Orthogonality

Cosines and Projections onto lines

Vectors: Algebraic Approach

- An n-dimensional coordinate vector is an element of $\mathbf{R}^n$, i.e., an ordered n-tuple $(\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_n)$ of real numbers.
- Let $\mathbf{a} = (\mathrm{a}_1, \mathrm{a}_2, \ldots, \mathrm{a}_n)$ and $\mathbf{b} = (\mathrm{b}_1, \mathrm{b}_2, \ldots, \mathrm{b}_n)$ be vectors, and $r \in \mathbf{R}$ be a scalar. Then, by definition,

$$\mathbf{a} + \mathbf{b} = (a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n),$$

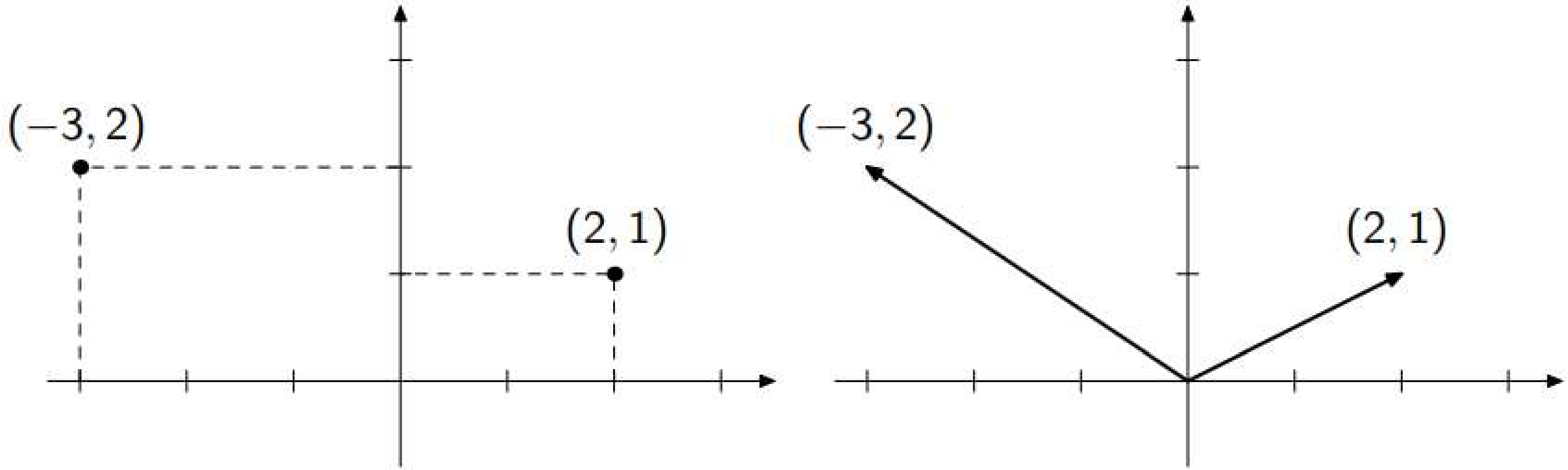
$$r\mathbf{a} = (ra_1, ra_2, \ldots, ra_n),$$

$$\mathbf{0} = (0, 0, \ldots, 0),$$

$$-\mathbf{b} = (-b_1, -b_2, \ldots, -b_n),$$

$$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b}) = (a_1 - b_1, a_2 - b_2, \ldots, a_n - b_n).$$

Cartesian coordinates: geometric meets algebraic



- Once we specify an origin O, each point A is associated a position vector $\overrightarrow{OA}$. Conversely, every vector has a unique representative with tail at O.
- Cartesian coordinates allow us to identify a line, a plane, and space with $\mathbf{R}, \mathbf{R}^2$ and $\mathbf{R}^3$, respectively.

Length and Distance

Definition. The **length** of a vector

$\mathbf{v} = (\mathrm{v}_1, \mathrm{v}_2, \dots, \mathrm{v}_n) \in \mathbf{R}^{\mathrm{n}}$ is

$$||\mathbf{v}|| = \sqrt{\mathrm{v}_1^2 + \mathrm{v}_2^2 + \cdots + \mathrm{v}_n^2}$$

The **distance** between vectors/points **x** and **y** is $||\mathbf{y} - \mathbf{x}||$.

Properties of length:

$||\mathbf{x}|| \geq 0, ||\mathbf{x}|| = 0 \text{ only if } \mathbf{x} = 0$ (positivity)

$||r\mathbf{x}|| = |r|\ ||\mathbf{x}||$ (homogeneity)

$||\mathbf{x} + \mathbf{y}|| \leq ||\mathbf{x}|| + ||\mathbf{y}||$ (triangle inequality)

Scalar Product

Definition. The **scalar product** of vectors $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ is

$$\boxed{\mathbf{x} \,.\, \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n}$$

Properties of scalar product:

$\mathbf{x} \,.\, \mathbf{x} \geq 0$, $\mathbf{x} \,.\, \mathbf{x} = 0$ only if $\mathbf{x} = 0$ (positivity)

$\mathbf{x} \,.\, \mathbf{y} = \mathbf{y} \,.\, \mathbf{x}$ (symmetry)

$(\mathbf{x} + \mathbf{y}).\mathbf{z} = \mathbf{x} \,.\, \mathbf{z} + \mathbf{y} \,.\, \mathbf{z}$ (distributive law)

$(r\mathbf{x}) \,.\, \mathbf{y} = r(\mathbf{x} \,.\, \mathbf{y})$ (homogeneity)

Angle

Cauchy-Schwarz inequality: $|\mathbf{x}\,.\,\mathbf{y}| \leq ||\mathbf{x}||\ ||\mathbf{y}||$

By the Cauchy-Schwarz inequality, for any nonzero vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^n$ we have

$$\cos\theta = \frac{\mathbf{x}\,.\mathbf{y}}{||\mathbf{x}||\ ||\mathbf{y}||} \text{ for a unique } 0 \leq \theta \leq \pi.$$

θ is called the **angle** between the vectors $\mathbf{x}$ & $\mathbf{y}$.

The vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be **orthogonal** (denoted $\mathbf{x} \perp \mathbf{y}$) if $\mathbf{x} \cdot \mathbf{y} = 0$ (i.e., $\theta = 90°$).

Problem. Find the angle θ between vectors $\mathbf{x} = (2, -1)$ and $\mathbf{y} = (3, 1)$.

$\mathbf{x} \,.\, \mathbf{y} = 5,\ \|\mathbf{x}\| = \sqrt{5},\ \|\mathbf{y}\| = \sqrt{10}.$

$$\cos\theta = \frac{\mathbf{x} \,.\, \mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|} = \frac{5}{\sqrt{5}\sqrt{10}} = \frac{1}{\sqrt{2}} \Rightarrow \theta = 45^{\circ}.$$

Problem. Find the angle ϕ between vectors $\mathbf{v} = (-2, 1, 3)$ and $\mathbf{w} = (4, 5, 1)$.

$\mathbf{v} \,.\, \mathbf{w} = 0 \Rightarrow \boldsymbol{v} \perp \boldsymbol{w} \Rightarrow \phi = 90°.$

Orthogonality

Definition 1. Vectors $\mathbf{x}, \mathbf{y} \in \mathbf{R}^{\mathrm{n}}$ are said to be **orthogonal** (denoted $\mathbf{x} \perp \mathbf{y}$) if $\mathbf{x} \cdot \mathbf{y} = 0$.

Definition 2. A vector $\mathbf{x} \in \mathbf{R}^n$ is said to be **orthogonal** to a nonempty set $Y \subset \mathbf{R}^n$ (denoted $\mathbf{x} \perp Y$) if $\mathbf{x} \cdot \mathbf{y} = 0$ for any $\mathbf{y} \in Y$.

Note: $\mathbf{x} \cdot \mathbf{y} = (\mathbf{x}^{\mathrm{T}}\mathbf{y})$.

Definition 3. Nonempty sets $X, Y \subset \mathbf{R}^n$ are said to be **orthogonal** (denoted $X \perp Y$) if $\mathbf{x} \cdot \mathbf{y} = 0$ for any $\mathbf{x} \in X$ and $\mathbf{y} \in Y$.

The line $x = y = 0$ is orthogonal to the line $y = z = 0$.

Indeed, if $\mathbf{v} = (0, 0, z)$ and $\mathbf{w} = (x, 0, 0)$ then $\mathbf{v} \cdot \mathbf{w} = 0$.

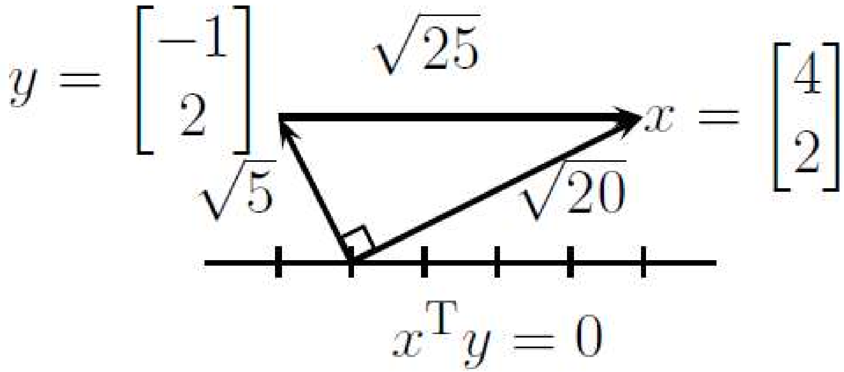
The line $x = y = 0$ is orthogonal to the plane $z = 0$.

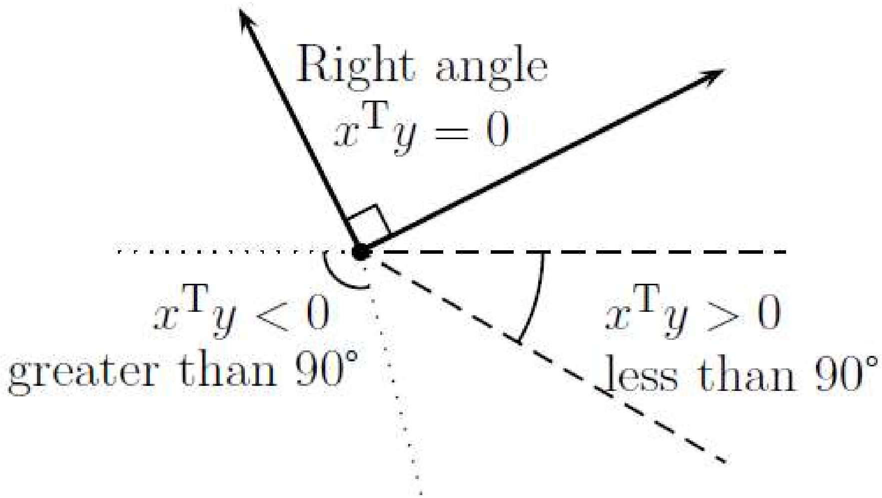
Indeed, if $\mathbf{v} = (0, 0, z)$ and $\mathbf{w} = (x, y, 0)$ then $\mathbf{v} \cdot \mathbf{w} = 0$.

The line $x = y = 0$ is not orthogonal to the plane $z = 1$.

The vector $\mathbf{v} = (0, 0, 1)$ belongs to both the line and the plane, and $\mathbf{v} \cdot \mathbf{v} = 1 \neq 0$.

Orthogonality - Geometric Interpretation





- Considering the figure on the left, the dot product of the two vectors $\mathbf{x}$ and $\mathbf{y}$ is equal to 0.
- This shows that the two vectors are orthogonal.
- If $\mathbf{x}^{\mathbf{T}}\mathbf{y} > \mathbf{0}$, the angle between the two vectors is less than 90°.
- If $\mathbf{x}^{\mathbf{T}}\mathbf{y} < \mathbf{0}$, the angle between the two vectors is greater than 90°.

Theorem 1: If nonzero vectors $\mathbf{v_1}; \ldots ; \mathbf{v_k}$ are mutually orthogonal (every vector is perpendicular to every other), then those vectors are linearly independent.

Proof:

- Suppose $c_1\mathbf{v_1} + \cdots + c_k\mathbf{v_k} = 0$. To show that c_1 must be zero, take the inner product of both sides with $\mathbf{v_1}$. Orthogonality of the $\mathbf{v}$'s leaves only one term:
$$\mathbf{v_1^T}(c_1\mathbf{v_1} + \cdots + c_k\mathbf{v_k}) = c_1\mathbf{v_1^T}\mathbf{v_1} = 0$$
- The vectors are nonzero, so $\mathbf{v_1^T}\mathbf{v_1} = 0$ and therefore $c_1 = 0$. The same is true of every c_i.
- The only combination of the $\mathbf{v}$'s producing zero has all $c_i = 0$
- This shows that the vectors are linearly independent.

- The coordinate vectors $\mathbf{e_1}; \ldots ; \mathbf{e_n}$ in $\mathbf{R}^{\mathrm{n}}$ are the most important orthogonal vectors.
- Those are the columns of the identity matrix. They form the simplest basis for $\mathbf{R}^{\mathrm{n}}$, and they are *unit vectors*—each has length $\|\mathbf{e_i}\| = 1$. They point along the coordinate axes.
- If these axes are rotated, the result is a new **orthonormal basis**: a new system of *mutually orthogonal unit vectors*.
- In $\mathbf{R}^2$ we have $\cos^2\theta + \sin^2\theta = 1$:
- Orthonormal vectors in $\boldsymbol{R}^2$
  $\mathbf{v_1} = (\cos\theta ; \sin\theta)$ and $\mathbf{v_2} = (-\sin\theta; \cos\theta)$

Orthogonal Subspaces

- Two subspaces $\boldsymbol{V}$ and $\boldsymbol{W}$ of the same space $\mathbf{R}^{\mathrm{n}}$ are *orthogonal* if every vector $\mathbf{v}$ in $\boldsymbol{V}$ is orthogonal to every vector $\mathbf{w}$ in $\boldsymbol{W}$: $\mathbf{v}^{\mathbf{T}}\mathbf{w} = 0$ for all $\mathbf{v}$ and $\mathbf{w}$.

- **Example** : Suppose $\boldsymbol{V}$ is the plane spanned by $\mathbf{v_1} = (1; 0; 0; 0)$ and $\mathbf{v_2} = (1; 1; 0; 0)$. If $\boldsymbol{W}$ is the line spanned by $\mathbf{w} = (0; 0; 4; 5)$, then $\mathbf{w}$ is orthogonal to both $\mathbf{v}$’s. The line $\boldsymbol{W}$ will be orthogonal to the whole plane $\boldsymbol{V}$.

Theorem 2. $N(A) \perp R(A)$. That is, the **null-space** of a matrix is **orthogonal** to its **row space**.

Also, $N(A^T) \perp C(A)$, ie. the **column space** of a matrix is **orthogonal** to the **left null space** of the matrix.

Example

- Suppose A has rank 1, so its row space and column space are lines:
- Rank-1 matrix $A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \\ 3 & 9 \end{bmatrix}$
- The rows are multiples of $(1; 3)$. The null space contains $\mathbf{x} = (-3; 1)$, which is orthogonal to all the rows.
- The null space and row space are perpendicular lines in $\mathbf{R}^2$:

$$[1\ 3]\begin{bmatrix} 3 \\ -1 \end{bmatrix} = 0 \text{ and } [2\ 6]\begin{bmatrix} 3 \\ -1 \end{bmatrix} = 0 \text{ and } [3\ 9]\begin{bmatrix} 3 \\ -1 \end{bmatrix} = 0.$$

- In contrast, the other two subspaces are in $\mathbf{R}^3$. The column space is the line through $(1; 2; 3)$. The left null space must be the *perpendicular plane* $y_1 + 2y_2 + 3y_3 = 0$. That equation is exactly the content of $\mathbf{y}^T A = 0$.

Orthogonal Complement

- ***Definition :*** Given a subspace $\boldsymbol{V}$ of $\mathbf{R}^{\mathrm{n}}$, the space of *all* vectors orthogonal to $\boldsymbol{V}$ is called the orthogonal complement of $\boldsymbol{V}$.
- It is denoted by $\boldsymbol{V}^{\perp} =$ "$\boldsymbol{V}$ $\boldsymbol{perp}$."

Theorem 3.

I. $\boldsymbol{V}^{\perp}$ is a subspace of $\mathbf{R}^{n}$.

II. $Span(\boldsymbol{V})^{\perp} = \boldsymbol{V}^{\perp}$.

III. $(\boldsymbol{V}^{\perp})^{\perp} = Span(\boldsymbol{V})$.

Theorem 4. If $\boldsymbol{V}$ is a subspace of $\mathbf{R}^{\mathrm{n}}$, then

I. $(\boldsymbol{V}^{\perp})^{\perp} = \boldsymbol{V}$.

II. $\boldsymbol{V} \cap \boldsymbol{V}^{\perp} = \{\mathbf{0}\}$.

Orthogonal Complement

Definition. Let $\boldsymbol{S} \subset \mathbf{R}^n$. The **orthogonal complement** of $\boldsymbol{S}$, denoted $\boldsymbol{S}^\perp$, is the set of all vectors $\mathbf{x} \in \mathbf{R}^n$ that are orthogonal to $\boldsymbol{S}$.

Theorem 2.

I. $\boldsymbol{S}^\perp$ is a subspace of $\mathbf{R}^n$.

II. $Span(\boldsymbol{S})^\perp = \boldsymbol{S}^\perp$.

III. $(\boldsymbol{S}^\perp)^\perp = Span(\boldsymbol{S})$.

Theorem 3. If $\boldsymbol{V}$ is a subspace of $\mathbf{R}^\mathrm{n}$, then

I. $(\boldsymbol{V}^\perp)^\perp = \boldsymbol{V}$.

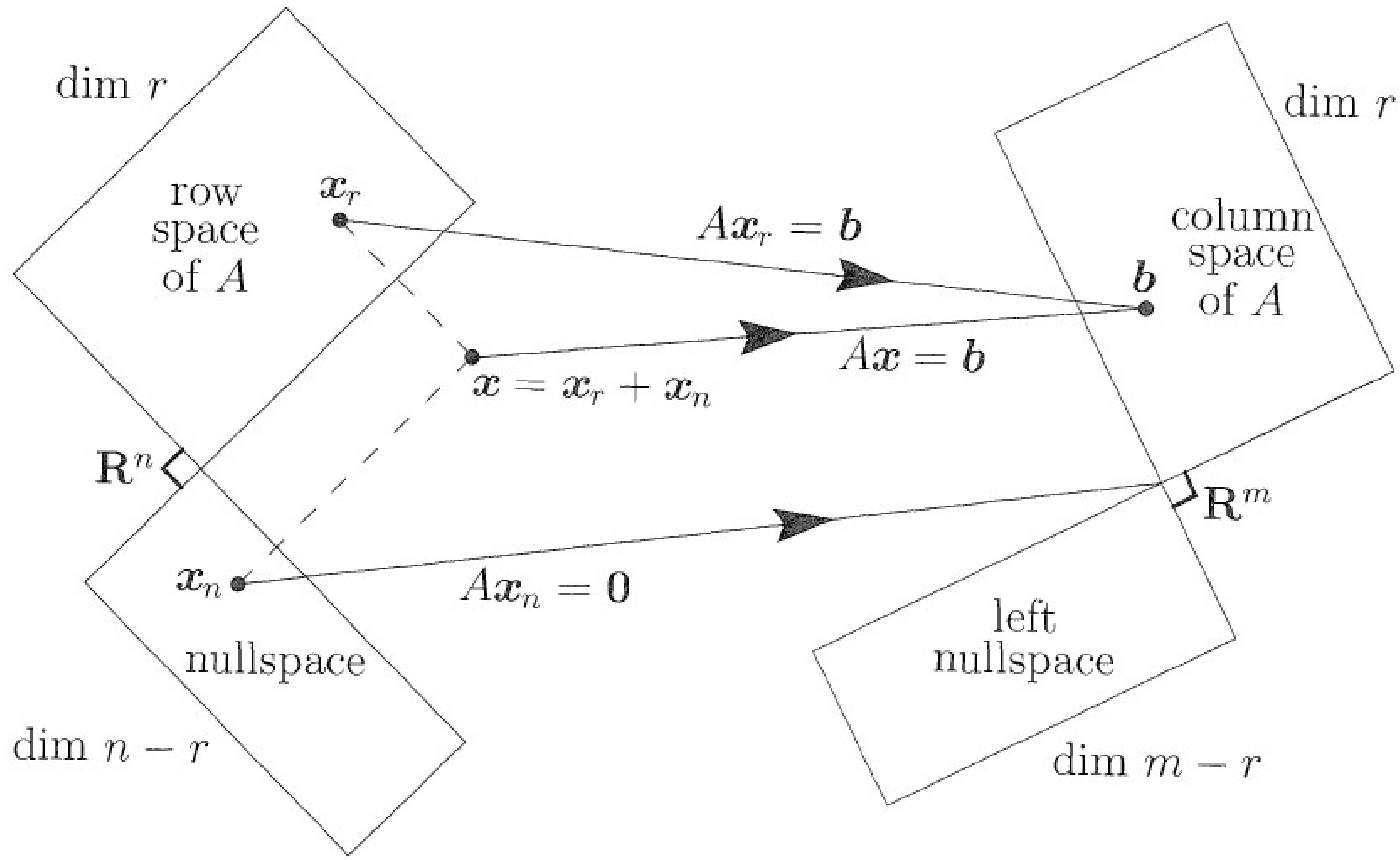
II. $\boldsymbol{V} \cap \boldsymbol{V}^\perp = \{\mathbf{0}\}$.

Orthogonal Complement

- The null space is the orthogonal complement of the row space: $N(A) = \left(C(A^T)\right)^{\perp}$.
- **Dimension formula :** dim(row space) + dim(null space) = number of columns.
- Every vector orthogonal to the null space is in the row space: $C(A^T) = \left(N(A)\right)^{\perp}$.
- The same reasoning applied to A^T produces the dual result: *The left null space $N(A^T)$ and the column space $C(A)$ are* ***orthogonal complements.***
- **Fundamental Theorem of Linear Algebra:** The null space is the *orthogonal complement* of the row space in $\mathbf{R}^n$.
- The left null space is the *orthogonal complement* of the column space in $\mathbf{R}^m$.

Matrix and the subspaces

- Two matrices V and W can be orthogonal without being complements. The line V spanned by $(0; 1; 0)$ is orthogonal to the line W spanned by $(0; 0; 1)$, but V is not $W^{\perp}$.
- The orthogonal complement of W is a two dimensional plane, and the line is only part of $W^{\perp}$.
- When the dimensions are right, orthogonal subspaces *are* necessarily orthogonal complements.
- If $W = V^{\perp}$ then $V = W^{\perp}$ and dim V + dim W = n.
- Splitting $\mathbf{R}^n$ into orthogonal parts will split every vector into $\mathbf{x} = \mathbf{v} + \mathbf{w}$. The vector $\mathbf{v}$ is the projection onto the subspace $\boldsymbol{V}$. The orthogonal component $\mathbf{w}$ is the projection of $\mathbf{x}$ onto $\boldsymbol{W}$.



Given figure summarizes the fundamental theorem of linear algebra

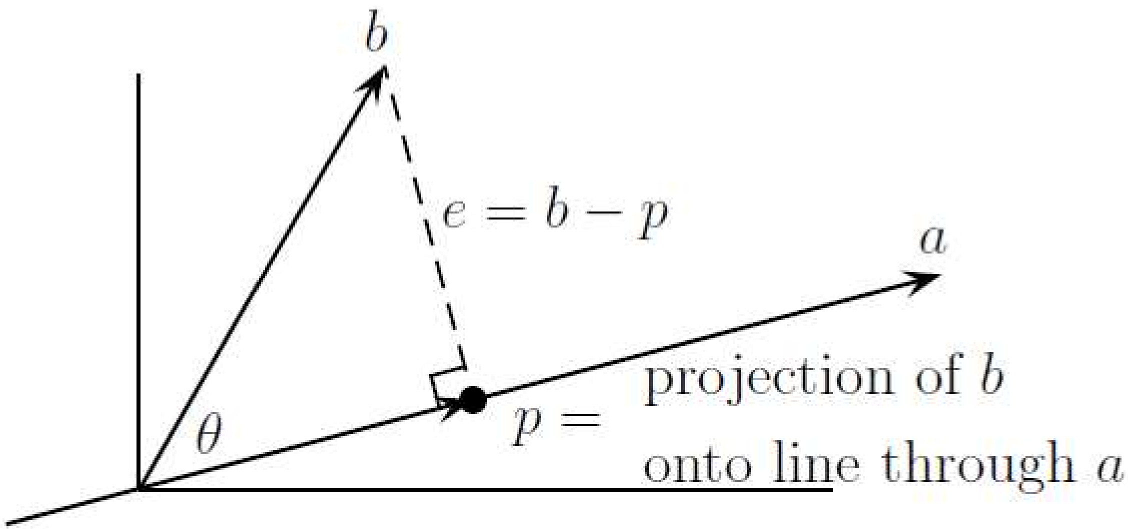
- The null space is carried to the zero vector. Every $A\mathbf{x}$ is in the column space. Nothing is carried to the left null space.
- A typical vector $\mathbf{x}$ has a “row space component” and a “null space component” with $\mathbf{x} = \mathbf{x}_r + \mathbf{x}_n$.
- When multiplied by A, this is $A\mathbf{x} = A\mathbf{x}_r + A\mathbf{x}_n$: the null space component goes to zero: $A\mathbf{x}_n = 0$.
- **The row space component goes to the column space**:

$$A\mathbf{x}_r = A\mathbf{x}.$$

- Everything goes to the column space—the matrix cannot do anything else.
- From the row space to the column space, A is actually invertible. Every vector $\mathbf{b}$ in the column space comes from exactly one vector $\mathbf{x}_r$ in the row space.

Cosines and Projections onto lines

- Suppose we want to find the distance from a point $\mathbf{b}$ to the line in the direction of the vector $\mathbf{a}$.
- We are looking along that line for the point p closest to $\mathbf{b}$.
- ***The line connecting* b *to* p *is perpendicular to* a.**
- *This fact will allow us to find the projection* **p**.



- Similarly, given a plane (or any subspace $\boldsymbol{S}$) instead of a line, the problem is to find the point $\mathbf{p}$ on that subspace that is closest to $\mathbf{b}$. ***This point* $\mathbf{p}$ *is the projection of* $\mathbf{b}$ *onto the subspace*.**
- A perpendicular line from $\mathbf{b}$ to $\boldsymbol{S}$ meets the subspace at $\mathbf{p}$. Geometrically, that gives the distance between points $\mathbf{b}$ and subspaces $\boldsymbol{S}$.
- This is exactly the problem of the ***least-squares solution*** *to an* ***overdetermined system***.
- The vector $\mathbf{b}$ represents the data from experiments or questionnaires, and it contains too many errors to be found in the subspace $\boldsymbol{S}$.
- When we try to write $\mathbf{b}$ as a combination of the basis vectors for $\boldsymbol{S}$, it cannot be done—the equations are inconsistent, and $A\mathbf{x} = \mathbf{b}$ has no solution.
- The least-squares method selects $\mathbf{p}$ as the best choice to replace $\mathbf{b}$.

Projection onto a line

- We want to find the projection point $\mathbf{p}$. This point must be some multiple $\mathbf{p} = \hat{\mathbf{x}}\mathbf{a}$ of the given vector $\mathbf{a}$—every point on the line is a multiple of $\mathbf{a}$.
- The problem is to compute the coefficient $\hat{\mathbf{x}}$. All we need is the geometrical fact that *the line from* $\mathbf{b}$ *to the closest point* $\mathbf{p} = \hat{\mathbf{x}}\mathbf{a}$ *is perpendicular to the vector* $\mathbf{a}$:

$$(\mathbf{b} - \hat{\mathbf{x}}\mathbf{a}) \perp \mathbf{a}, \text{ or } \mathbf{a}^{\mathrm{T}}(\mathbf{b} - \hat{\mathbf{x}}\mathbf{a}) = 0, \text{ or } \hat{\mathbf{x}} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}$$

  - That gives the formula for the number $\hat{\mathbf{x}}$ and the projection $\mathbf{p}$:
  - **Projection onto a line** $\mathbf{p} = \hat{\mathbf{x}}\mathbf{a} = \frac{\mathbf{a}^{\mathrm{T}}\mathbf{b}}{\mathbf{a}^{\mathrm{T}}\mathbf{a}}\mathbf{a}$

Projection Matrix of Rank 1

- Projection onto a line is carried out by a *projection matrix* P, and written in this new order we can see what it is. P *is the matrix that multiplies* $\mathbf{b}$ *and produces* $\mathbf{p}$:

$$P = \mathbf{a}\frac{\mathbf{a^T b}}{\mathbf{a^T a}} \text{ so the projection matrix is } P = \frac{\mathbf{aa^T}}{\mathbf{a^T a}}$$

- **Example:** The matrix that projects onto the line through $\mathbf{a} = (1; 1; 1)$ is:

$$P = \frac{\mathbf{aa^T}}{\mathbf{a^T a}} = \frac{1}{3}\begin{bmatrix}1\\1\\1\end{bmatrix}[1\ 1\ 1] = \begin{bmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{bmatrix}$$

- This matrix has two properties that we will see as typical of projections:
  1. P **is a symmetric matrix.**
  2. **Its square is itself:** $P^2 = P$.
- **Example**: Project onto the "θ-direction" in the $x - y$ plane. The line goes through $a = (\cos\theta\ ; \sin\theta\ )$ and the matrix is symmetric with $P^2 = P$:

$$P = \frac{\mathbf{a}\mathbf{a}^{\mathbf{T}}}{\mathbf{a}^{\mathbf{T}}\mathbf{a}} = \frac{\begin{bmatrix} c \\ s \end{bmatrix} [c\ s]}{[c\ s] \begin{bmatrix} c \\ s \end{bmatrix}} = \begin{bmatrix} c^2 & cs \\ cs & s^2 \end{bmatrix}$$

- Here c is $\cos\theta$, s is $\sin\theta$, and $c^2 + s^2 = 1$ in the denominator.
- **To project b onto a, multiply by the projection matrix** P: $\mathbf{p} = P\mathbf{b}$.

Problem. Find the distance from the point $\mathbf{x} = (3, 1)$ to the line spanned by $\mathbf{y} = (2, -1)$.

Consider the decomposition $\mathbf{x} = \mathbf{p} + \mathbf{o}$, where $\mathbf{p}$ is parallel to $\mathbf{y}$ while $\mathbf{o} \perp \mathbf{y}$. The required distance is the length of the orthogonal component $\mathbf{o}$.

$$\mathbf{p} = \frac{\mathbf{x}\,.\,\mathbf{y}}{\mathbf{y}\,.\,\mathbf{y}}\mathbf{y} = \frac{5}{5}(2, -1) = (2, -1)$$

$$\mathbf{o} = \mathbf{x} - \mathbf{p} = (3,1) - (2, -1) = (1,2),$$

$$||\mathbf{o}|| = \sqrt{5}$$

Problem. Find the point on the line $y = -x$ that is closest to the point $(3, 4)$.

The required point is the projection $\mathbf{p}$ of $\mathbf{v} = (3, 4)$ on the vector $\mathbf{w} = (1, -1)$ spanning the line $y = -x$.

$$\mathbf{p} = \frac{\mathbf{v}\,.\,\mathbf{w}}{\mathbf{w}\,.\,\mathbf{w}} w = \frac{-1}{2}(1, -1) = \left(-\frac{1}{2}, \frac{1}{2}\right)$$