then $\underline{q}_1, \dots, \underline{q}_k$ is an orthonormal basis of $\text{span}\{\underline{a}_1, \dots, \underline{a}_k\}$ (Definition 0.10).*

i Proof

Proof. Parts 1 and 2, by induction on i . Before step 1, the list of $\underline{q}$'s is empty, which is orthonormal, and both spans are $\{0\}$ (Definition 5 in Subspaces and Rank). Suppose part 1 holds after step $i-1$, and write

$$\begin{aligned}\mathcal{S}_{i-1} &= \text{span}\{\underline{a}_1, \dots, \underline{a}_{i-1}\} \\ &= \text{span}\{\underline{q}_1, \dots, \underline{q}_{i-1}\}.\end{aligned}$$

The orthonormal vectors $\underline{q}_1, \dots, \underline{q}_{i-1}$ are linearly independent (Theorem 0.8) and span $\mathcal{S}_{i-1}$, so they are an orthonormal basis of it.

$\underline{q}'_i$ is orthogonal to the earlier $\underline{q}$'s. For each $l < i$,

$$\begin{aligned}\underline{q}_l \cdot \underline{q}'_i &= \underline{q}_l \cdot \left(1 \underline{a}_i + \sum_{j=1}^{i-1} (-\underline{q}_j \cdot \underline{a}_i) \underline{q}_j \right) && \text{(Definition 0.11, as a linear combination)} \\ &= 1 (\underline{q}_l \cdot \underline{a}_i) + \sum_{j=1}^{i-1} (-\underline{q}_j \cdot \underline{a}_i) (\underline{q}_l \cdot \underline{q}_j) && \text{(Theorem 0.7)} \\ &= \underline{q}_l \cdot \underline{a}_i - \sum_{j=1}^{i-1} (\underline{q}_j \cdot \underline{a}_i) (\underline{q}_l \cdot \underline{q}_j) && (1z = z, \text{ and pull the minus signs out of the sum}) \\ &= \underline{q}_l \cdot \underline{a}_i - \underline{q}_l \cdot \underline{a}_i && \text{(only the } j = l \text{ term survives, Definition 23 in Vectors)} \\ &= 0. && \text{(arithmetic)}\end{aligned}$$

Part 2 at step i . If $\underline{q}'_i = 0$, then $\underline{a}_i = \sum_{j < i} (\underline{q}_j \cdot \underline{a}_i) \underline{q}_j \in \mathcal{S}_{i-1}$. Conversely, if $\underline{a}_i \in \mathcal{S}_{i-1}$, then $\underline{a}_i = \sum_{j < i} (\underline{q}_j \cdot \underline{a}_i) \underline{q}_j$ (Theorem 0.9), so $\underline{q}'_i = 0$.

Part 1 at step i , if the process does not stop. Then $\underline{q}'_i \neq 0$, so $\|\underline{q}'_i\| > 0$, and $\underline{q}_i = \underline{q}'_i / \|\underline{q}'_i\|$ has norm $\|\underline{q}'_i\| / \|\underline{q}'_i\| = 1$ (Theorem 0.1) and for $l < i$

$$\begin{aligned} \underline{q}_l \cdot \underline{q}_i &= \frac{1}{\|\underline{q}'_i\|} (\underline{q}_l \cdot \underline{q}'_i) \quad (\text{Theorem 5 in Direct Sums and Orthogonal Complements, special case } b = 0) \\ &= 0, \quad (\underline{q}'_i \text{ is orthogonal to } \underline{q}_l) \end{aligned}$$

and $\underline{q}_i \cdot \underline{q}_l = 0$ too (Theorem 1 in Vectors). So $\underline{q}_1, \dots, \underline{q}_i$ are orthonormal. For the spans:

- $\underline{a}_i = \|\underline{q}'_i\| \underline{q}_i + \sum_{j < i} (\underline{q}_j \cdot \underline{a}_i) \underline{q}_j$ is in $\text{span}\{\underline{q}_1, \dots, \underline{q}_i\}$, and so is each of $\underline{a}_1, \dots, \underline{a}_{i-1}$, which lie in $\mathcal{S}_{i-1} = \text{span}\{\underline{q}_1, \dots, \underline{q}_{i-1}\} \subseteq \text{span}\{\underline{q}_1, \dots, \underline{q}_i\}$.
- $\underline{q}_i = \left(\underline{a}_i - \sum_{j < i} (\underline{q}_j \cdot \underline{a}_i) \underline{q}_j \right) / \|\underline{q}'_i\|$ is in $\text{span}\{\underline{a}_1, \dots, \underline{a}_i\}$, because each $\underline{q}_j$ with $j < i$ lies in $\mathcal{S}_{i-1}$; and so is each of $\underline{q}_1, \dots, \underline{q}_{i-1}$.

A span is a subspace (Theorem 2 in Subspaces and Rank), so it contains every linear combination of vectors in it; each span therefore contains the other, and they are equal.

Part 3. If the process completes, then $\underline{a}_1 \notin \text{span}$ of the empty list, and for each $i \geq 2$, $\underline{a}_i \notin \text{span}\{\underline{a}_1, \dots, \underline{a}_{i-1}\}$ (part 2). Applying Theorem 8 in Subspaces and Rank k times, starting from the empty list, shows that $\underline{a}_1, \dots, \underline{a}_k$ are linearly independent. Conversely, if they are linearly independent, no $\underline{a}_i$ is a linear combination $\sum_{j < i} c_j \underline{a}_j$ of the earlier ones, because then $\sum_{j < i} c_j \underline{a}_j - \underline{a}_i = \underline{0}$ would be a combination with a nonzero coefficient; so by part 2 the process never stops. When it completes, $\underline{q}_1, \dots, \underline{q}_k$ are orthonormal, hence linearly independent (Theorem 0.8), and $\text{span}\{\underline{a}_1, \dots, \underline{a}_k\} = \text{span}\{\underline{q}_1, \dots, \underline{q}_k\}$ (part 1), so they are an orthonormal basis of it. $\square$

Example 0.24 (Finishing an orthonormal basis of $\mathbb{R}^3$). Continue Example 0.22 with $\underline{a}_3 = (0, 1, 1)$. $\underline{q}_1 \cdot \underline{a}_3 = \frac{1}{\sqrt{2}}$ and

$$\begin{aligned} \underline{q}_2 \cdot \underline{a}_3 &= \frac{1}{\sqrt{6}} (0 - 1 + 2) \\ &= \frac{1}{\sqrt{6}}, \end{aligned}$$

so

$$\begin{aligned} \underline{q}'_3 &= (0, 1, 1) - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} (1, 1, 0) - \frac{1}{\sqrt{6}} \cdot \frac{1}{\sqrt{6}} (1, -1, 2) \quad (\text{orthogonalize}) \\ &= (0, 1, 1) - \frac{1}{2} (1, 1, 0) - \frac{1}{6} (1, -1, 2) \quad (\text{multiply the scalars}) \\ &= \left(0 - \frac{1}{2} - \frac{1}{6}, 1 - \frac{1}{2} + \frac{1}{6}, 1 - 0 - \frac{2}{6}\right) \quad (\text{subtract entrywise}) \\ &= \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right), \quad (\text{arithmetic}) \end{aligned}$$

with norm $\sqrt{3 \cdot \frac{4}{9}} = \frac{2}{\sqrt{3}}$, so $\underline{q}_3 = \frac{1}{\sqrt{3}} (-1, 1, 1)$. The process did not stop, so by Theorem 0.10 $\underline{a}_1, \underline{a}_2, \underline{a}_3$ are linearly independent and $\underline{q}_1, \underline{q}_2, \underline{q}_3$ is an orthonormal basis of their span. That span therefore has dimension 3 (Definition 9 in Subspaces and Rank), so it is $\mathbb{R}^3$ (Theorem 11 in Subspaces and Rank, Example 16 in Subspaces and Rank). As a

check,

$$\begin{aligned} \underline{q}_1 \cdot \underline{q}_3 &= \frac{1}{\sqrt{6}} (-1 + 1 + 0) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned} \underline{q}_2 \cdot \underline{q}_3 &= \frac{1}{\sqrt{18}} (-1 - 1 + 2) \\ &= 0. \end{aligned}$$

Corollary 0.4 (Every subspace has an orthonormal basis). *Let $\mathcal{S}$ be a subspace of $\mathbb{R}^p$. Every orthonormal list of vectors in $\mathcal{S}$ extends to an orthonormal basis of $\mathcal{S}$ (Definition 0.10). Starting from the empty list shows that $\mathcal{S}$ has an orthonormal basis.*

Proof

Proof. Let $\underline{u}_1, \dots, \underline{u}_r \in \mathcal{S}$ be orthonormal. They are linearly independent (Theorem 0.8), so they extend to a basis $\underline{u}_1, \dots, \underline{u}_r, \underline{u}_{r+1}, \dots, \underline{u}_d$ of $\mathcal{S}$ (Theorem 9 in Subspaces and Rank). Run the Gram-Schmidt process on this basis. It is linearly independent, so the process completes, and the output is an orthonormal basis of its span, which is $\mathcal{S}$ (Theorem 0.10, part 3).

The first r outputs are $\underline{u}_1, \dots, \underline{u}_r$ themselves, by induction on $i \leq r$. Suppose the outputs before step i are $\underline{u}_1, \dots, \underline{u}_{i-1}$ (for $i = 1$ there are none, and the sum below is empty). Then

$$\begin{aligned} \underline{q}'_i &= \underline{u}_i - \sum_{j < i} (\underline{u}_j \cdot \underline{u}_i) \underline{u}_j \\ &= \underline{u}_i, \end{aligned}$$

because each $\underline{u}_j \cdot \underline{u}_i = 0$; $\underline{q}'_i = \underline{u}_i \neq 0$, because $\|\underline{u}_i\| = 1$, so the process does not stop; and

$$\begin{aligned} \underline{q}_i &= \underline{u}_i / \|\underline{u}_i\| \\ &= \underline{u}_i. \end{aligned}$$

So the orthonormal basis contains the starting list. □

Example 0.25 (Extending one unit vector to an orthonormal basis of $\mathbb{R}^3$). Start from the unit vector $\underline{u}_1 = \frac{1}{\sqrt{2}} (1, 1, 0)$. Extend it to a basis of $\mathbb{R}^3$ as in Example 18 in Subspaces and Rank, starting from the empty list: $\underline{u}_1$ is not in the span of the empty list, $\{0\}$, because $\|\underline{u}_1\| = 1$; $(1, 0, 0)$ is not in $\text{span}\{\underline{u}_1\}$, because every multiple of $\underline{u}_1$ has equal first and second entries, and $(0, 0, 1)$ is not in $\text{span}\{\underline{u}_1, (1, 0, 0)\}$, because every combination of those two vectors has third entry 0. So $\underline{u}_1, (1, 0, 0), (0, 0, 1)$ are linearly independent

([Theorem 8 in Subspaces and Rank](#)), and three linearly independent vectors in $\mathbb{R}^3$ are a basis of it (they are a basis of their span, which has dimension 3 ([Definition 9 in Subspaces and Rank](#)) and so is $\mathbb{R}^3$ ([Theorem 11 in Subspaces and Rank](#), [Example 16 in Subspaces and Rank](#))). Gram-Schmidt on $\underline{u}_1, (1, 0, 0), (0, 0, 1)$:

1. $\underline{q}_1 = \underline{u}_1$.
2. $\underline{q}_1 \cdot (1, 0, 0) = \frac{1}{\sqrt{2}}$, so

$$\begin{aligned}\underline{q}'_2 &= (1, 0, 0) - \frac{1}{2} (1, 1, 0) \\ &= \left(\frac{1}{2}, -\frac{1}{2}, 0\right),\end{aligned}$$

with norm $\frac{1}{\sqrt{2}}$, and $\underline{q}_2 = \frac{1}{\sqrt{2}} (1, -1, 0)$.

3. $(0, 0, 1)$ is orthogonal to $\underline{q}_1$ and $\underline{q}_2$, so $\underline{q}'_3 = (0, 0, 1)$, which already has norm 1, and $\underline{q}_3 = (0, 0, 1)$.

The result is the orthonormal basis of [Example 0.19](#), in a different order.

Outer product

Definition 0.12 (Outer product). The **outer product** [14] of a vector $\underline{u}$ of length m and a vector $\underline{v}$ of length n is the $m \times n$ matrix product ([Definition 7 in Matrices](#)) $\underline{u} \underline{v}^\top$ of the column vector $\underline{u}$ with the row vector $\underline{v}^\top$. Its entries are

$$\left(\begin{matrix} \underline{u} & \underline{v}^\top \\ m \times 1 & 1 \times n \end{matrix} \right)_{ij} = u_i v_j \quad (\text{definition of matrix multiplication; the sum has one term})$$

Example 0.26 (An outer product). For $\underline{u} = (1, 2, 3)$ and $\underline{v} = (4, 5)$:

$$\begin{aligned}
\underset{\sim}{u} \underset{\sim}{v}^\top &= \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 4 & 5 \end{bmatrix} && \text{(definition of the outer product)} \\
&= \begin{bmatrix} 1 \cdot 4 & 1 \cdot 5 \\ 2 \cdot 4 & 2 \cdot 5 \\ 3 \cdot 4 & 3 \cdot 5 \end{bmatrix} && \text{(entry } (i, j) \text{ is } u_i v_j) \\
&= \begin{bmatrix} 4 & 5 \\ 8 & 10 \\ 12 & 15 \end{bmatrix} && \text{(multiply)}
\end{aligned}$$

i Remark (Outer product and dot product)

Remark 0.6 (Outer product and dot product). The dot product $\underset{\sim}{u}^\top \underset{\sim}{v}$ ([Example 6 in Matrices](#)) needs two vectors of the same length and gives a number. The outer product $\underset{\sim}{u} \underset{\sim}{v}^\top$ takes vectors of any two lengths and gives a matrix [3, Ch. 1], p. 11. For example, $\underset{\sim}{u} = (1, 2, 3)$ and $\underset{\sim}{v} = (4, 5)$ in [Example 0.26](#) have different lengths, so they have no dot product, but their outer product is a 3×2 matrix. For $\underset{\sim}{a} = (1, 2)$ and $\underset{\sim}{b} = (3, 4)$, which have the same length, both products exist:

$$\begin{aligned}
\underset{\sim}{a}^\top \underset{\sim}{b} &= 1 \cdot 3 + 2 \cdot 4 \\
&= 11 \\
\underset{\sim}{a} \underset{\sim}{b}^\top &= \begin{bmatrix} 1 \cdot 3 & 1 \cdot 4 \\ 2 \cdot 3 & 2 \cdot 4 \end{bmatrix} \\
&= \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix}
\end{aligned}$$

Theorem 0.11 (An outer product has rank one). *If $\underset{\sim}{u}$ is a nonzero vector of length m and $\underset{\sim}{v}$ is a nonzero vector of length n , then $\text{rank}(\underset{\sim}{u} \underset{\sim}{v}^\top) = 1$ ([Definition 2 in Subspaces and Rank](#)).*

i Proof

Proof. Column j of $\underset{\sim}{u} \underset{\sim}{v}^\top$ has entries $u_1 v_j, \dots, u_m v_j$, so it is the vector $v_j \underset{\sim}{u}$.

The rank is at least one. Because $\underset{\sim}{v} \neq \underset{\sim}{0}$, some entry v_j is nonzero. Then column j , $v_j \underset{\sim}{u}$, is a nonzero vector, and a single nonzero vector is linearly independent ([Definition 1 in Subspaces and Rank](#)): $c(v_j \underset{\sim}{u}) = \underset{\sim}{0}$ forces $c = 0$.

The rank is at most one. Take any two columns $j \neq k$. If

$$\begin{aligned} v_j &= v_k \\ &= 0, \end{aligned}$$

both columns are $\underline{0}$, and $1 \cdot (v_j \underline{u}) + 1 \cdot (v_k \underline{u}) = \underline{0}$. Otherwise, use the coefficients $c_j = v_k$ and $c_k = -v_j$, which are not both zero:

$$\begin{aligned} v_k (v_j \underline{u}) - v_j (v_k \underline{u}) &= (v_k v_j - v_j v_k) \underline{u} && \text{(collect the scalar multiples of } \underline{u} \text{)} \\ &= 0 \cdot \underline{u} && \text{(multiplication of numbers is commutative)} \\ &= \underline{0} && \text{(a zero multiple of a vector is } \underline{0} \text{)} \end{aligned}$$

Either way, every pair of columns has a combination with coefficients that are not all zero and that equals $\underline{0}$. Now take any set of two or more columns. It contains a pair $j \neq k$. Use that pair's coefficients for columns j and k , and the coefficient 0 for every other column in the set. This combination equals $\underline{0}$, and its coefficients are not all zero, so the set is not linearly independent ([Definition 1 in Subspaces and Rank](#)). The largest linearly independent set of columns therefore has one column. $\square$

Example 0.27 (The rank of an outer product). In [Example 0.26](#), the second column $(5, 10, 15)$ of $\underline{u} \underline{v}^\top$ is $\frac{5}{4}$ times the first column $(4, 8, 12)$, because the columns are $v_1 \underline{u} = 4\underline{u}$ and $v_2 \underline{u} = 5\underline{u}$. So the two columns are not linearly independent, and the 3×2 matrix has rank 1.

- [1] Wikipedia contributors, “Inner product space — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Inner_product_space
- [2] S. Axler, *Linear algebra done right*, 4th ed. in Undergraduate texts in mathematics. Springer, 2024. doi: [10.1007/978-3-031-41026-0](https://doi.org/10.1007/978-3-031-41026-0).
- [3] S. Banerjee and A. Roy, *Linear algebra and matrix analysis for statistics*, vol. 181. Crc Press Boca Raton, 2014. Available: <https://www.routledge.com/Linear-Algebra-and-Matrix-Analysis-for-Statistics/Banerjee-Roy/p/book/9781420095388>
- [4] H. Zhou, “Vectors.” Lecture notes for Biostat 216, Mathematical Methods for Biostatistics, University of California, Los Angeles, 2024. Accessed: Oct. 05, 2026. [Online]. Available: <https://ucla-biostat-216.github.io/2024fall/slides/02-vector/02-vector.html>
- [5] Wikipedia contributors, “Pythagorean theorem — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Pythagorean_theorem
- [6] Wikipedia contributors, “Cauchy–schwarz inequality — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Cauchy%E2%80%93Schwarz_inequality
- [7] Wikipedia contributors, “Norm (mathematics) — Wikipedia, the free encyclopedia.” 2026. Available: [https://en.wikipedia.org/wiki/Norm_\(mathematics\)](https://en.wikipedia.org/wiki/Norm_(mathematics))

- [8] Wikipedia contributors, “Triangle inequality — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Triangle_inequality
- [9] Wikipedia contributors, “Lp space — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Lp_space
- [10] Wikipedia contributors, “Uniform norm — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Uniform_norm
- [11] Wikipedia contributors, “Cosine similarity — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Cosine_similarity
- [12] Wikipedia contributors, “Orthonormal basis — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Orthonormal_basis
- [13] Wikipedia contributors, “Gram–schmidt process — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Gram%E2%80%93Schmidt_process
- [14] Wikipedia contributors, “Outer product — Wikipedia, the free encyclopedia.” 2026. Available: https://en.wikipedia.org/wiki/Outer_product