Orthogonal Bases (cont.)

D. H. Luecking

MASC

5 April 2024

An example of the Cauchy-Schwarz Inequality:

An example of the Cauchy-Schwarz Inequality:

Let's consider the following setup:
The vector space will be $\mathbb{R}^3$
The inner product will be $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + 2x_2 y_2 + 3x_3 y_3$.

An example of the Cauchy-Schwarz Inequality:

Let's consider the following setup:
The vector space will be $\mathbb{R}^3$
The inner product will be $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3$.

(Exercise: verify that this is an inner product.)

An example of the Cauchy-Schwarz Inequality:

Let's consider the following setup:
The vector space will be $\mathbb{R}^3$
The inner product will be $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3$.
(Exercise: verify that this is an inner product.)

Consider the two vectors $\mathbf{x} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$.

An example of the Cauchy-Schwarz Inequality:

Let's consider the following setup:
The vector space will be $\mathbb{R}^3$
The inner product will be $\langle \mathbf{x}, \mathbf{y} \rangle = x_1y_1 + 2x_2y_2 + 3x_3y_3$.

(Exercise: verify that this is an inner product.)

Consider the two vectors $\mathbf{x} = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} -1 \\ 2 \\ 2 \end{pmatrix}$. Then

$$\langle \mathbf{x}, \mathbf{y} \rangle = 3 \cdot (-1) + 2(2 \cdot 2) + 3(1 \cdot 2) = 11$$
$$\langle \mathbf{x}, \mathbf{x} \rangle = 3^2 + 2(2)^2 + 3(1)^2 = 20$$
$$\langle \mathbf{y}, \mathbf{y} \rangle = (-1)^2 + 2(2)^2 + 3(2)^2 = 21$$

Then the Cauchy-Schwarz Inequality predicts $11 \leq \sqrt{20}\sqrt{21} \approx 20.5$.

If we let $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix}$, then this inner product is the same as $\langle \mathbf{x}, \mathbf{y} \rangle = (A\mathbf{x})^T A\mathbf{y}$.

If we let $A = \left(\begin{array}{ccc} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{array} \right)$, then this inner product is the same as $\langle \mathbf{x}, \mathbf{y} \rangle = (A\mathbf{x})^T A\mathbf{y}$.

The triangle inequality for the same inner product and the same vectors:

$$\|\mathbf{x} + \mathbf{y}\| = \sqrt{2^2 + 2(4)^2 + 3(3)^2} = \sqrt{63} \approx 7.94$$
$$\|\mathbf{x}\| = \sqrt{3^2 + 2(2)^2 + 3(1)^2} = \sqrt{20} \approx 4.47$$
$$\|\mathbf{y}\| = \sqrt{(-1)^2 + 2(2)^2 + 3(2^2)} = \sqrt{21} \approx 4.58$$

If we let $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{2} & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix}$, then this inner product is the same as $\langle \mathbf{x}, \mathbf{y} \rangle = (A\mathbf{x})^T A\mathbf{y}$.

The triangle inequality for the same inner product and the same vectors:

$$\|\mathbf{x}+\mathbf{y}\| = \sqrt{2^2 + 2(4)^2 + 3(3)^2} = \sqrt{63} \approx 7.94$$
$$\|\mathbf{x}\| = \sqrt{3^2 + 2(2)^2 + 3(1)^2} = \sqrt{20} \approx 4.47$$
$$\|\mathbf{y}\| = \sqrt{(-1)^2 + 2(2)^2 + 3(2^2)} = \sqrt{21} \approx 4.58$$

and $7.94 \leq 4.47 + 4.58$.

If we use the scalar product for our inner product, we get

$$\mathbf{x}^T\mathbf{y} = 3 \cdot (-1) + 2 \cdot 2 + 1 \cdot 2 = 3$$
$$\mathbf{x}^T\mathbf{x} = 3^2 + (2)^2 + (1)^2 = 14$$
$$\mathbf{y}^T\mathbf{y} = (-1)^2 + (2)^2 + (2)^2 = 9$$

If we use the scalar product for our inner product, we get

$$\mathbf{x}^T\mathbf{y} = 3 \cdot (-1) + 2 \cdot 2 + 1 \cdot 2 = 3$$
$$\mathbf{x}^T\mathbf{x} = 3^2 + (2)^2 + (1)^2 = 14$$
$$\mathbf{y}^T\mathbf{y} = (-1)^2 + (2)^2 + (2)^2 = 9$$

and then the Cauchy-Schwarz Inequality predicts $3 \leq \sqrt{14}\sqrt{9} \approx 11.2$.

If we use the scalar product for our inner product, we get

$$\mathbf{x}^T\mathbf{y} = 3 \cdot (-1) + 2 \cdot 2 + 1 \cdot 2 = 3$$
$$\mathbf{x}^T\mathbf{x} = 3^2 + (2)^2 + (1)^2 = 14$$
$$\mathbf{y}^T\mathbf{y} = (-1)^2 + (2)^2 + (2)^2 = 9$$

and then the Cauchy-Schwarz Inequality predicts $3 \leq \sqrt{14}\sqrt{9} \approx 11.2$. The triangle inequality predicts $\|\mathbf{x}+\mathbf{y}\| = \sqrt{29} \leq \sqrt{14} + \sqrt{9}$ or, approximately, $5.39 \leq 6.74$.

If we use the scalar product for our inner product, we get

$$\begin{aligned} \mathbf{x}^T\mathbf{y} &= 3\cdot(-1) + 2\cdot 2 + 1\cdot 2 = 3 \\ \mathbf{x}^T\mathbf{x} &= 3^2 + (2)^2 + (1)^2 = 14 \\ \mathbf{y}^T\mathbf{y} &= (-1)^2 + (2)^2 + (2)^2 = 9 \end{aligned}$$

and then the Cauchy-Schwarz Inequality predicts $3 \le \sqrt{14}\sqrt{9} \approx 11.2$. The triangle inequality predicts $\|\mathbf{x}+\mathbf{y}\| = \sqrt{29} \le \sqrt{14} + \sqrt{9}$ or, approximately, $5.39 \le 6.74$.

Here is a third example of an inner product for $\mathcal{P}_3$. Define

$$\langle p, q\rangle = p(0)q(0) + p(1)q(1) + p(-1)q(-1).$$

If we use the scalar product for our inner product, we get

$$\mathbf{x}^T\mathbf{y} = 3 \cdot (-1) + 2 \cdot 2 + 1 \cdot 2 = 3$$
$$\mathbf{x}^T\mathbf{x} = 3^2 + (2)^2 + (1)^2 = 14$$
$$\mathbf{y}^T\mathbf{y} = (-1)^2 + (2)^2 + (2)^2 = 9$$

and then the Cauchy-Schwarz Inequality predicts $3 \leq \sqrt{14}\sqrt{9} \approx 11.2$. The triangle inequality predicts $\|\mathbf{x} + \mathbf{y}\| = \sqrt{29} \leq \sqrt{14} + \sqrt{9}$ or, approximately, $5.39 \leq 6.74$.

Here is a third example of an inner product for $\mathcal{P}_3$. Define

$$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(-1)q(-1).$$

This is an inner product. I will leave the verification of the 2nd and 3rd condition as exercises.

If we use the scalar product for our inner product, we get

$$\mathbf{x}^T\mathbf{y} = 3 \cdot (-1) + 2 \cdot 2 + 1 \cdot 2 = 3$$
$$\mathbf{x}^T\mathbf{x} = 3^2 + (2)^2 + (1)^2 = 14$$
$$\mathbf{y}^T\mathbf{y} = (-1)^2 + (2)^2 + (2)^2 = 9$$

and then the Cauchy-Schwarz Inequality predicts $3 \leq \sqrt{14}\sqrt{9} \approx 11.2$. The triangle inequality predicts $\|\mathbf{x}+\mathbf{y}\| = \sqrt{29} \leq \sqrt{14} + \sqrt{9}$ or, approximately, $5.39 \leq 6.74$.

Here is a third example of an inner product for $\mathcal{P}_3$. Define

$$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(-1)q(-1).$$

This is an inner product. I will leave the verification of the 2nd and 3rd condition as exercises. For the first condition consider

$$\langle p, p \rangle = p(0)^2 + p(1)^2 + p(-1)^2 \geq 0$$

If we use the scalar product for our inner product, we get

$$\mathbf{x}^T\mathbf{y} = 3 \cdot (-1) + 2 \cdot 2 + 1 \cdot 2 = 3$$
$$\mathbf{x}^T\mathbf{x} = 3^2 + (2)^2 + (1)^2 = 14$$
$$\mathbf{y}^T\mathbf{y} = (-1)^2 + (2)^2 + (2)^2 = 9$$

and then the Cauchy-Schwarz Inequality predicts $3 \leq \sqrt{14}\sqrt{9} \approx 11.2$. The triangle inequality predicts $\|\mathbf{x} + \mathbf{y}\| = \sqrt{29} \leq \sqrt{14} + \sqrt{9}$ or, approximately, $5.39 \leq 6.74$.

Here is a third example of an inner product for $\mathcal{P}_3$. Define

$$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(-1)q(-1).$$

This is an inner product. I will leave the verification of the 2nd and 3rd condition as exercises. For the first condition consider

$$\langle p, p \rangle = p(0)^2 + p(1)^2 + p(-1)^2 \geq 0$$

It can only be zero if $p(x)$ is zero at 3 points.

If we use the scalar product for our inner product, we get

$$\mathbf{x}^T\mathbf{y} = 3 \cdot (-1) + 2 \cdot 2 + 1 \cdot 2 = 3$$
$$\mathbf{x}^T\mathbf{x} = 3^2 + (2)^2 + (1)^2 = 14$$
$$\mathbf{y}^T\mathbf{y} = (-1)^2 + (2)^2 + (2)^2 = 9$$

and then the Cauchy-Schwarz Inequality predicts $3 \leq \sqrt{14}\sqrt{9} \approx 11.2$. The triangle inequality predicts $\|\mathbf{x} + \mathbf{y}\| = \sqrt{29} \leq \sqrt{14} + \sqrt{9}$ or, approximately, $5.39 \leq 6.74$.

Here is a third example of an inner product for $\mathcal{P}_3$. Define

$$\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(-1)q(-1).$$

This is an inner product. I will leave the verification of the 2nd and 3rd condition as exercises. For the first condition consider

$$\langle p, p \rangle = p(0)^2 + p(1)^2 + p(-1)^2 \geq 0$$

It can only be zero if $p(x)$ is zero at 3 points. But if p is not the zero function, $p(x) = 0$ can't have 3 roots since its degree is less than 3.

Some examples:

Some examples:

If $p(x) = 1 + x + x^2$ and $q(x) = 1 - x - x^2$, then

Some examples:

If $p(x) = 1 + x + x^2$ and $q(x) = 1 - x - x^2$, then
$\langle p, p \rangle = 1^2 + 3^2 + 1^2 = 11$,

Some examples:

If $p(x) = 1 + x + x^2$ and $q(x) = 1 - x - x^2$, then
$\langle p, p \rangle = 1^2 + 3^2 + 1^2 = 11$,
$\langle q, q \rangle = 1^2 + (-1)^2 + 1^2 = 3$

Some examples:

If $p(x) = 1 + x + x^2$ and $q(x) = 1 - x - x^2$, then
$\langle p, p \rangle = 1^2 + 3^2 + 1^2 = 11$,
$\langle q, q \rangle = 1^2 + (-1)^2 + 1^2 = 3$ and
$\langle p, q \rangle = 1(1) + 3(-1) + 1(1) = -1$.

Some examples:

If $p(x) = 1 + x + x^2$ and $q(x) = 1 - x - x^2$, then
$\langle p, p \rangle = 1^2 + 3^2 + 1^2 = 11$,
$\langle q, q \rangle = 1^2 + (-1)^2 + 1^2 = 3$ and
$\langle p, q \rangle = 1(1) + 3(-1) + 1(1) = -1$.

Exercise: calculate $\|p + q\|$ for this inner product.

Some examples:

If $p(x) = 1 + x + x^2$ and $q(x) = 1 - x - x^2$, then
$\langle p, p \rangle = 1^2 + 3^2 + 1^2 = 11$,
$\langle q, q \rangle = 1^2 + (-1)^2 + 1^2 = 3$ and
$\langle p, q \rangle = 1(1) + 3(-1) + 1(1) = -1$.

Exercise: calculate $\|p + q\|$ for this inner product.

Let V be an inner product space. Recall that $\mathbf{x} \perp \mathbf{y}$ means $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

Definition

Two subspaces S and T of V are orthogonal if $\mathbf{x} \perp \mathbf{y}$ for every $\mathbf{x}$ in S and $\mathbf{y}$ in T.

Some examples:

If $p(x) = 1 + x + x^2$ and $q(x) = 1 - x - x^2$, then
$\langle p, p \rangle = 1^2 + 3^2 + 1^2 = 11$,
$\langle q, q \rangle = 1^2 + (-1)^2 + 1^2 = 3$ and
$\langle p, q \rangle = 1(1) + 3(-1) + 1(1) = -1$.

Exercise: calculate $\|p + q\|$ for this inner product.

Let V be an inner product space. Recall that $\mathbf{x} \perp \mathbf{y}$ means $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.

Definition

Two subspaces S and T of V are orthogonal if $\mathbf{x} \perp \mathbf{y}$ for every $\mathbf{x}$ in S and $\mathbf{y}$ in T.
The orthogonal complement of S is the set of vectors $\mathbf{x}$ in V such that $\mathbf{x} \perp \mathbf{y}$ for every vector $\mathbf{y}$ in S.

Just as in $\mathbb{R}^n$ with the scalar product, we have

1. If S and T are orthogonal and $\mathbf{x}$ belongs to both, then $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ and so $\mathbf{x} = \mathbf{0}$

Just as in $\mathbb{R}^n$ with the scalar product, we have

1. If S and T are orthogonal and $\mathbf{x}$ belongs to both, then $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ and so $\mathbf{x} = \mathbf{0}$
2. $S^\perp$ is a subspace of V: if $\mathbf{u}$ and $\mathbf{v}$ belong to $S^\perp$ and $\mathbf{x} \in S$, then $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{x} \rangle = \alpha \langle \mathbf{u}, \mathbf{x} \rangle + \beta \langle \mathbf{v}, \mathbf{x} \rangle = 0 + 0$. Thus $\alpha\mathbf{u} + \beta\mathbf{v} \in S^\perp$.

Just as in $\mathbb{R}^n$ with the scalar product, we have

1. If S and T are orthogonal and $\mathbf{x}$ belongs to both, then $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ and so $\mathbf{x} = \mathbf{0}$
2. $S^\perp$ is a subspace of V: if $\mathbf{u}$ and $\mathbf{v}$ belong to $S^\perp$ and $\mathbf{x} \in S$, then $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{x} \rangle = \alpha \langle \mathbf{u}, \mathbf{x} \rangle + \beta \langle \mathbf{v}, \mathbf{x} \rangle = 0 + 0$. Thus $\alpha\mathbf{u} + \beta\mathbf{v} \in S^\perp$.
3. $\dim S + \dim S^\perp = \dim V$.

Just as in $\mathbb{R}^n$ with the scalar product, we have

1. If S and T are orthogonal and $\mathbf{x}$ belongs to both, then $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ and so $\mathbf{x} = \mathbf{0}$
2. $S^\perp$ is a subspace of V: if $\mathbf{u}$ and $\mathbf{v}$ belong to $S^\perp$ and $\mathbf{x} \in S$, then $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{x} \rangle = \alpha \langle \mathbf{u}, \mathbf{x} \rangle + \beta \langle \mathbf{v}, \mathbf{x} \rangle = 0 + 0$. Thus $\alpha\mathbf{u} + \beta\mathbf{v} \in S^\perp$.
3. $\dim S + \dim S^\perp = \dim V$.
4. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a basis for S and $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is a basis for $S^\perp$ where $n = \dim V$ then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V.

Just as in $\mathbb{R}^n$ with the scalar product, we have

1. If S and T are orthogonal and $\mathbf{x}$ belongs to both, then $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ and so $\mathbf{x} = \mathbf{0}$
2. $S^\perp$ is a subspace of V: if $\mathbf{u}$ and $\mathbf{v}$ belong to $S^\perp$ and $\mathbf{x} \in S$, then $\langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{x} \rangle = \alpha \langle \mathbf{u}, \mathbf{x} \rangle + \beta \langle \mathbf{v}, \mathbf{x} \rangle = 0 + 0$. Thus $\alpha\mathbf{u} + \beta\mathbf{v} \in S^\perp$.
3. $\dim S + \dim S^\perp = \dim V$.
4. If $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a basis for S and $\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n\}$ is a basis for $S^\perp$ where $n = \dim V$ then $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ is a basis for V.
5. If $\mathbf{x}$ is any vector in V then there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{x} = \mathbf{u} + \mathbf{v}$. The vector $\mathbf{u}$ is the closest vector in S to $\mathbf{x}$. That is, $\|\mathbf{x} - \mathbf{u}\| \leq \|\mathbf{x} - \mathbf{y}\|$ for all $\mathbf{y} \in S$.

Definition

For vectors $\mathbf{x}, \mathbf{y}$ in V, $\mathbf{y} \neq \mathbf{0}$, the *scalar projection of $\mathbf{x}$ onto $\mathbf{y}$* is

$$\alpha = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|}$$

Definition

For vectors $\mathbf{x}, \mathbf{y}$ in V, $\mathbf{y} \neq \mathbf{0}$, the *scalar projection of $\mathbf{x}$ onto $\mathbf{y}$* is

$$\alpha = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|}$$

and the *vector projection of $\mathbf{x}$ onto $\mathbf{y}$* is

$$\mathbf{p} = \alpha \frac{1}{\|\mathbf{y}\|} \mathbf{y} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y}$$

Definition

For vectors $\mathbf{x}, \mathbf{y}$ in V, $\mathbf{y} \neq \mathbf{0}$, the *scalar projection of $\mathbf{x}$ onto $\mathbf{y}$* is

$$\alpha = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{y}\|}$$

and the *vector projection of $\mathbf{x}$ onto $\mathbf{y}$* is

$$\mathbf{p} = \alpha \frac{1}{\|\mathbf{y}\|} \mathbf{y} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y}$$

Just as in $\mathbb{R}^3$, $\mathbf{x} - \mathbf{p}$ is orthogonal to $\mathbf{y}$:

$$\begin{aligned}
\langle \mathbf{y}, \mathbf{x} - \mathbf{p} \rangle &= \langle \mathbf{y}, \mathbf{x} \rangle - \langle \mathbf{y}, \mathbf{p} \rangle \\
&= \langle \mathbf{y}, \mathbf{x} \rangle - \left\langle \mathbf{y}, \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y} \right\rangle \\
&= \langle \mathbf{y}, \mathbf{x} \rangle - \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \langle \mathbf{y}, \mathbf{y} \rangle \\
&= \langle \mathbf{y}, \mathbf{x} \rangle - \langle \mathbf{x}, \mathbf{y} \rangle = 0
\end{aligned}$$

Definition

A set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is said to be *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for each $i \neq j$. If it is orthogonal and $\|\mathbf{v}_j\| = 1$ for each j, we call it *orthonormal*.

Just as in $\mathbb{R}^n$ we have

1. Every finite dimensional subset of V has an orthonomal basis.

Definition

A set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is said to be *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for each $i \neq j$. If it is orthogonal and $\|\mathbf{v}_j\| = 1$ for each j, we call it *orthonormal*.

Just as in $\mathbb{R}^n$ we have

1. Every finite dimensional subset of V has an orthonomal basis.
2. If S is a subspace of V with an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ and $\mathbf{x}$ is any vector in V. Then the closest vector in S to $\mathbf{x}$ is

$$\mathbf{p} = \sum_{j=1}^{r} \frac{\langle \mathbf{x}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \mathbf{v}_j$$

   and $\mathbf{x} - \mathbf{p}$ is orthogonal to S.

Definition

A set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is said to be *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for each $i \neq j$. If it is orthogonal and $\|\mathbf{v}_j\| = 1$ for each j, we call it *orthonormal*.

Just as in $\mathbb{R}^n$ we have

1. Every finite dimensional subset of V has an orthonomal basis.
2. If S is a subspace of V with an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ and $\mathbf{x}$ is any vector in V. Then the closest vector in S to $\mathbf{x}$ is

$$\mathbf{p} = \sum_{j=1}^{r} \frac{\langle \mathbf{x}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \mathbf{v}_j$$

   and $\mathbf{x} - \mathbf{p}$ is orthogonal to S.
3. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S then $\mathbf{p} = \sum_{j=1}^{r} \langle \mathbf{x}, \mathbf{u}_j \rangle \mathbf{u}_j$

Definition

A set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is said to be *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for each $i \neq j$. If it is orthogonal and $\|\mathbf{v}_j\| = 1$ for each j, we call it *orthonormal*.

Just as in $\mathbb{R}^n$ we have

1. Every finite dimensional subset of V has an orthonomal basis.
2. If S is a subspace of V with an orthogonal basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ and $\mathbf{x}$ is any vector in V. Then the closest vector in S to $\mathbf{x}$ is

$$\mathbf{p} = \sum_{j=1}^{r} \frac{\langle \mathbf{x}, \mathbf{v}_j \rangle}{\langle \mathbf{v}_j, \mathbf{v}_j \rangle} \mathbf{v}_j$$

   and $\mathbf{x} - \mathbf{p}$ is orthogonal to S.
3. If $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S then $\mathbf{p} = \sum_{j=1}^{r} \langle \mathbf{x}, \mathbf{u}_j \rangle \mathbf{u}_j$
4. Every orthogonal set of nonzero vectors is independent.

Matrices with orthogonal columns

In this discussion, all vectors will be in $\mathbb{R}^n$ for some n. Orthogonality will mean $\mathbf{x}^T\mathbf{y} = 0$.

Matrices with orthogonal columns

In this discussion, all vectors will be in $\mathbb{R}^n$ for some n. Orthogonality will mean $\mathbf{x}^T\mathbf{y} = 0$.

If A is an $n \times k$ matrix with orthonormal columns, then $k \leq n$ because the columns are independent so there can't be more than n of them.

Matrices with orthogonal columns

In this discussion, all vectors will be in $\mathbb{R}^n$ for some n. Orthogonality will mean $\mathbf{x}^T\mathbf{y} = 0$.

If A is an $n \times k$ matrix with orthonormal columns, then $k \leq n$ because the columns are independent so there can't be more than n of them. Also, the rank of A is k and the nullity is 0.

Matrices with orthogonal columns

In this discussion, all vectors will be in $\mathbb{R}^n$ for some n. Orthogonality will mean $\mathbf{x}^T\mathbf{y} = 0$.

If A is an $n \times k$ matrix with orthonormal columns, then $k \leq n$ because the columns are independent so there can't be more than n of them. Also, the rank of A is k and the nullity is 0.

If the columns of A are $\mathbf{a}_j$ then the rows of A^T are $\mathbf{a}_j^T$ and we get the product $A^TA = (\mathbf{a}_i^T\mathbf{a}_j)_{k\times k}$.

Matrices with orthogonal columns

In this discussion, all vectors will be in $\mathbb{R}^n$ for some n. Orthogonality will mean $\mathbf{x}^T\mathbf{y} = 0$.

If A is an $n \times k$ matrix with orthonormal columns, then $k \leq n$ because the columns are independent so there can't be more than n of them. Also, the rank of A is k and the nullity is 0.

If the columns of A are $\mathbf{a}_j$ then the rows of A^T are $\mathbf{a}_j^T$ and we get the product $A^TA = (\mathbf{a}_i^T\mathbf{a}_j)_{k\times k}$. Because the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is orthonormal we have

$$\mathbf{a}_i^T\mathbf{a}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

This means that $A^TA = I$, the $k \times k$ identity matrix.

Matrices with orthogonal columns

In this discussion, all vectors will be in $\mathbb{R}^n$ for some n. Orthogonality will mean $\mathbf{x}^T\mathbf{y} = 0$.

If A is an $n \times k$ matrix with orthonormal columns, then $k \leq n$ because the columns are independent so there can't be more than n of them. Also, the rank of A is k and the nullity is 0.

If the columns of A are $\mathbf{a}_j$ then the rows of A^T are $\mathbf{a}_j^T$ and we get the product $A^TA = (\mathbf{a}_i^T\mathbf{a}_j)_{k\times k}$. Because the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is orthonormal we have

$$\mathbf{a}_i^T\mathbf{a}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

This means that $A^TA = I$, the $k \times k$ identity matrix.

If it happens that $n = k$ so that A is a square matrix, then this tells us that A is invertible and $A^{-1} = A^T$.

Matrices with orthogonal columns

In this discussion, all vectors will be in $\mathbb{R}^n$ for some n. Orthogonality will mean $\mathbf{x}^T\mathbf{y} = 0$.

If A is an $n \times k$ matrix with orthonormal columns, then $k \leq n$ because the columns are independent so there can't be more than n of them. Also, the rank of A is k and the nullity is 0.

If the columns of A are $\mathbf{a}_j$ then the rows of A^T are $\mathbf{a}_j^T$ and we get the product $A^T A = (\mathbf{a}_i^T \mathbf{a}_j)_{k \times k}$. Because the set $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ is orthonormal we have

$$\mathbf{a}_i^T \mathbf{a}_j = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

This means that $A^T A = I$, the $k \times k$ identity matrix.

If it happens that $n = k$ so that A is a square matrix, then this tells us that A is invertible and $A^{-1} = A^T$. In this case only, we also have $AA^T = I$ and this tells us that A^T also has orthonormal columns.

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns.

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns. It is not enough to have only orthgonal columns.

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns. It is not enough to have only orthgonal columns. The reason for the terminology is the following.

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns. It is not enough to have only orthgonal columns. The reason for the terminology is the following.

Theorem

A square matrix Q is orthogonal if and only if it satisfies both the following:

1. *for every vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \perp \mathbf{y}$ implies $Q\mathbf{x} \perp Q\mathbf{y}$.*

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns. It is not enough to have only orthgonal columns. The reason for the terminology is the following.

Theorem

A square matrix Q is orthogonal if and only if it satisfies both the following:

1. *for every vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \perp \mathbf{y}$ implies $Q\mathbf{x} \perp Q\mathbf{y}$.*
2. *for every vector $\mathbf{x}$, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.*

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns. It is not enough to have only orthgonal columns. The reason for the terminology is the following.

Theorem

A square matrix Q is orthogonal if and only if it satisfies both the following:

1. *for every vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \perp \mathbf{y}$ implies $Q\mathbf{x} \perp Q\mathbf{y}$.*
2. *for every vector $\mathbf{x}$, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.*

If Q is an orthogonal matrix then $(Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T \mathbf{y}$.

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns. It is not enough to have only orthgonal columns. The reason for the terminology is the following.

Theorem

A square matrix Q is orthogonal if and only if it satisfies both the following:

1. *for every vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \perp \mathbf{y}$ implies $Q\mathbf{x} \perp Q\mathbf{y}$.*
2. *for every vector $\mathbf{x}$, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.*

If Q is an orthogonal matrix then $(Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T\mathbf{y}$. Applying this with $\mathbf{x} \perp \mathbf{y}$ we see that $Q\mathbf{x} \perp Q\mathbf{y}$.

Definition

An *orthogonal matrix* is a square matrix with orthonormal columns. Equivalently, an orthogonal matrix A is a square matrix satisfying $A^{-1} = A^T$.

Note that an ortho**gon**al matrix has ortho**norm**al columns. It is not enough to have only orthgonal columns. The reason for the terminology is the following.

Theorem

A square matrix Q is orthogonal if and only if it satisfies both the following:

1. *for every vectors $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x} \perp \mathbf{y}$ implies $Q\mathbf{x} \perp Q\mathbf{y}$.*
2. *for every vector $\mathbf{x}$, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.*

If Q is an orthogonal matrix then $(Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T\mathbf{y}$. Applying this with $\mathbf{x} \perp \mathbf{y}$ we see that $Q\mathbf{x} \perp Q\mathbf{y}$. Applying this with $\mathbf{y} = \mathbf{x}$ we see that $\|Q\mathbf{x}\|^2 = \|\mathbf{x}\|^2$.

Now suppose we have a matrix Q that transforms orthogonal vectors to orthogonal vectors and preserves norms.

Now suppose we have a matrix Q that transforms orthogonal vectors to orthogonal vectors and preserves norms. Then $\{Q\mathbf{e}_1, \ldots, Q\mathbf{e}_n\}$ must be orthonormal, and these are the columns of Q.

Now suppose we have a matrix Q that transforms orthogonal vectors to orthogonal vectors and preserves norms. Then $\{Q\mathbf{e}_1, \ldots, Q\mathbf{e}_n\}$ must be orthonormal, and these are the columns of Q.

It should be pointed out that the product of two (or more) orthogonal matrices is an orthogonal matrices. This can be deduced from the property $Q^{-1} = Q^T$ or from the above preservation properties.

Now suppose we have a matrix Q that transforms orthogonal vectors to orthogonal vectors and preserves norms. Then $\{Q\mathbf{e}_1, \ldots, Q\mathbf{e}_n\}$ must be orthonormal, and these are the columns of Q.

It should be pointed out that the product of two (or more) orthogonal matrices is an orthogonal matrices. This can be deduced from the property $Q^{-1} = Q^T$ or from the above preservation properties.

Note that because Q also preserves the scalar product, it preserves all angles between vectors (in contexts where angles make sense):

Now suppose we have a matrix Q that transforms orthogonal vectors to orthogonal vectors and preserves norms. Then $\{Q\mathbf{e}_1, \ldots, Q\mathbf{e}_n\}$ must be orthonormal, and these are the columns of Q.

It should be pointed out that the product of two (or more) orthogonal matrices is an orthogonal matrices. This can be deduced from the property $Q^{-1} = Q^T$ or from the above preservation properties.

Note that because Q also preserves the scalar product, it preserves all angles between vectors (in contexts where angles make sense): If θ is the angle between $Q\mathbf{x}$ and $Q\mathbf{y}$ then

$$\cos\theta = \frac{\langle Q\mathbf{x}, Q\mathbf{y}\rangle}{\|Q\mathbf{x}\|\,\|Q\mathbf{y}\|} = \frac{\langle \mathbf{x}, \mathbf{y}\rangle}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$

and that is the cosine of the angle between $\mathbf{x}$ and $\mathbf{y}$.

If $\mathcal{B}$ is an orthonormal basis, and Q is the matrix whose columns are the elements of $\mathcal{B}$, then Q is the transition matrix from $\mathcal{B}$ to $\mathcal{E}$ and Q^T is the transition matrix from $\mathcal{E}$ to $\mathcal{B}$.

If $\mathcal{B}$ is an orthonormal basis, and Q is the matrix whose columns are the elements of $\mathcal{B}$, then Q is the transition matrix from $\mathcal{B}$ to $\mathcal{E}$ and Q^T is the transition matrix from $\mathcal{E}$ to $\mathcal{B}$.

The following can be used instead of a transition matrix, and it works for any inner product space:

If $\mathcal{B}$ is an orthonormal basis, and Q is the matrix whose columns are the elements of $\mathcal{B}$, then Q is the transition matrix from $\mathcal{B}$ to $\mathcal{E}$ and Q^T is the transition matrix from $\mathcal{E}$ to $\mathcal{B}$.

The following can be used instead of a transition matrix, and it works for any inner product space:

Theorem

If $\mathcal{B} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ is an orthonormal basis in an inner product space V and $\mathbf{v}$ is any vector in V, then

$$\mathbf{v} = \langle \mathbf{v}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{v}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \cdots \langle \mathbf{v}, \mathbf{u}_n \rangle \mathbf{u}_n$$

If $\mathcal{B}$ is an orthonormal basis, and Q is the matrix whose columns are the elements of $\mathcal{B}$, then Q is the transition matrix from $\mathcal{B}$ to $\mathcal{E}$ and Q^T is the transition matrix from $\mathcal{E}$ to $\mathcal{B}$.

The following can be used instead of a transition matrix, and it works for any inner product space:

Theorem

If $\mathcal{B} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ is an orthonormal basis in an inner product space V and $\mathbf{v}$ is any vector in V, then

$$\mathbf{v} = \langle \mathbf{v}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{v}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \cdots \langle \mathbf{v}, \mathbf{u}_n \rangle \mathbf{u}_n$$

As we have seen before, if $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$ then

$$\langle \mathbf{v}, \mathbf{u}_j \rangle = c_1 \langle \mathbf{u}_1, \mathbf{u}_j \rangle + \cdots + c_n \langle \mathbf{u}_n, \mathbf{u}_j \rangle = c_j$$