Definition

Orthogonal vs Orthonormal vectors A collection of vectors $a_1, ..., a_k$ is orthogonal or mutually orthogonal if

$$a_i \perp a_j \qquad i \neq j, \;\; i, j = 1, 2, \ldots, k$$

A collection of vectors $a_1, ..., a_k$ is orthonormal if it is orthogonal and

$$\|a_i\|_2 = 1, \qquad i = 1, \ldots, n.$$

In other words (based on inner product), the collection $a_1, ..., a_k$ is orthonormal if

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Definition

Orthogonal vs Orthonormal vectors A collection of vectors $a_1, ..., a_k$ is orthogonal or mutually orthogonal if

$$a_i \perp a_j \qquad i \neq j, \;\; i,j = 1, 2, \ldots, k$$

A collection of vectors $a_1, ..., a_k$ is orthonormal if it is orthogonal and

$$\|a_i\|_2 = 1, \qquad i = 1, \ldots, n.$$

In other words (based on inner product), the collection $a_1, ..., a_k$ is orthonormal if

$$a_i^T a_j = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- The standard unit n vectors $e_1, \ldots, e_n$ are orthonormal.
- The following vectors are orthonormal

$$\begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \qquad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

Linear independence of orthonormal vectors

Recall: the collection of vectors $a_1, ..., a_k$ are linearly independent if and only if

$$\beta_1 a_1 + \cdots + \beta_k a_k = 0 \quad \rightarrow \quad \beta_1 = \cdots = \beta_k = 0.$$

Linear independence of orthonormal vectors

Recall: the collection of vectors $a_1, ..., a_k$ are linearly independent if and only if

$$\beta_1 a_1 + \cdots + \beta_k a_k = 0 \quad \rightarrow \quad \beta_1 = \cdots = \beta_k = 0.$$

Proof. Suppose $a_1, ..., a_k$ are othonormal vectors, and $\beta_1 a_1 + \cdots + \beta_k a_k = 0$, for some scalars $\beta_i, i = 1, \ldots, k$.

Taking the inner product of this equality with a_i yields

$$\begin{aligned} 0 &= a_i^T(\beta_1 a_1 + \cdots + \beta_i a_i + \cdots + \beta_k a_k) \\ &= \beta_1(a_i^T a_1) + \cdots + \beta_i(a_i^T a_i) + \cdots + \beta_k(a_i^T a_k) \\ &= 0 + \cdots + \beta_i + \cdots + 0 = \beta_i \end{aligned}$$

Since $a_i^T a_j = 0$ for $j \neq i$ and $a_i^T a_i = 1$. Thus, the only linear combination of $a_1, \ldots, a_k$ that is zero is the one with all coefficients zero.

Linear combinations of orthonormal vectors

Theorem

If a vector x is a linear combination of **orthonormal vectors** $a_1, \ldots, a_k$, *i.e.,*

$$x = \beta_1 a_1 + \cdots + \beta_k a_k$$

Then

$$\beta_i = a_i^T x, \quad i = 1, \ldots, n.$$

Proof. Taking the inner product of the left-hand and right-hand sides of this equation with a_i yields

$$a_i^T x = a_i^T(\beta_1 a_1 + \cdots + \beta_k a_k) = \beta_i$$

Since $a_i^T a_j = 0$ for $j \neq i$ and $a_i^T a_i = 1$.

For any x that is a linear combination of orthonormal vectors $a_1, \ldots, a_k$, we have the identity

$$x = (a_1^T x) a_1 + \cdots + (a_k^T x) a_k$$

Orthonormal basis

Definition

If the n-vectors $a_1, \ldots, a_k$ are orthonormal $\rightarrow$ they form a basis, then for any vector $x \in \mathbb{R}^n$ we have

$$x = (a_1^T x)a_1 + \cdots + (a_k^T x)a_k$$

Orthonormal basis

Definition

If the n-vectors $a_1, \ldots, a_k$ are orthonormal $\rightarrow$ they form a basis, then for any vector $x \in \mathbb{R}^n$ we have

$$x = (a_1^T x)a_1 + \cdots + (a_k^T x)a_k$$

Example. $x = [1, 2, 3]^T$ can be expressed as a linear combinations of the orthonormal basis

$$a_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} \qquad a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}.$$

The inner products of x with these vectors are

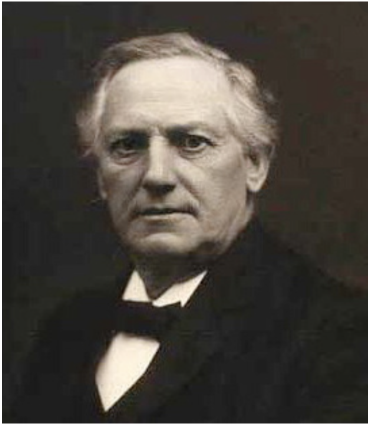
$$a_1^T x = -3 \qquad a_2^T x = \frac{3}{\sqrt{2}}, \qquad a_3^T x = \frac{-1}{\sqrt{2}}.$$

It can be verified that the expansion of x in this basis is

$$x = (-3)a_1 + \frac{3}{\sqrt{2}} a_2 + \frac{-1}{\sqrt{2}} a_3.$$

Gram–Schmidt Algorithm

It is an algorithm that can be used to determine if a list of n-vectors $a_1, \ldots, a_k$ is linearly independent. The algorithm is named after the mathematicians Jørgen Pedersen Gram (left) and Erhard Schmidt (right).



How Gram–Schmidt works!

Given a set of vectors $a_1, \ldots, a_k \in \mathbb{R}^n$.

- If the vectors are linearly independent, the Gram–Schmidt algorithm produces an **orthonormal collection** of vectors $q_1, \ldots, q_k$ with the following property

$$span\{a_1, \ldots, a_i\} = span\{q_1, \ldots, q_i\}$$

- If the vectors $a_1, \ldots, a_{j-1}$ are linearly independent, but $a_1, \ldots, a_j$ are linearly dependent, the algorithm detects this and terminates. In other words, the Gram–Schmidt algorithm finds the **first vector** a_j that is a linear combination of previous vectors $a_1, \ldots, a_{j-1}$.

Algorithm 5.1 GRAM–SCHMIDT ALGORITHM

given n-vectors $a_1, \ldots, a_k$

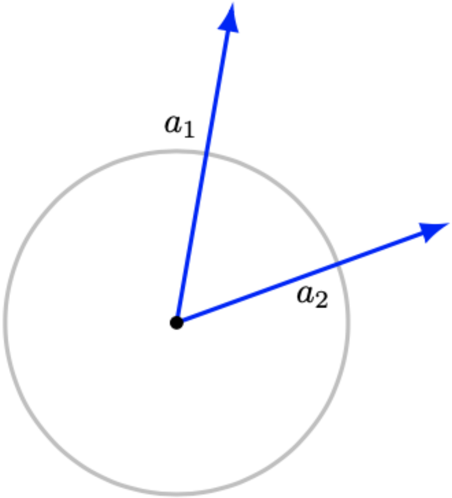
for $i = 1, \ldots, k$,

1. *Orthogonalization.* $\tilde{q}_i = a_i - (q_1^T a_i)q_1 - \cdots - (q_{i-1}^T a_i)q_{i-1}$
2. *Test for linear dependence.* if $\tilde{q}_i = 0$, quit.
3. *Normalization.* $q_i = \tilde{q}_i / \|\tilde{q}_i\|$

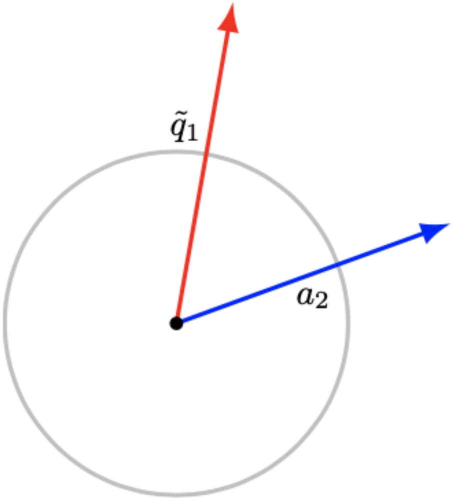
- $\tilde{q}_1 = a_1 \quad \text{if } \tilde{q}_1 \neq 0 \quad \rightarrow \qquad q_1 = \frac{\tilde{q}_1}{\|\tilde{q}_1\|}$
- $\tilde{q}_2 = a_2 - (q_1^T a_2)q_1 \quad \text{if } \tilde{q}_2 \neq 0 \quad \rightarrow \qquad q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$
- $\tilde{q}_3 = a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2 \quad \text{if } \tilde{q}_3 \neq 0 \quad \rightarrow \qquad q_3 = \frac{\tilde{q}_3}{\|\tilde{q}_3\|}$

  $\vdots$
- $\tilde{q}_k = a_k - (q_1^T a_k)q_1 - \cdots - (q_{k-1}^T a_k)q_{k-1} \text{ if } \tilde{q}_k \neq 0 \quad \rightarrow \qquad q_k = \frac{\tilde{q}_k}{\|\tilde{q}_k\|}$

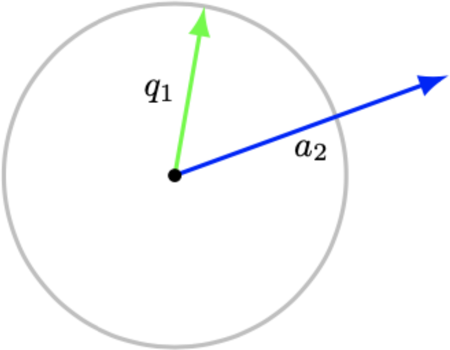
Gram–Schmidt illustration for two vectors in $\mathbb{R}^2$



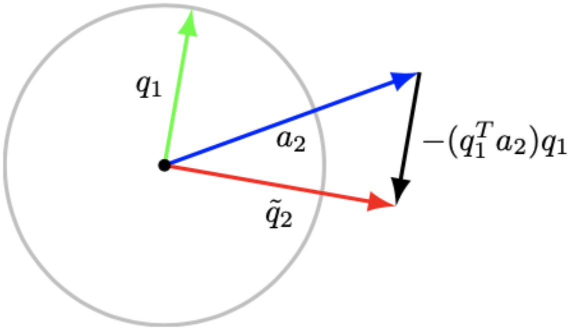
Gram–Schmidt illustration for two vectors in $\mathbb{R}^2$



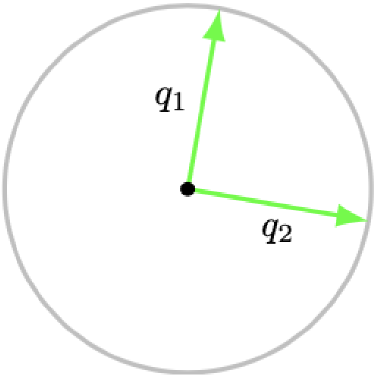
Gram–Schmidt illustration for two vectors in $\mathbb{R}^2$



Gram–Schmidt illustration for two vectors in $\mathbb{R}^2$



Gram–Schmidt illustration for two vectors in $\mathbb{R}^2$



Example

We define three vectors:

$$a_1 = (-1, 1, -1, 1)^T, \quad a_2 = (-1, 3, -1, 3)^T, \quad a_3 = (1, 3, 5, 7)^T$$

Applying the Gram–Schmidt algorithm gives the following results.

- $i = 1$. We have $\tilde{q}_1 = a_1$, and $\|\tilde{q}_1\| = \sqrt{4} = 2$, so

$$q_1 = \frac{1}{\|\tilde{q}_1\|}\tilde{q}_1 = (-1/2, 1/2, -1/2, 1/2)^T.$$

  which is simply a_1 normalized.

- $i = 2$. We have $q_1^T a_2 = 4$, so

$$\tilde{q}_2 = a_2 - (q_1^T a_2)q_1 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 3 \end{bmatrix} - 4 \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

  which is indeed orthogonal to q_1 (and a_1). It has norm $\|\tilde{q}_2\| = 2$; normalizing it gives

$$q_2 = \frac{1}{\|\tilde{q}_2\|}\tilde{q}_2 = (1/2, 1/2, 1/2, 1/2)^T.$$

- $i = 3$. We have $q_1^T a_3 = 2$ and $q_2^T a_3 = 8$, so

$$\begin{aligned}
\tilde{q}_3 &= a_3 - (q_1^T a_3) q_1 - (q_2^T a_3) q_2 \\
&= \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} - 2 \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} - 8 \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} \\
&= \begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \end{bmatrix},
\end{aligned}$$

which is indeed orthogonal to q_1 and q_2 (and a_1 and a_2). It has norm $\|\tilde{q}_3\| = 4$; normalizing it gives

$$q_3 = \frac{1}{\|\tilde{q}_3\|} \tilde{q}_3 = (-1/2, -1/2, 1/2, 1/2)^T.$$

Completion of the Gram–Schmidt algorithm without early termination tells us that the vectors a_1, a_2, a_3 are linearly independent.