Review of Chapter 5, part 1

Daniel H. Luecking

19 March 2024

Definition (Scalar product)

If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ then the *scalar product of* $\mathbf{x}$ *and* $\mathbf{y}$ is $\mathbf{x}^T\mathbf{y}$

Definition (Scalar product)

If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ then the *scalar product of* $\mathbf{x}$ *and* $\mathbf{y}$ is $\mathbf{x}^T\mathbf{y}$

In terms of the coordinates, if $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, then

$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \mathbf{y}^T\mathbf{x}$.

Definition (Scalar product)

If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ then the *scalar product of* $\mathbf{x}$ *and* $\mathbf{y}$ is $\mathbf{x}^T\mathbf{y}$

In terms of the coordinates, if $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, then

$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \mathbf{y}^T\mathbf{x}$.

Definition (Norm)

We write $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2}$ and call this the *norm of* $\mathbf{x}$.

Definition (Scalar product)

If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ then the *scalar product of* $\mathbf{x}$ *and* $\mathbf{y}$ is $\mathbf{x}^T\mathbf{y}$

In terms of the coordinates, if $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, then

$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \mathbf{y}^T\mathbf{x}$.

Definition (Norm)

We write $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2}$ and call this the *norm of* $\mathbf{x}$.

In terms of coordinates $\|\mathbf{x}\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$.

Definition (Scalar product)

If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ then the *scalar product of* $\mathbf{x}$ *and* $\mathbf{y}$ is $\mathbf{x}^T\mathbf{y}$

In terms of the coordinates, if $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, then

$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \mathbf{y}^T\mathbf{x}$.

Definition (Norm)

We write $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2}$ and call this the *norm of* $\mathbf{x}$.

In terms of coordinates $\|\mathbf{x}\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$.

1. The scalar product of $\alpha\mathbf{x}$ and $\beta\mathbf{y}$ is $(\alpha\mathbf{x})^T(\beta\mathbf{y}) = \alpha\beta(\mathbf{x}^T\mathbf{y})$

Definition (Scalar product)

If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ then the *scalar product of* $\mathbf{x}$ *and* $\mathbf{y}$ is $\mathbf{x}^T\mathbf{y}$

In terms of the coordinates, if $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, then

$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots + x_ny_n = \mathbf{y}^T\mathbf{x}$.

Definition (Norm)

We write $\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2}$ and call this the *norm of* $\mathbf{x}$.

In terms of coordinates $\|\mathbf{x}\| = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$.

1. The scalar product of $\alpha\mathbf{x}$ and $\beta\mathbf{y}$ is $(\alpha\mathbf{x})^T(\beta\mathbf{y}) = \alpha\beta(\mathbf{x}^T\mathbf{y})$
2. The norm of $\alpha\mathbf{x}$ is $\|\alpha\mathbf{x}\| = |\alpha|\,\|\mathbf{x}\|$. Note that if $\alpha = 1/\|\mathbf{x}\|$ the $\|\alpha\mathbf{x}\| = 1$.

Theorem (Angle between two vectors)

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ *or* $\mathbb{R}^3$. *Suppose* $\|\mathbf{x}\| \neq 0$ *and* $\|\mathbf{y}\| \neq 0$. *Let* θ *be the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *with* $0 \leq \theta \leq 180°$. *Then*

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\| \, \|\mathbf{y}\|}.$$

Theorem (Angle between two vectors)

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ *or* $\mathbb{R}^3$. *Suppose* $\|\mathbf{x}\| \neq 0$ *and* $\|\mathbf{y}\| \neq 0$. *Let* θ *be the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *with* $0 \leq \theta \leq 180°$. *Then*

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

Theorem (The Cauchy-Schwarz Inequality)

If $\mathbf{x}$ *and* $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$, *then* $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|$.

Theorem (Angle between two vectors)

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ *or* $\mathbb{R}^3$*. Suppose* $\|\mathbf{x}\| \neq 0$ *and* $\|\mathbf{y}\| \neq 0$*. Let* θ *be the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *with* $0 \leq \theta \leq 180°$*. Then*

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

Theorem (The Cauchy-Schwarz Inequality)

If $\mathbf{x}$ *and* $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$*, then* $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|$*.*

Definition (Orthogonal)

If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$, we say that $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$ if $\mathbf{x}^T\mathbf{y} = 0$. We denote this by writing $\mathbf{x} \perp \mathbf{y}$.

Theorem (Angle between two vectors)

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ *or* $\mathbb{R}^3$. *Suppose* $\|\mathbf{x}\| \neq 0$ *and* $\|\mathbf{y}\| \neq 0$. *Let* θ *be the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *with* $0 \leq \theta \leq 180°$. *Then*

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

Theorem (The Cauchy-Schwarz Inequality)

If $\mathbf{x}$ *and* $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$, *then* $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|$.

Definition (Orthogonal)

If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$, we say that $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$ if $\mathbf{x}^T\mathbf{y} = 0$. We denote this by writing $\mathbf{x} \perp \mathbf{y}$.

1. If $\mathbf{x} \in \mathbb{R}^n$ then the set of vectors orthogonal to $\mathbf{x}$ is a subspace of $\mathbb{R}^n$.

Theorem (Angle between two vectors)

Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^2$ *or* $\mathbb{R}^3$*. Suppose* $\|\mathbf{x}\| \neq 0$ *and* $\|\mathbf{y}\| \neq 0$*. Let* θ *be the angle between* $\mathbf{x}$ *and* $\mathbf{y}$ *with* $0 \leq \theta \leq 180°$*. Then*

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}.$$

Theorem (The Cauchy-Schwarz Inequality)

If $\mathbf{x}$ *and* $\mathbf{y}$ *are vectors in* $\mathbb{R}^n$*, then* $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\,\|\mathbf{y}\|$*.*

Definition (Orthogonal)

If $\mathbf{x}$ and $\mathbf{y}$ are in $\mathbb{R}^n$, we say that $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$ if $\mathbf{x}^T\mathbf{y} = 0$. We denote this by writing $\mathbf{x} \perp \mathbf{y}$.

1. If $\mathbf{x} \in \mathbb{R}^n$ then the set of vectors orthogonal to $\mathbf{x}$ is a subspace of $\mathbb{R}^n$.
2. $\mathbf{x} \perp \mathbf{x}$ if and only if $\mathbf{x} = \mathbf{0}$.

Example: find a nonzero vector orthogonal to both $\mathbf{a}_1 = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$ and $\mathbf{a}_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$.

Example: find a nonzero vector orthogonal to both $\mathbf{a}_1 = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$ and $\mathbf{a}_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$. Solution: solve the system $\mathbf{a}_1^T \mathbf{x} = 0$, $\mathbf{a}_2^T \mathbf{x} = 0$:

Example: find a nonzero vector orthogonal to both $\mathbf{a}_1 = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$ and $\mathbf{a}_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$. Solution: solve the system $\mathbf{a}_1^T \mathbf{x} = 0$, $\mathbf{a}_2^T \mathbf{x} = 0$: This system is equivalent to $A\mathbf{x} = \mathbf{0}$ where $A = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \end{pmatrix}$. We solve that by row-reducing A:

$$\begin{pmatrix} 1 & -1 & 3 \\ -1 & 2 & -1 \end{pmatrix} \xrightarrow{\text{2EROs}} \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 2 \end{pmatrix}$$

Example: find a nonzero vector orthogonal to both $\mathbf{a}_1 = \begin{pmatrix} 1 \\ -1 \\ 3 \end{pmatrix}$ and $\mathbf{a}_2 = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$. Solution: solve the system $\mathbf{a}_1^T\mathbf{x} = 0$, $\mathbf{a}_2^T\mathbf{x} = 0$: This system is equivalent to $A\mathbf{x} = \mathbf{0}$ where $A = \begin{pmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \end{pmatrix}$. We solve that by row-reducing A:

$$\begin{pmatrix} 1 & -1 & 3 \\ -1 & 2 & -1 \end{pmatrix} \xrightarrow{\text{2EROs}} \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 2 \end{pmatrix}$$

Thus, $x_1 = -5x_3$ and $x_2 = -2x_3$ and so the vectors $\begin{pmatrix} -5\alpha \\ -2\alpha \\ \alpha \end{pmatrix}$ are orthogonal to $\mathbf{a}_1$ and $\mathbf{a}_2$, for any choice of α.

Definition (Scalar and vector projections)

If $\mathbf{x}$ and $\mathbf{y}$ belong to $\mathbb{R}^n$ then:

The number $\alpha = \dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$ is called the *scalar projection of $\mathbf{x}$ onto $\mathbf{y}$*.

The vector $\mathbf{p} = \dfrac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$ is called the *vector projection of $\mathbf{x}$ onto $\mathbf{y}$*

Definition (Scalar and vector projections)

If $\mathbf{x}$ and $\mathbf{y}$ belong to $\mathbb{R}^n$ then:

The number $\alpha = \dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$ is called the *scalar projection of $\mathbf{x}$ onto $\mathbf{y}$*.

The vector $\mathbf{p} = \dfrac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$ is called the *vector projection of $\mathbf{x}$ onto $\mathbf{y}$*

Definition (Orthogonal subspaces)

If V and W are subspaces of $\mathbb{R}^n$ then we say they are *orthogonal subspaces* if every vector in V is orthogonal to every vector in W.

Definition (Scalar and vector projections)

If $\mathbf{x}$ and $\mathbf{y}$ belong to $\mathbb{R}^n$ then:

The number $\alpha = \dfrac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}$ is called the *scalar projection of* $\mathbf{x}$ *onto* $\mathbf{y}$.

The vector $\mathbf{p} = \dfrac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$ is called the *vector projection of* $\mathbf{x}$ *onto* $\mathbf{y}$

Definition (Orthogonal subspaces)

If V and W are subspaces of $\mathbb{R}^n$ then we say they are *orthogonal subspaces* if every vector in V is orthogonal to every vector in W.

Definition (Orthogonal complement)

If S is a subspace of $\mathbb{R}^n$ then the *orthogonal complement of* S is the set of all vectors that are orthogonal to every vector in S. We denote this set $S^\perp$. Formally:

$$S^\perp = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x}^T\mathbf{y} = 0 \text{ for all } \mathbf{y} \in S\}.$$

Properties of orthogonal subspaces

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.
2. For any subspace S, $S^\perp$ is also a subspace.

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.
2. For any subspace S, $S^\perp$ is also a subspace.

We have seen that any product like $A\mathbf{x}$ is a linear combination of the columns of A.

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.
2. For any subspace S, $S^\perp$ is also a subspace.

We have seen that any product like $A\mathbf{x}$ is a linear combination of the columns of A. If A and $n \times k$ matrix, we define

$$\mathcal{R}(A) = \{\mathbf{b} \in \mathbb{R}^n \mid \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^k\}$$

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.
2. For any subspace S, $S^{\perp}$ is also a subspace.

We have seen that any product like $A\mathbf{x}$ is a linear combination of the columns of A. If A and $n \times k$ matrix, we define

$$\mathcal{R}(A) = \{\mathbf{b} \in \mathbb{R}^n \mid \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^k\}$$

This is the column space of A.

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.
2. For any subspace S, $S^\perp$ is also a subspace.

We have seen that any product like $A\mathbf{x}$ is a linear combination of the columns of A. If A and $n \times k$ matrix, we define

$$\mathcal{R}(A) = \{\mathbf{b} \in \mathbb{R}^n \mid \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^k\}$$

This is the column space of A.

What is $\mathcal{R}(A)^\perp$?

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.
2. For any subspace S, $S^\perp$ is also a subspace.

We have seen that any product like $A\mathbf{x}$ is a linear combination of the columns of A. If A and $n \times k$ matrix, we define

$$\mathcal{R}(A) = \{\mathbf{b} \in \mathbb{R}^n \mid \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^k\}$$

This is the column space of A.

What is $\mathcal{R}(A)^\perp$?

The solution space of a homogeneous $k \times n$ system of linear equations is a subspace of $\mathbb{R}^n$. It is also the nullspace $\mathcal{N}(A)$ if A is the system matrix.

Properties of orthogonal subspaces

1. If V and W are orthogonal subspaces of $\mathbb{R}^n$ then the only vector that belongs to both V and W is $\mathbf{0}$.
2. For any subspace S, $S^\perp$ is also a subspace.

We have seen that any product like $A\mathbf{x}$ is a linear combination of the columns of A. If A and $n \times k$ matrix, we define

$$\mathcal{R}(A) = \{\mathbf{b} \in \mathbb{R}^n \mid \mathbf{b} = A\mathbf{x} \text{ for some } \mathbf{x} \in \mathbb{R}^k\}$$

This is the column space of A.

What is $\mathcal{R}(A)^\perp$?

The solution space of a homogeneous $k \times n$ system of linear equations is a subspace of $\mathbb{R}^n$. It is also the nullspace $\mathcal{N}(A)$ if A is the system matrix.

What is $\mathcal{N}(A)^\perp$?

Theorem (Orthogonal complements of matrix spaces)

$\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ *and* $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$.

Theorem (Orthogonal complements of matrix spaces)

$\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ *and* $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$.

Theorem (Dimension sum)

If S *is a subspace of* $\mathbb{R}^n$ *then* $\dim S + \dim S^\perp = n$. *Furthermore, if* $\{\mathbf{x}_1, \ldots, \mathbf{x}_r\}$ *is a basis for* S *and* $\{\mathbf{x}_{r+1}, \ldots, \mathbf{x}_n\}$ *is a basis for* $S^\perp$*, then* $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ *is a basis for* $\mathbb{R}^n$.

Theorem (Orthogonal decomposition)

If S *is a subspace of* $\mathbb{R}^n$ *then every vector* $\mathbf{x}$ *in* $\mathbb{R}^n$ *can be written uniquely as a sum* $\mathbf{x} = \mathbf{u} + \mathbf{v}$ *with* $\mathbf{u} \in S$ *and* $\mathbf{v} \in S^\perp$.

Theorem (The double $\perp$)

If S *is a subspace of* $\mathbb{R}^n$ *then* $(S^\perp)^\perp = S$.

Application: for which vectors $\mathbf{b}$ does the equation $A\mathbf{x} = \mathbf{b}$ have a solution?

Application: for which vectors $\mathbf{b}$ does the equation $A\mathbf{x} = \mathbf{b}$ have a solution? The existence of $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$ means that $\mathbf{b}$ is in the column space of A.

Application: for which vectors $\mathbf{b}$ does the equation $A\mathbf{x} = \mathbf{b}$ have a solution? The existence of $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$ means that $\mathbf{b}$ is in the column space of A. By the above, $\mathbf{b}$ must be orthogonal to the null space of A^T.

Application: for which vectors $\mathbf{b}$ does the equation $A\mathbf{x} = \mathbf{b}$ have a solution? The existence of $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$ means that $\mathbf{b}$ is in the column space of A. By the above, $\mathbf{b}$ must be orthogonal to the null space of A^T. If we find a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for the null space of A^T, then the condition for $A\mathbf{x} = \mathbf{b}$ to have a solution is that $\mathbf{v}_1^T\mathbf{b} = 0, \ldots, \mathbf{v}_k^T\mathbf{b} = 0$.

Example: For what vectors $\mathbf{b}$ does the following have a solution?

$$\begin{aligned} 2x_1 + 2x_2 + 4x_3 &= b_1 \\ x_1 + + x_3 &= b_2 \\ x_2 + x_3 &= b_3 \end{aligned}$$

Application: for which vectors $\mathbf{b}$ does the equation $A\mathbf{x} = \mathbf{b}$ have a solution? The existence of $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$ means that $\mathbf{b}$ is in the column space of A. By the above, $\mathbf{b}$ must be orthogonal to the null space of A^T. If we find a basis $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ for the null space of A^T, then the condition for $A\mathbf{x} = \mathbf{b}$ to have a solution is that $\mathbf{v}_1^T \mathbf{b} = 0, \dots, \mathbf{v}_k^T \mathbf{b} = 0$.

Example: For what vectors $\mathbf{b}$ does the following have a solution?

$$\begin{aligned} 2x_1 + 2x_2 + 4x_3 &= b_1 \\ x_1 + \quad\;\; + \; x_3 &= b_2 \\ x_2 + \; x_3 &= b_3 \end{aligned}$$

Take the system matrix A, transpose it, and find the null space of A^T

Application: for which vectors $\mathbf{b}$ does the equation $A\mathbf{x} = \mathbf{b}$ have a solution? The existence of $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$ means that $\mathbf{b}$ is in the column space of A. By the above, $\mathbf{b}$ must be orthogonal to the null space of A^T. If we find a basis $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ for the null space of A^T, then the condition for $A\mathbf{x} = \mathbf{b}$ to have a solution is that $\mathbf{v}_1^T\mathbf{b} = 0, \ldots, \mathbf{v}_k^T\mathbf{b} = 0$.

Example: For what vectors $\mathbf{b}$ does the following have a solution?

$$\begin{aligned} 2x_1 + 2x_2 + 4x_3 &= b_1 \\ x_1 + \qquad + \ x_3 &= b_2 \\ x_2 + \ x_3 &= b_3 \end{aligned}$$

Take the system matrix A, transpose it, and find the null space of A^T

$$\begin{pmatrix} 2 & 1 & 0 \\ 2 & 0 & 1 \\ 4 & 1 & 1 \end{pmatrix} \xrightarrow{\text{6 EROs}} \begin{pmatrix} 1 & 0 & 1/2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{i.e.} \quad \begin{cases} x_1 + (1/2)x_3 = 0 \\ \qquad x_2 - x_3 = 0 \end{cases}$$

Then $\mathcal{N}(A^T)$ is spanned by one vector $\begin{pmatrix} -1/2 \\ 1 \\ 1 \end{pmatrix}$.

Then $\mathcal{N}(A^T)$ is spanned by one vector $\begin{pmatrix} -1/2 \\ 1 \\ 1 \end{pmatrix}$. So, we conclude that a solution exists if and only if $\mathbf{b}$ is orthogonal to this:

$$\begin{pmatrix} -1/2 & 1 & 1 \end{pmatrix} \mathbf{b} = -(1/2)b_1 + b_2 + b_3 = 0$$

Then $\mathcal{N}(A^T)$ is spanned by one vector $\begin{pmatrix} -1/2 \\ 1 \\ 1 \end{pmatrix}$. So, we conclude that a solution exists if and only if $\mathbf{b}$ is orthogonal to this:

$$\begin{pmatrix} -1/2 & 1 & 1 \end{pmatrix} \mathbf{b} = -(1/2)b_1 + b_2 + b_3 = 0$$

Another application:

If we have an inconsistent equation $A\mathbf{x} = \mathbf{b}$ (i.e., one which has no solution), we can multiply it by A^T and we get $A^T A\mathbf{x} = A^T\mathbf{b}$ ...

Then $\mathcal{N}(A^T)$ is spanned by one vector $\begin{pmatrix} -1/2 \\ 1 \\ 1 \end{pmatrix}$. So, we conclude that a solution exists if and only if $\mathbf{b}$ is orthogonal to this:

$$\begin{pmatrix} -1/2 & 1 & 1 \end{pmatrix} \mathbf{b} = -(1/2)b_1 + b_2 + b_3 = 0$$

Another application:

If we have an inconsistent equation $A\mathbf{x} = \mathbf{b}$ (i.e., one which has no solution), we can multiply it by A^T and we get $A^T A\mathbf{x} = A^T\mathbf{b}$ ... which must have a solution.

Then $\mathcal{N}(A^T)$ is spanned by one vector $\begin{pmatrix} -1/2 \\ 1 \\ 1 \end{pmatrix}$. So, we conclude that a solution exists if and only if $\mathbf{b}$ is orthogonal to this:

$$\begin{pmatrix} -1/2 & 1 & 1 \end{pmatrix} \mathbf{b} = -(1/2)b_1 + b_2 + b_3 = 0$$

Another application:

If we have an inconsistent equation $A\mathbf{x} = \mathbf{b}$ (i.e., one which has no solution), we can multiply it by A^T and we get $A^T A\mathbf{x} = A^T\mathbf{b}$ . . . which must have a solution. That solution $\hat{\mathbf{x}}$ has the property that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the column space of A and has the smallest norm among all $\mathbf{x}$ in $\mathbb{R}^n$.

Then $\mathcal{N}(A^T)$ is spanned by one vector $\begin{pmatrix} -1/2 \\ 1 \\ 1 \end{pmatrix}$. So, we conclude that a solution exists if and only if $\mathbf{b}$ is orthogonal to this:

$$\begin{pmatrix} -1/2 & 1 & 1 \end{pmatrix} \mathbf{b} = -(1/2)b_1 + b_2 + b_3 = 0$$

Another application:

If we have an inconsistent equation $A\mathbf{x} = \mathbf{b}$ (i.e., one which has no solution), we can multiply it by A^T and we get $A^T A\mathbf{x} = A^T\mathbf{b}$ ... which must have a solution. That solution $\hat{\mathbf{x}}$ has the property that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the column space of A and has the smallest norm among all $\mathbf{x}$ in $\mathbb{R}^n$.

This vector $\hat{\mathbf{x}}$ is called the *least squares solution* to $A\mathbf{x} = \mathbf{b}$.

An example:

An example:

Find the least squares solution of the following system:

$$\begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

An example:

Find the least squares solution of the following system:

$$\begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

Multiplying by the transpose gives

$$\begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

An example:

Find the least squares solution of the following system:

$$\begin{pmatrix} 1 & 1 \\ -2 & 3 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

Multiplying by the transpose gives

$$\begin{pmatrix} 9 & -7 \\ -7 & 11 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

This has a solution $\hat{\mathbf{x}} = \begin{pmatrix} 83/50 \\ 71/50 \end{pmatrix} = \begin{pmatrix} 1.66 \\ 1.42 \end{pmatrix}$.

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3, 1, 2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3, 1, 2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

In any problem $A\mathbf{x} = \mathbf{b}$, the difference $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is called the *residual vector* assosiated to $\mathbf{x}$.

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3, 1, 2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

In any problem $A\mathbf{x} = \mathbf{b}$, the difference $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is called the *residual vector* assosiated to $\mathbf{x}$. The least squares solution is a vector $\hat{\mathbf{x}}$ that gives the residual vector the smallest possible norm.

Using the method of least squares to get vectors orthogonal to a subspace.

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3, 1, 2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

In any problem $A\mathbf{x} = \mathbf{b}$, the difference $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is called the *residual vector* assosiated to $\mathbf{x}$. The least squares solution is a vector $\hat{\mathbf{x}}$ that gives the residual vector the smallest possible norm.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

If we want to see how close we have come, we find

$$A\hat{\mathbf{x}} = \begin{pmatrix} 3.08 \\ 0.94 \\ 1.9 \end{pmatrix}$$

these values differ from $(3,1,2)^T$ by $(-0.08, 0.06, 0.1)^T$ which has norm $\sqrt{0.02} \approx 0.1414$

In any problem $A\mathbf{x} = \mathbf{b}$, the difference $r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$ is called the *residual vector* assosiated to $\mathbf{x}$. The least squares solution is a vector $\hat{\mathbf{x}}$ that gives the residual vector the smallest possible norm.

Using the method of least squares to get vectors orthogonal to a subspace.

If S is a subspace of $\mathbb{R}^n$ and $\mathbf{b}$ is a vector in $\mathbb{R}^n$, we know that there exist unique vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$. The method of least squares allows us to find them.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$)

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$) has the following property: $A\hat{\mathbf{x}}$ is the closest element of $\mathcal{R}(A)$ to $\mathbf{b}$ and the residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathcal{R}(A)$. That is $\mathbf{u} = A\hat{\mathbf{x}}$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}}$.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$) has the following property: $A\hat{\mathbf{x}}$ is the closest element of $\mathcal{R}(A)$ to $\mathbf{b}$ and the residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathcal{R}(A)$. That is $\mathbf{u} = A\hat{\mathbf{x}}$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}}$.

But every subspace S of $\mathbb{R}^n$ is the column space of some matrix: take any basis of S (or any set of vectors whose span is S) and make them the columns of a matrix A. Then $S = \mathcal{R}(A)$.

Example: Let S be the span of $(1, 1, 2, 0)^T$ and $(0, 1, 2, -2)^T$ and let $\mathbf{b} = (1, 1, 1, 1)^T$. Find the vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^{\perp}$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

If S is $\mathcal{R}(A)$ for some matrix then, by what we have seen, the least squares solution $\hat{\mathbf{x}}$ of $A\mathbf{x} = \mathbf{b}$ (which satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$) has the following property: $A\hat{\mathbf{x}}$ is the closest element of $\mathcal{R}(A)$ to $\mathbf{b}$ and the residual vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to $\mathcal{R}(A)$. That is $\mathbf{u} = A\hat{\mathbf{x}}$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}}$.

But every subspace S of $\mathbb{R}^n$ is the column space of some matrix: take any basis of S (or any set of vectors whose span is S) and make them the columns of a matrix A. Then $S = \mathcal{R}(A)$.

Example: Let S be the span of $(1, 1, 2, 0)^T$ and $(0, 1, 2, -2)^T$ and let $\mathbf{b} = (1, 1, 1, 1)^T$. Find the vectors $\mathbf{u} \in S$ and $\mathbf{v} \in S^\perp$ such that $\mathbf{b} = \mathbf{u} + \mathbf{v}$.

Now S is the column space of the 4×2 matrix A below and we need the least squares solution of

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 2 & 2 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$.

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then
$\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then $\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$ and $\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}} = (-2/29, 12/29, -5/29, 1/29)$.

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then
$\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$ and
$\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}} = (-2/29, 12/29, -5/29, 1/29)$.

Definition (Orthogonal/orthonormal set)

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is called *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for every $i \neq j$.

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then
$\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$ and
$\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}} = (-2/29, 12/29, -5/29, 1/29)$.

Definition (Orthogonal/orthonormal set)

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is called *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for every $i \neq j$.
It is called *orthonormal* if it is orthogonal and also $\|\mathbf{v}_j\| = 1$ for every j.

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then
$\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$ and
$\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}} = (-2/29, 12/29, -5/29, 1/29)$.

Definition (Orthogonal/orthonormal set)

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is called *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for every $i \neq j$.
It is called *orthonormal* if it is orthogonal and also $\|\mathbf{v}_j\| = 1$ for every j.

If a set of nonzero vectors is orthogonal, we can make it orthonormal simply by multiplying each $\mathbf{v}_j$ by $(1/\|\mathbf{v}_j\|)$.

Multiplying by A^T we get

$$\begin{pmatrix} 6 & 5 \\ 5 & 9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \end{pmatrix}$$

which has solution $\hat{\mathbf{x}} = (31/29, -14/29)^T$. Then
$\mathbf{u} = A\hat{\mathbf{x}} = (31/29, 17/29, 34/29, 28/29)^T$ and
$\mathbf{v} = \mathbf{b} - A\hat{\mathbf{x}} = (-2/29, 12/29, -5/29, 1/29)$.

Definition (Orthogonal/orthonormal set)

A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ is called *orthogonal* if $\mathbf{v}_i \perp \mathbf{v}_j$ for every $i \neq j$.
It is called *orthonormal* if it is orthogonal and also $\|\mathbf{v}_j\| = 1$ for every j.

If a set of nonzero vectors is orthogonal, we can make it orthonormal simply by multiplying each $\mathbf{v}_j$ by $(1/\|\mathbf{v}_j\|)$.

Theorem (Orthogonal $\Rightarrow$ independent)

If a set of nonzero vectors is orthogonal, then it is independent.

Theorem (Orthogonal bases)

Every nonzero subspace of $\mathbb{R}^n$ *has a basis that is orthonormal.*

Theorem (Orthogonal bases)

Every nonzero subspace of $\mathbb{R}^n$ has a basis that is orthonormal.

Theorem (Formula for the projection)

Let S is a subspace of $\mathbb{R}^n$ and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be an orthogonal basis for S. If $\mathbf{v}$ is any vector in $\mathbb{R}^n$ and $\mathbf{p}$ the element of S that is closest to $\mathbf{v}$, then

$$\mathbf{p} = \sum_{j=1}^{r} \frac{\mathbf{v}^T \mathbf{v}_j}{\mathbf{v}_j^T \mathbf{v}_j} \mathbf{v}_j$$

Theorem (Orthogonal bases)

Every nonzero subspace of $\mathbb{R}^n$ has a basis that is orthonormal.

Theorem (Formula for the projection)

Let S is a subspace of $\mathbb{R}^n$ and let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ be an orthogonal basis for S. If $\mathbf{v}$ is any vector in $\mathbb{R}^n$ and $\mathbf{p}$ the element of S that is closest to $\mathbf{v}$, then

$$\mathbf{p} = \sum_{j=1}^{r} \frac{\mathbf{v}^T \mathbf{v}_j}{\mathbf{v}_j^T \mathbf{v}_j} \mathbf{v}_j$$

Notice that each term in the sum $\sum_{j=1}^{r} \frac{\mathbf{v}^T \mathbf{v}_j}{\mathbf{v}_j^T \mathbf{v}_j} \mathbf{v}_j$ has the same formula as the vector projection of $\mathbf{v}$ onto $\mathbf{v}_j$.

Theorem (Orthogonal bases)

Every nonzero subspace of $\mathbb{R}^n$ *has a basis that is orthonormal.*

Theorem (Formula for the projection)

Let S *is a subspace of* $\mathbb{R}^n$ *and let* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$ *be an orthogonal basis for* S. *If* $\mathbf{v}$ *is any vector in* $\mathbb{R}^n$ *and* $\mathbf{p}$ *the element of* S *that is closest to* $\mathbf{v}$, *then*

$$\mathbf{p} = \sum_{j=1}^{r} \frac{\mathbf{v}^T \mathbf{v}_j}{\mathbf{v}_j^T \mathbf{v}_j} \mathbf{v}_j$$

Notice that each term in the sum $\sum_{j=1}^{r} \frac{\mathbf{v}^T \mathbf{v}_j}{\mathbf{v}_j^T \mathbf{v}_j} \mathbf{v}_j$ has the same formula as the vector projection of $\mathbf{v}$ onto $\mathbf{v}_j$. We will call $\mathbf{p}$ the *projection of* $\mathbf{v}$ *onto* $\mathrm{Span}(\mathbf{v}_1, \ldots, \mathbf{v}_r)$.

There are many vector spaces with a type of product similar to the scalar product in $\mathbb{R}^n$.

Definition (Inner product)

If V is a vector space, then an *inner product* is an operation that assigns, to any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, satisfying

There are many vector spaces with a type of product similar to the scalar product in $\mathbb{R}^n$.

Definition (Inner product)

If V is a vector space, then an *inner product* is an operation that assigns, to any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, satisfying

1. $\langle \mathbf{0}, \mathbf{0} \rangle = 0$, and if $\mathbf{x} \neq \mathbf{0}$ then $\langle \mathbf{x}, \mathbf{x} \rangle > 0$.

There are many vector spaces with a type of product similar to the scalar product in $\mathbb{R}^n$.

Definition (Inner product)

If V is a vector space, then an *inner product* is an operation that assigns, to any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, satisfying

1. $\langle \mathbf{0}, \mathbf{0} \rangle = 0$, and if $\mathbf{x} \neq \mathbf{0}$ then $\langle \mathbf{x}, \mathbf{x} \rangle > 0$.
2. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}$ in V

There are many vector spaces with a type of product similar to the scalar product in $\mathbb{R}^n$.

Definition (Inner product)

If V is a vector space, then an *inner product* is an operation that assigns, to any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, satisfying

1. $\langle \mathbf{0}, \mathbf{0} \rangle = 0$, and if $\mathbf{x} \neq \mathbf{0}$ then $\langle \mathbf{x}, \mathbf{x} \rangle > 0$.
2. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}$ in V
3. $\langle \alpha \mathbf{y} + \beta \mathbf{z}, \mathbf{x} \rangle = \alpha \langle \mathbf{y}, \mathbf{x} \rangle + \beta \langle \mathbf{z}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in V and every α, β in $\mathbb{R}$.

There are many vector spaces with a type of product similar to the scalar product in $\mathbb{R}^n$.

Definition (Inner product)

If V is a vector space, then an *inner product* is an operation that assigns, to any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, satisfying

1. $\langle \mathbf{0}, \mathbf{0} \rangle = 0$, and if $\mathbf{x} \neq \mathbf{0}$ then $\langle \mathbf{x}, \mathbf{x} \rangle > 0$.
2. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}$ in V
3. $\langle \alpha\mathbf{y} + \beta\mathbf{z}, \mathbf{x} \rangle = \alpha \langle \mathbf{y}, \mathbf{x} \rangle + \beta \langle \mathbf{z}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in V and every α, β in $\mathbb{R}$.

A vector space with an inner product defined on it is called an *inner product space*.

One way to create an inner product is to produce a linear one-to-one correspondence from V to a vector space that already has an inner product.

There are many vector spaces with a type of product similar to the scalar product in $\mathbb{R}^n$.

Definition (Inner product)

If V is a vector space, then an *inner product* is an operation that assigns, to any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, satisfying

1. $\langle \mathbf{0}, \mathbf{0} \rangle = 0$, and if $\mathbf{x} \neq \mathbf{0}$ then $\langle \mathbf{x}, \mathbf{x} \rangle > 0$.
2. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}$ in V
3. $\langle \alpha \mathbf{y} + \beta \mathbf{z}, \mathbf{x} \rangle = \alpha \langle \mathbf{y}, \mathbf{x} \rangle + \beta \langle \mathbf{z}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in V and every α, β in $\mathbb{R}$.

A vector space with an inner product defined on it is called an *inner product space*.

One way to create an inner product is to produce a linear one-to-one correspondence from V to a vector space that already has an inner product.

$\mathbb{R}^n$ can have inner products different from the scalar product.

There are many vector spaces with a type of product similar to the scalar product in $\mathbb{R}^n$.

Definition (Inner product)

If V is a vector space, then an *inner product* is an operation that assigns, to any pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number, denoted $\langle \mathbf{x}, \mathbf{y} \rangle$, satisfying

1. $\langle \mathbf{0}, \mathbf{0} \rangle = 0$, and if $\mathbf{x} \neq \mathbf{0}$ then $\langle \mathbf{x}, \mathbf{x} \rangle > 0$.
2. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}$ in V
3. $\langle \alpha\mathbf{y} + \beta\mathbf{z}, \mathbf{x} \rangle = \alpha \langle \mathbf{y}, \mathbf{x} \rangle + \beta \langle \mathbf{z}, \mathbf{x} \rangle$, for every $\mathbf{x}, \mathbf{y}, \mathbf{z}$ in V and every α, β in $\mathbb{R}$.

A vector space with an inner product defined on it is called an *inner product space*.

One way to create an inner product is to produce a linear one-to-one correspondence from V to a vector space that already has an inner product.

$\mathbb{R}^n$ can have inner products different from the scalar product. If A is any invertible $n \times n$ matrix, then

$$\langle \mathbf{x}, \mathbf{y} \rangle = (A\mathbf{x})^T A\mathbf{y} = \mathbf{x}^T A^T A\mathbf{y}$$

is an example of an inner product.

Definition (Induced norm)

If V is a vector space with an inner product $\langle \mathbf{x}, \mathbf{y} \rangle$, then $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. This is called the *norm induced by* this inner product.

Definition (Induced norm)

If V is a vector space with an inner product $\langle \mathbf{x}, \mathbf{y} \rangle$, then $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. This is called the *norm induced by* this inner product.

Definition (Orthogonal)

If V is a vector space with an inner product $\langle \mathbf{x}, \mathbf{y} \rangle$, then we say $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$ if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. and we express this by $\mathbf{x} \perp \mathbf{y}$.

Definition (Induced norm)

If V is a vector space with an inner product $\langle \mathbf{x}, \mathbf{y} \rangle$, then $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$. This is called the *norm induced by* this inner product.

Definition (Orthogonal)

If V is a vector space with an inner product $\langle \mathbf{x}, \mathbf{y} \rangle$, then we say $\mathbf{x}$ *is orthogonal to* $\mathbf{y}$ if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$. and we express this by $\mathbf{x} \perp \mathbf{y}$.

Theorem (Pythagorean Formula)

If $\mathbf{x} \perp \mathbf{y}$ *then* $\|\mathbf{x} + \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$.

Theorem (Cauchy-Schwarz Inequality)

For any $\mathbf{x}$ *and* $\mathbf{y}$ *in an inner product space,*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$$

Theorem (Triangle Inequality)

For any pair of vectors $\mathbf{x}$ *and* $\mathbf{y}$, $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

Theorem (Triangle Inequality)

For any pair of vectors $\mathbf{x}$ *and* $\mathbf{y}$, $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

This allows us to express the distance between vectors by $d(\mathbf{x},\mathbf{y}) = \|\mathbf{x}-\mathbf{y}\|$

Theorem (Triangle Inequality)

For any pair of vectors $\mathbf{x}$ *and* $\mathbf{y}$*,* $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.

This allows us to express the distance between vectors by $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$

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

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

Theorem (Triangle Inequality)

For any pair of vectors $\mathbf{x}$ *and* $\mathbf{y}$*,* $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$*.*

This allows us to express the distance between vectors by $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|$

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

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

In this inner product space, the polynomials x and x^2 are orthogonal to each other. Also the constant polynomial 1 is orthogonal to x but not to x^2. Exercise: $1 - x^2$ is orthogonal to both x and x^2.

Definition (Scalar/vector projection)

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 (Scalar/vector projection)

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 (Orthogonal/orthonormal set)

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 (Orthogonal/orthonormal set)

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 (Orthogonal/orthonormal set)

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 (Orthogonal/orthonormal set)

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.

The columns of A being orthonormal means $A^T A = I$, the $k \times k$ identity matrix. The matrix AA^T doesn't have to be the identity unless A is a square matrix. In general, AA^T is the projection matrix for $\mathcal{R}(A)$: $AA^T\mathbf{b}$ is the closest vector in $\mathcal{R}(A)$ to $\mathbf{b}$.

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.

The columns of A being orthonormal means $A^TA = I$, the $k \times k$ identity matrix. The matrix AA^T doesn't have to be the identity unless A is a square matrix. In general, AA^T is the projection matrix for $\mathcal{R}(A)$: $AA^T\mathbf{b}$ is the closest vector in $\mathcal{R}(A)$ to $\mathbf{b}$.

If it happens that $n = k$ so that A is a square matrix, then $A^TA = I$ 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.

The columns of A being orthonormal means $A^T A = I$, the $k \times k$ identity matrix. The matrix AA^T doesn't have to be the identity unless A is a square matrix. In general, AA^T is the projection matrix for $\mathcal{R}(A)$: $AA^T\mathbf{b}$ is the closest vector in $\mathcal{R}(A)$ to $\mathbf{b}$.

If it happens that $n = k$ so that A is a square matrix, then $A^T A = I$ 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 (Orthogonal matrix)

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 (Orthogonal matrix)

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

The product of two (or more) orthogonal matrices is an orthogonal matrices.

Definition (Orthogonal matrix)

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

The product of two (or more) orthogonal matrices is an orthogonal matrices.

For an orthonormal basis $\mathcal{B}$, if Q is the matrix whose columns are the vectors in $\mathcal{B}$, then Q is the transition matrix from $\mathcal{B}$ to $\mathcal{E}$ and Q^T is the transition matrix from $\mathcal{E}$ to $\mathcal{B}$.

Definition (Orthogonal matrix)

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

The product of two (or more) orthogonal matrices is an orthogonal matrices.

For an orthonormal basis $\mathcal{B}$, if Q is the matrix whose columns are the vectors in $\mathcal{B}$, 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:

Definition (Orthogonal matrix)

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

The product of two (or more) orthogonal matrices is an orthogonal matrices.

For an orthonormal basis $\mathcal{B}$, if Q is the matrix whose columns are the vectors in $\mathcal{B}$, 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 (Coordinates for orthonormal bases)

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*

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

Definition (Orthogonal matrix)

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

The product of two (or more) orthogonal matrices is an orthogonal matrices.

For an orthonormal basis $\mathcal{B}$, if Q is the matrix whose columns are the vectors in $\mathcal{B}$, 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 (Coordinates for orthonormal bases)

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

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

Some examples of orthonormal sets and orthogonal matrices

Some examples of orthonormal sets and orthogonal matrices

The set $\mathcal{B} = \left[\mathbf{u}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} \right]$ is an orthonormal basis in $\mathbb{R}^2$.

Some examples of orthonormal sets and orthogonal matrices

The set $\mathcal{B} = \left[\mathbf{u}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}\right]$ is an orthonormal basis in $\mathbb{R}^2$.

If we take any other vector, for example $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$,

Some examples of orthonormal sets and orthogonal matrices

The set $\mathcal{B} = \left[\mathbf{u}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} \right]$ is an orthonormal basis in $\mathbb{R}^2$.

If we take any other vector, for example $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, we can write

$$\mathbf{x} = (\mathbf{x}^T \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}^T \mathbf{u}_2)\mathbf{u}_2 = \sqrt{2}\, \mathbf{u}_1 - 2\sqrt{2}\, \mathbf{u}_2.$$

Some examples of orthonormal sets and orthogonal matrices

The set $\mathcal{B} = \left[\mathbf{u}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} \right]$ is an orthonormal basis in $\mathbb{R}^2$.

If we take any other vector, for example $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, we can write

$\mathbf{x} = (\mathbf{x}^T \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}^T \mathbf{u}_2)\mathbf{u}_2 = \sqrt{2}\,\mathbf{u}_1 - 2\sqrt{2}\,\mathbf{u}_2$. That is, $[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} \sqrt{2} \\ -2\sqrt{2} \end{pmatrix}$

Some examples of orthonormal sets and orthogonal matrices

The set $\mathcal{B} = \left[\mathbf{u}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix} \right]$ is an orthonormal basis in $\mathbb{R}^2$.

If we take any other vector, for example $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, we can write

$\mathbf{x} = (\mathbf{x}^T \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}^T \mathbf{u}_2)\mathbf{u}_2 = \sqrt{2}\,\mathbf{u}_1 - 2\sqrt{2}\,\mathbf{u}_2$. That is, $[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} \sqrt{2} \\ -2\sqrt{2} \end{pmatrix}$

Moreover, $Q = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$ is an orthogonal matrix

Some examples of orthonormal sets and orthogonal matrices

The set $\mathcal{B} = \left[\mathbf{u}_1 = \begin{pmatrix} \sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} -\sqrt{2}/2 \\ \sqrt{2}/2 \end{pmatrix}\right]$ is an orthonormal basis in $\mathbb{R}^2$.

If we take any other vector, for example $\mathbf{x} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$, we can write

$\mathbf{x} = (\mathbf{x}^T\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}^T\mathbf{u}_2)\mathbf{u}_2 = \sqrt{2}\,\mathbf{u}_1 - 2\sqrt{2}\,\mathbf{u}_2$. That is, $[\mathbf{x}]_{\mathcal{B}} = \begin{pmatrix} \sqrt{2} \\ -2\sqrt{2} \end{pmatrix}$

Moreover, $Q = \begin{pmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$ is an orthogonal matrix and

$Q^T = \begin{pmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{pmatrix}$ is its inverse. Finally, note that $Q^T\mathbf{x} = [\mathbf{x}]_{\mathcal{B}}$

A three-dimensional example:

A three-dimensional example:

Here is an orthogonal set: $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$.

A three-dimensional example:

Here is an orthogonal set: $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$. Therefore, $\mathcal{B} = \left[\mathbf{u}_1 = (1/\sqrt{6})\mathbf{v}_1, \mathbf{u}_2 = (1/\sqrt{3})\mathbf{v}_2, \mathbf{u}_3 = (1/\sqrt{2})\mathbf{v}_3\right]$ is an orthonormal basis and

A three-dimensional example:

Here is an orthogonal set: $\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$. Therefore, $\mathcal{B} = \left[\mathbf{u}_1 = (1/\sqrt{6})\mathbf{v}_1, \mathbf{u}_2 = (1/\sqrt{3})\mathbf{v}_2, \mathbf{u}_3 = (1/\sqrt{2})\mathbf{v}_3\right]$ is an orthonormal basis and

$$Q = \begin{pmatrix} 1/\sqrt{6} & 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{6} & 1/\sqrt{3} & -1/\sqrt{2} \\ 2/\sqrt{6} & -1/\sqrt{3} & 0 \end{pmatrix}$$

is an orthogonal matrix.

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T \mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T \mathbf{v} \\ \mathbf{u}_2^T \mathbf{v} \\ \mathbf{u}_3^T \mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T\mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v} \\ \mathbf{u}_2^T\mathbf{v} \\ \mathbf{u}_3^T\mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

The projection matrix

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T\mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v} \\ \mathbf{u}_2^T\mathbf{v} \\ \mathbf{u}_3^T\mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

The projection matrix

Let S be a subspace of $\mathbb{R}^n$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S.

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T\mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v} \\ \mathbf{u}_2^T\mathbf{v} \\ \mathbf{u}_3^T\mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

The projection matrix

Let S be a subspace of $\mathbb{R}^n$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S. For any vector $\mathbf{b} \in \mathbb{R}^n$ we can get the closest vector to S by solving a least squares problem.

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T\mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v} \\ \mathbf{u}_2^T\mathbf{v} \\ \mathbf{u}_3^T\mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

The projection matrix

Let S be a subspace of $\mathbb{R}^n$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S. For any vector $\mathbf{b} \in \mathbb{R}^n$ we can get the closest vector to S by solving a least squares problem.

Recall: If U is the matrix whose columns are $\mathbf{u}_1, \ldots, \mathbf{u}_r$, then we solve $U^TU\mathbf{x} = U^T\mathbf{b}$ to get $\hat{\mathbf{x}}$,

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T\mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v} \\ \mathbf{u}_2^T\mathbf{v} \\ \mathbf{u}_3^T\mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

The projection matrix

Let S be a subspace of $\mathbb{R}^n$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S. For any vector $\mathbf{b} \in \mathbb{R}^n$ we can get the closest vector to S by solving a least squares problem.

Recall: If U is the matrix whose columns are $\mathbf{u}_1, \ldots, \mathbf{u}_r$, then we solve $U^TU\mathbf{x} = U^T\mathbf{b}$ to get $\hat{\mathbf{x}}$, and then $U\hat{\mathbf{x}}$ is then the closest vector to $\mathbf{b}$ in the column space of U (which is S).

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T\mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v} \\ \mathbf{u}_2^T\mathbf{v} \\ \mathbf{u}_3^T\mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

The projection matrix

Let S be a subspace of $\mathbb{R}^n$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S. For any vector $\mathbf{b} \in \mathbb{R}^n$ we can get the closest vector to S by solving a least squares problem.

Recall: If U is the matrix whose columns are $\mathbf{u}_1, \ldots, \mathbf{u}_r$, then we solve $U^TU\mathbf{x} = U^T\mathbf{b}$ to get $\hat{\mathbf{x}}$, and then $U\hat{\mathbf{x}}$ is then the closest vector to $\mathbf{b}$ in the column space of U (which is S).

But, since U has orthogonal columns, the $U^TU = I$ $(r \times r)$ and so the solution of $U^TU\mathbf{x} = U^T\mathbf{b}$ is $\hat{\mathbf{x}} = U^T\mathbf{b}$ and the closest vector in S to $\mathbf{b}$ is $U\hat{\mathbf{x}} = UU^T\mathbf{b}$.

Taking any vector, for example $\mathbf{v} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ we get

$$[\mathbf{v}]_{\mathcal{B}} = Q^T\mathbf{v} = \begin{pmatrix} \mathbf{u}_1^T\mathbf{v} \\ \mathbf{u}_2^T\mathbf{v} \\ \mathbf{u}_3^T\mathbf{v} \end{pmatrix} = \begin{pmatrix} 5/\sqrt{6} \\ 2/\sqrt{3} \\ 1/\sqrt{2} \end{pmatrix}$$

The projection matrix

Let S be a subspace of $\mathbb{R}^n$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_r\}$ an orthonormal basis for S. For any vector $\mathbf{b} \in \mathbb{R}^n$ we can get the closest vector to S by solving a least squares problem.

Recall: If U is the matrix whose columns are $\mathbf{u}_1, \ldots, \mathbf{u}_r$, then we solve $U^TU\mathbf{x} = U^T\mathbf{b}$ to get $\hat{\mathbf{x}}$, and then $U\hat{\mathbf{x}}$ is then the closest vector to $\mathbf{b}$ in the column space of U (which is S).

But, since U has orthogonal columns, the $U^TU = I$ $(r \times r)$ and so the solution of $U^TU\mathbf{x} = U^T\mathbf{b}$ is $\hat{\mathbf{x}} = U^T\mathbf{b}$ and the closest vector in S to $\mathbf{b}$ is $U\hat{\mathbf{x}} = UU^T\mathbf{b}$.

We call the matrix $P = UU^T$ the *projection matrix*.

Here is an example. Find the projection matrix for the span of

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Here is an example. Find the projection matrix for the span of

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Use it to find the projection of $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ onto this span.

Here is an example. Find the projection matrix for the span of

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}.$$

Use it to find the projection of $\mathbf{v} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ onto this span.

These are orthogonal, but not orthonormal, so we normalize them:

$\mathbf{u}_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$, and we get the projection matrix

$$\begin{aligned} P = UU^T &= \begin{pmatrix} 1/3 & 0 \\ 2/3 & -1/\sqrt{2} \\ 2/3 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 & 2/3 \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \\ &= \begin{pmatrix} 1/9 & 2/9 & 2/9 \\ 2/9 & 17/18 & -1/18 \\ 2/9 & -1/18 & 17/18 \end{pmatrix} \end{aligned}$$

Then $P\mathbf{v} = \begin{pmatrix} 1/3 \\ 7/6 \\ 1/6 \end{pmatrix}$.

Then $P\mathbf{v} = \begin{pmatrix} 1/3 \\ 7/6 \\ 1/6 \end{pmatrix}$. We can check our work (in part) by determining whether the difference $\mathbf{v} - P\mathbf{v} = \begin{pmatrix} 2/3 \\ -1/6 \\ -1/6 \end{pmatrix}$ is orthogonal to both $\mathbf{v}_1$ and $\mathbf{v}_2$.