MA580H Matrix Computations

Lectures 3 & 4: Orthogonal vectors and matrices

Rafikul Alam
Department of Mathematics
IIT Guwahati

Outline

- Orthogonal vectors and orthogonal subspaces
- Orthogonal matrices
- Orthogonal decomposition theorem

Inner product

Angle between two n-vectors can be described by using inner product (dot product).

Definition: If $\mathbf{u} := [u_1, \ldots, u_n]^\top$ and $\mathbf{v} := [v_1, \ldots, v_n]^\top$ are n-vectors then the inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ is defined by

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &:= u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \mathbf{v}^\top \mathbf{u} \quad \text{when } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n \\ \langle \mathbf{u}, \mathbf{v} \rangle &:= u_1 \bar{v}_1 + u_2 \bar{v}_2 + \cdots + u_n \bar{v}_n = \mathbf{v}^* \mathbf{u} \quad \text{when } \mathbf{u}, \mathbf{v} \in \mathbb{C}^n. \end{aligned}$$

The inner product $\langle \mathbf{u}, \mathbf{v} \rangle$ is also called dot product and is written as $\mathbf{u} \bullet \mathbf{v}$.

Example: If $\mathbf{u} := [1, 2, -3]^\top$ and $\mathbf{v} := [-3, 5, 2]^\top$ then

$$\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot (-3) + 2 \cdot 5 + (-3) \cdot 2 = 1.$$

Inner product

Weights, features, and score. Let $\mathbf{f} := \begin{bmatrix} f_1 & \cdots & f_n \end{bmatrix} \in \mathbb{R}^n$ be a feature vector of an object and $\mathbf{w} := \begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} \in \mathbb{R}^n$ be a weight vector. Then the inner product

$$\langle \mathbf{f}, \mathbf{w} \rangle = w_1 f_1 + \cdots + w_n f_n$$

is the sum of the feature values, scaled by the weights, and is called a score.

Examples:

- Credit score: Let f be a feature vector associated with a loan applicant (e.g., age, income, . . .). Then we might interpret $\langle \mathbf{f}, \mathbf{w} \rangle$ as a credit score, where w_i is the weight given to feature f_i in forming the score.
- Co-occurrence. Let $\mathbf{x}$ and $\mathbf{y}$ be Boolean n-vectors (each entry is either 0 or 1) that describe occurrence. Then the inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ gives the total number of co-occurrences.
  For $\mathbf{x} := [0, 1, 1, 1, 1, 1, 1]^\top$ and $\mathbf{y} := [1, 0, 1, 0, 1, 0, 0]^\top$, we have $\langle \mathbf{x}, \mathbf{y} \rangle = 2$, which is the number of common occurrences.

Properties of inner product

Theorem: Let $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ be vectors in $\mathbb{C}^n$ and let $\alpha \in \mathbb{C}$. Then

1. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = \mathbf{0}$.
2. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$
3. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
4. $\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle$.

Definition: The norm (or length) of a vector $\mathbf{v} := [v_1, \ldots, v_n]^\top$ in $\mathbb{C}^n$ is a nonnegative number $\|\mathbf{v}\|$ defined by

$$\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle} = \sqrt{|v_1|^2 + \cdots + |v_n|^2}.$$

Theorem (Cauchy-Schwarz Inequality): Let $\mathbf{u}$ and $\mathbf{v}$ be n-vectors. Then

$$|\langle \mathbf{u}, \mathbf{v} \rangle| \leq \|\mathbf{u}\| \, \|\mathbf{v}\|.$$

Proof for $\mathbb{R}^n$: $p(t) := \|\mathbf{u} + t\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2t\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 t^2 \geq 0$ for all $t \in \mathbb{R}$. Hence discriminant of $p(t)$ is non-positive which yields the result. ■

Unit vectors

Definition: A vector $\mathbf{v}$ in $\mathbb{C}^n$ or $\mathbb{R}^n$ is called a unit vector if $\|\mathbf{v}\| = 1$. If $\mathbf{u}$ is a nonzero vector then $\mathbf{v} := \frac{1}{\|\mathbf{u}\|}\mathbf{u}$ is a unit vector in the direction of $\mathbf{u}$. Indeed,

$$\|\mathbf{v}\| = \|(1/\|\mathbf{u}\|)\mathbf{u}\| = \frac{1}{\|\mathbf{u}\|}\|\mathbf{u}\| = 1.$$

The vector $\mathbf{v}$ is referred to as a normalization of $\mathbf{u}$.

Example: Let $\mathbf{u} := \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$. Then $\|\mathbf{u}\| = \sqrt{4+1+9} = \sqrt{14}$ and

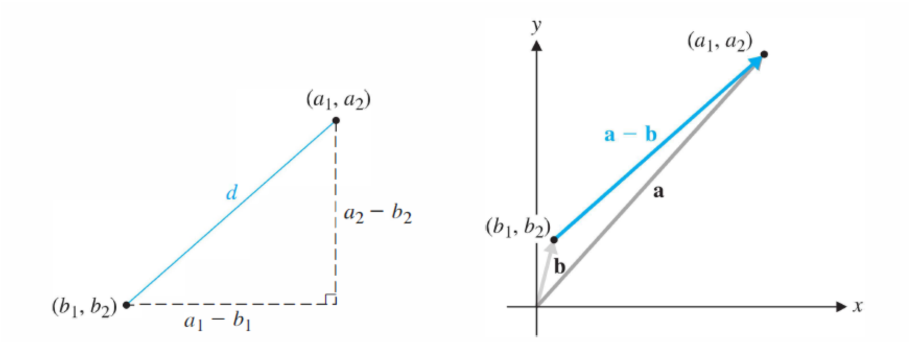
$$\mathbf{v} := \frac{1}{\sqrt{14}}\mathbf{u} = \frac{1}{\sqrt{14}}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 2/\sqrt{14} \\ -1/\sqrt{14} \\ 3/\sqrt{14} \end{bmatrix}.$$

Standard unit vectors: The vectors $\mathbf{e}_1 := [1,0,0]^\top$, $\mathbf{e}_2 := [0,1,0]^\top$ and $\mathbf{e}_3 := [0,0,1]^\top$ are unit vectors in $\mathbb{R}^3$ and are called standard unit vectors. The canonical vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$ are standard unit vectors.

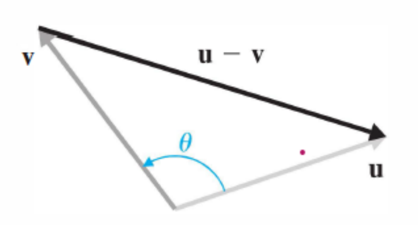
Distance

Distance: The distance $\mathrm{d}(\mathbf{u},\mathbf{v})$ between two vectors $\mathbf{u} := [u_1,\ldots,u_n]^\top$ and $\mathbf{v} := [v_1,\ldots,v_n]^\top$ in $\mathbb{R}^n$ or $\mathbb{C}^n$ is defined by

$$\mathrm{d}(\mathbf{u},\mathbf{v}) := \|\mathbf{u}-\mathbf{v}\| = \sqrt{|u_1-v_1|^2+\cdots+|u_n-v_n|^2}.$$



Angle between two vectors in $\mathbb{R}^2$



Consider the triangle in $\mathbb{R}^2$ with sides $\mathbf{u}, \mathbf{v}$ and $\mathbf{u} - \mathbf{v}$. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Then by the law of cosines there is unique $\theta \in [0, \pi]$ such that

$$\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\|\mathbf{u}\|\,\|\mathbf{v}\| \cos\theta.$$

Expanding $\|\mathbf{u} - \mathbf{v}\|^2 = \|\mathbf{u}\|^2 - 2\langle \mathbf{u}, \mathbf{v}\rangle + \|\mathbf{v}\|^2$ gives us

$$\langle \mathbf{u}, \mathbf{v}\rangle = \|\mathbf{u}\|\,\|\mathbf{v}\| \cos\theta \implies \cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|}.$$

Angle between two n-vectors

Definition: Let $\mathbf{u}$ and $\mathbf{v}$ be nonzero n-vectors. Then

$$\cos\theta \;:=\; \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \Longrightarrow \langle \mathbf{u}, \mathbf{v}\rangle = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta \;\text{ for } \theta \in [0,\pi] \text{ when } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n$$

$$\cos\theta \;:=\; \frac{|\langle \mathbf{u}, \mathbf{v}\rangle|}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \Longrightarrow |\langle \mathbf{u}, \mathbf{v}\rangle| = \|\mathbf{u}\|\,\|\mathbf{v}\|\cos\theta \;\text{ for } \theta \in [0,\frac{\pi}{2}] \text{ when } \mathbf{u}, \mathbf{v} \in \mathbb{C}^n.$$

Example: Let $\mathbf{u} := \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{v} := \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$. Then $\langle \mathbf{u}, \mathbf{v}\rangle = 1\cdot 0 + 0 \cdot 1 + 1 \cdot 1 = 1$. We have $\|\mathbf{u}\| = \sqrt{1+1} = \sqrt{2}$ and $\|\mathbf{v}\| = \sqrt{1+1} = \sqrt{2}$. Hence

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} = \frac{1}{\sqrt{2}\sqrt{2}} = \frac{1}{2} \Longrightarrow \theta = \pi/3 \;\text{ radians.} \;\blacksquare$$

Orthogonal vectors

Definition: Two n-vectors $\mathbf{u}$ and $\mathbf{v}$ are said to be mutually orthogonal if $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. We write $\mathbf{u} \perp \mathbf{v}$ when $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. If, in addition, $\|\mathbf{u}\| = \|\mathbf{v}\| = 1$ then $\mathbf{u}$ and $\mathbf{v}$ are called orthonormal.

Remark: The zero vector $\mathbf{0}$ is orthogonal to all vectors in $\mathbb{R}^n$ as $\langle \mathbf{0}, \mathbf{v} \rangle = 0$ for all $\mathbf{v} \in \mathbb{R}^n$.

Example: The vectors $\mathbf{u} := [1, 1, -2]^\top$ and $\mathbf{v} := [3, 1, 2]^\top$ in $\mathbb{R}^3$ are orthogonal as $\langle \mathbf{u}, \mathbf{v} \rangle = 1 \cdot 3 + 1 \cdot 1 + (-2) \cdot 2 = 0$.

Pythagoras' Theorem: Let $\mathbf{u}$ and $\mathbf{v}$ be n-vectors. Then

$$\begin{aligned} \langle \mathbf{u}, \mathbf{v} \rangle &= 0 \iff \|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \text{ when } \mathbf{u}, \mathbf{v} \in \mathbb{R}^n \\ \langle \mathbf{u}, \mathbf{v} \rangle &= 0 \implies \|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \text{ when } \mathbf{u}, \mathbf{v} \in \mathbb{C}^n \end{aligned}$$

Proof: We have $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\langle \mathbf{u}, \mathbf{v} \rangle + \|\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 \iff \langle \mathbf{u}, \mathbf{v} \rangle = 0$ when $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$. We have $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + 2\,\mathrm{Re}(\langle \mathbf{u}, \mathbf{v} \rangle) + \|\mathbf{v}\|^2$ when $\mathbf{u}, \mathbf{v} \in \mathbb{C}^n$. Hence $\langle \mathbf{u}, \mathbf{v} \rangle = 0 \implies \|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$. ■

Vectors in information retrieval

Problem: Given a few key words, retrieve relevant information from a large database.

Document vectors: Document vectors are used in information retrieval. Consider the five documents.

Doc. 1:	The Google matrix G is a model of the Internet.
Doc. 2:	G_{ij} is nonzero if there is a link from web page j to i.
Doc. 3:	The Google matrix G is used to rank all web pages.
Doc. 4:	The ranking is done by solving a matrix eigenvalue problem.
Doc. 5:	England dropped out of the top 10 in the FIFA ranking.

The blue colored texts are the key words or terms. The set of terms is called a Dictionary. Counting the frequency of terms in each document, we obtain document vectors.

Term-document matrix

Term	Doc. 1	Doc. 2	Doc. 3	Doc. 4	Doc. 5
eigenvalue	0	0	0	1	0
England	0	0	0	0	1
FIFA	0	0	0	0	1
Google	1	0	1	0	0
Internet	1	0	0	0	0
link	0	1	0	0	0
matrix	1	0	1	1	0
page	0	1	1	0	0
rank	0	0	1	1	1
web	0	1	1	0	1

Each document is a vector in $\mathbb{R}^{10}$ and is represented by a column of the term-document matrix.

Query vector

Suppose that we want to find all documents that are relevant to the query ranking of web pages. This is represented by a query vector, constructed in the way as the document vectors, using the same dictionary:

$$\mathbf{v} := \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}^\top \in \mathbb{R}^{10}.$$

Thus the query itself is a document. The information retrieval task can now be formulated as a mathematical problem.

Problem: Find the document vectors (columns of the term of document matrix) that are close (in some sense) to the query vector $\mathbf{v}$.

Query matching (use of dot product)

Query matching is the process of finding all documents that are relevant to a particular query $\mathbf{v}$. The cosine of angle between two vectors is often used to determine relevant documents:

$$\cos\theta_j := \frac{|\langle \mathbf{d}_j, \mathbf{v}\rangle|}{\|\mathbf{v}\|\,\|\mathbf{d}_j\|} > \mathrm{tol}$$

where $\mathbf{d}_j$ is the j-th document vector (j-th column of the term-document matrix) and tol is a predefined tolerance. Thus $\cos\theta_j > \mathrm{tol} \Rightarrow \mathbf{d}_j$ is relevant.

For the document vectors $\mathbf{d}_1, \ldots, \mathbf{d}_5$ and the query ("ranking of web pages") vector $\mathbf{v}$, the cosines measures of the query and the original data are given by

$$[0,\ 0.6667,\ 0.7746,\ 0.3333,\ 0.3333]^\top$$

which shows that Doc 2 and Doc 3 are most relevant.

Orthogonality in $\mathbb{C}^n$

Let $\mathbf{u} := [u_1, \ldots, u_n]^\top$ and $\mathbf{v} := [v_1, \ldots, v_n]^\top$ be vectors in $\mathbb{C}^n$. Then recall that

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 \bar{v}_1 + \cdots + u_n \bar{v}_n = \mathbf{v}^* \mathbf{u} \text{ and } |\langle \mathbf{u}, \mathbf{v} \rangle| = \|\mathbf{u}\| \, \|\mathbf{v}\| \cos\theta, \text{ where } \theta \in [0, \pi/2].$$

If $\langle \mathbf{u}, \mathbf{v} \rangle = 0$ then $\mathbf{u}$ and $\mathbf{v}$ are called mutually orthogonal and is written as $\mathbf{u} \perp \mathbf{v}$.

Definition: A set of vectors $S := \{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset \mathbb{C}^n$ is called an orthogonal set if the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are mutually orthogonal, that is, $\langle \mathbf{u}_i, \mathbf{u}_j \rangle = 0$ for all $i \neq j$. If S is an orthogonal set then the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_m$ are called orthogonal vectors.

If S is an orthogonal set and $\|\mathbf{u}_j\| = 1$ for $j = 1 : m$ then S is called an orthonormal set (ONS) and the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ are called orthonormal vectors. ■

Example: The standard unit vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ in $\mathbb{R}^n$ are orthonormal vectors. The vectors $\mathbf{u}_1 := [2, 1, -1]^\top, \mathbf{u}_2 := [0, 1, 1]^\top, \mathbf{u}_3 := [1, -1, 1]^\top$ in $\mathbb{R}^3$ are orthogonal vectors.

Orthonormal basis

Fact: If $S := \{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset \mathbb{C}^n$ is an orthonormal set then S is linearly independent.

Proof: $c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m = \mathbf{0} \Longrightarrow c_j = \mathbf{u}_j^*(c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m) = 0$ for $j = 1 : m$. ■

Definition: Let $\mathcal{V}$ be a subspace of $\mathbb{C}^n$ and $\mathcal{B} := \{\mathbf{u}_1, \ldots, \mathbf{u}_m\} \subset \mathcal{V}$. Then $\mathcal{B}$ is called an orthonormal basis (ONB) of $\mathcal{V}$ if $\mathcal{B}$ is an orthonormal set and $\mathrm{span}(\mathcal{B}) = \mathcal{V}$.

If $\mathcal{B}$ is an orthogonal set and is a basis of $\mathcal{V}$ then $\mathcal{B}$ is called an orthogonal basis of $\mathcal{V}$. ■

Example: The vectors $\mathbf{u}_1 := [2, 1, -1]^\top, \mathbf{u}_2 := [0, 1, 1]^\top, \mathbf{u}_3 := [1, -1, 1]^\top$ in $\mathbb{R}^3$ are orthogonal and linearly independent. Hence $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ is an orthogonal basis of $\mathbb{R}^3$.

Theorem: Let $\mathcal{V}$ be a subspace of $\mathbb{C}^n$ and $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be an ONB of $\mathcal{V}$. Let $\mathbf{v} \in \mathcal{V}$. Then $\mathbf{v}$ can be expressed uniquely as

$$\mathbf{v} = \langle \mathbf{v}, \mathbf{u}_1\rangle \mathbf{u}_1 + \cdots + \langle \mathbf{v}, \mathbf{u}_m\rangle \mathbf{u}_m = \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_m\end{bmatrix} \begin{bmatrix}\mathbf{u}_1 & \cdots & \mathbf{u}_m\end{bmatrix}^* \mathbf{v}.$$

Proof: There exist unique scalars $c_1, \ldots, c_m$ in $\mathbb{C}$ such that $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m$ $\Longrightarrow \mathbf{u}_j^*\mathbf{v} = \mathbf{u}_j^*(c_1\mathbf{u}_1 + \cdots + c_m\mathbf{u}_m) = c_j \Longrightarrow c_j = \mathbf{u}_j^*\mathbf{v} = \langle \mathbf{v}, \mathbf{u}_j\rangle$ for $j = 1 : m$. ■

Unitary and orthogonal matrices

Definition: A matrix $U \in \mathbb{C}^{n\times n}$ is called unitary if $U^*U = UU^* = I_n$. A matrix $V \in \mathbb{C}^{m\times n}$ is called an isometry if $V^*V = I_n$. A matrix $Q \in \mathbb{R}^{n\times n}$ is called an orthogonal matrix if $Q^\top Q = QQ^\top = I_n$.

Remark: A matrix $Q \in \mathbb{R}^{n\times n}$ is orthogonal if and only if $Q^\top = Q^{-1}$.

Fact: A matrix $U \in \mathbb{C}^{m\times n}$ is an isometry $\iff$ columns of U are orthonormal.

Proof: If $U := \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$ then $U^*U = [\mathbf{u}_i^*\mathbf{u}_j]_{n\times n} = I_n \iff \langle \mathbf{u}_j, \mathbf{u}_i\rangle = \mathbf{u}_i^*\mathbf{u}_j = \delta_{ij}$. ■

Example: The rotation matrix $A := \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is an orthogonal matrix.

Theorem: Let $U \in \mathbb{C}^{n\times n}$. Then the following statements are equivalent.

(a) U is unitary.

(b) $\|U\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{C}^n$.

(c) $\langle U\mathbf{x}, U\mathbf{y}\rangle = \langle \mathbf{x}, \mathbf{y}\rangle$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{C}^n$.

Orthogonal subspaces in $\mathbb{R}^n$

Definition: Two subspaces $\mathcal{X}$ and $\mathcal{Y}$ of $\mathbb{R}^n$ are said to be orthogonal if $\mathbf{y}^\top \mathbf{x} = 0$ for all $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$. We write $\mathcal{X} \perp \mathcal{Y}$ when $\mathcal{X}$ and $\mathcal{Y}$ are orthogonal. In particular, we write $\mathbf{x} \perp \mathcal{Y}$ when $\mathbf{y}^\top \mathbf{x} = 0$ for all $\mathbf{y} \in \mathcal{Y}$. ■

Consider $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ in $\mathbb{R}^3$. Let $\mathcal{X} := \mathrm{span}(\mathbf{e}_1, \mathbf{e}_2)$ and $\mathcal{Y} := \mathrm{span}(\mathbf{e}_3)$. Then $\mathcal{X} \perp \mathcal{Y}$.

Let $A \in \mathbb{R}^{m \times n}$ and let $A^\top = \begin{bmatrix} \mathbf{y}_1 & \cdots & \mathbf{y}_m \end{bmatrix}$. Then $A\mathbf{x} = \begin{bmatrix} \mathbf{y}_1^\top \\ \vdots \\ \mathbf{y}_m^\top \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{y}_1^\top \mathbf{x} \\ \vdots \\ \mathbf{y}_m^\top \mathbf{x} \end{bmatrix}$.

Consider $N(A) := \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = \mathbf{0}\} \subset \mathbb{R}^n$ and $R(A) := \{A\mathbf{x} : \mathbf{x} \in \mathbb{R}^n\} \subset \mathbb{R}^m$. Then $\mathbf{x} \in N(A) \iff A\mathbf{x} = \mathbf{0} \iff \mathbf{y}_j^\top \mathbf{x} = 0$ for $j = 1 : m \iff \mathbf{x} \perp R(A^\top)$.

Fact: Let $A \in \mathbb{R}^{m \times n}$. Then $N(A)$ and $R(A^\top)$ are mutually orthogonal subspaces of $\mathbb{R}^n$, that is, $N(A) \perp R(A^\top)$. Similarly, $N(A^\top) \perp R(A)$.

Orthogonal Decomposition Theorem

Definition: Let $\mathcal{X}$ be a subspace of $\mathbb{R}^n$. Define $\mathcal{X}^\perp := \{\mathbf{y} \in \mathbb{R}^n : \mathbf{y} \perp \mathcal{X}\}$. The set $\mathcal{X}^\perp$ is called the orthogonal complement of $\mathcal{X}$. ■

Fact: If $\mathcal{X}$ is a subspace of $\mathbb{R}^n$ then $\mathcal{X}^\perp$ is a subspace of $\mathbb{R}^n$ and $\mathcal{X} \cap \mathcal{X}^\perp = \{\mathbf{0}\}$.

Theorem: Let $\mathcal{X}$ be a subspace of $\mathbb{R}^n$ and let $\mathbf{v} \in \mathbb{R}^n$. Then there exist unique $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{X}^\perp$ such that $\mathbf{v} = \mathbf{x} + \mathbf{y}$. Equivalently, $\mathbb{R}^n = \mathcal{X} \oplus \mathcal{X}^\perp$.

Proof: Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be an orthonormal basis of $\mathcal{X}$. Let $\mathbf{v} \in \mathbb{R}^n$. Define $\mathbf{x} := \langle \mathbf{v}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \cdots + \langle \mathbf{v}, \mathbf{u}_m \rangle \mathbf{u}_m$ and $\mathbf{y} := \mathbf{v} - \mathbf{x}$. Then $\mathbf{v} = \mathbf{x} + \mathbf{y}$ and $\mathbf{x} \in \mathcal{X}$.

Note that $\langle \mathbf{y}, \mathbf{u}_j \rangle = \langle \mathbf{v}, \mathbf{u}_j \rangle - \langle \mathbf{x}, \mathbf{u}_j \rangle = 0$ for $j = 1 : m \Longrightarrow \mathbf{y} \perp \mathcal{X} \Longrightarrow \mathbf{y} \in \mathcal{X}^\perp$.
Thus $\mathbf{v} = \mathbf{x} + \mathbf{y}$ with $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{X}^\perp$. Since $\mathcal{X} \cap \mathcal{X}^\perp = \{\mathbf{0}\}$, the result follows. ■

Fact: Let $\mathcal{X}$ be a supspace of $\mathbb{R}^n$. Then $(\mathcal{X}^\perp)^\perp = \mathcal{X}$.

Proof: $\mathcal{X} \perp \mathcal{X}^\perp \Longrightarrow \mathcal{X} \subset (\mathcal{X}^\perp)^\perp$. Let $\mathbf{v} \in (\mathcal{X}^\perp)^\perp$. By projection theorem, $\mathbf{v} = \mathbf{x} + \mathbf{y}$ with $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{X}^\perp \Longrightarrow \mathbf{y} \perp \{\mathbf{x}, \mathbf{v}\} \Longrightarrow 0 = \mathbf{y}^\top \mathbf{v} = \mathbf{y}^\top \mathbf{x} + \mathbf{y}^\top \mathbf{y} = \mathbf{y}^\top \mathbf{y} \Longrightarrow \mathbf{y} = \mathbf{0}$.
Hence $\mathbf{v} = \mathbf{x} \in \mathcal{X} \Longrightarrow (\mathcal{X}^\perp)^\perp \subset \mathcal{X}$. ■

Fundamental Subspace Theorem

Remark: The orthogonal decomposition theorem is also called Projection Theorem.

Theorem: Let $A \in \mathbb{R}^{m\times n}$. Then $N(A)^\perp = R(A^\top)$ and $N(A^\top) = R(A)^\perp$. Further,

$$\begin{aligned} \mathbb{R}^n &= N(A) \oplus R(A^\top) \text{ and } N(A) \perp R(A^\top), \\ \mathbb{R}^m &= N(A^\top) \oplus R(A) \text{ and } N(A^\top) \perp R(A). \end{aligned}$$

Proof: We have seen that $N(A) \perp R(A^\top)$ which imples that $N(A) \subset R(A^\top)^\perp$. Now, $\mathbf{x} \in R(A^\top)^\perp \Longrightarrow \mathbf{x} \perp R(A^\top) \Longrightarrow \mathbf{x} \perp A^\top \mathbf{e}_j$ for $j = 1 : m \Longrightarrow (A^\top \mathbf{e}_j)^\top \mathbf{x} = \mathbf{e}_j^\top A\mathbf{x} = \mathbf{0}$ for $j = 1 : m \Longrightarrow A\mathbf{x} = \mathbf{0} \Longrightarrow \mathbf{x} \in N(A) \Longrightarrow R(A^\top)^\perp \subset N(A)$.

This proves $N(A) = R(A^\top)^\perp$. Now replacing A with $A^\top$ yields $N(A^\top) = R(A)^\perp$. Finally, by orthogonal decomposition theorem

$$\begin{aligned} \mathbb{R}^n &= N(A) \oplus N(A)^\perp = N(A) \oplus (R(A^\top)^\perp)^\perp = N(A) \oplus R(A^\top), \\ \mathbb{R}^m &= N(A^\top) \oplus N(A^\top)^\perp = N(A^\top) \oplus (R(A)^\perp)^\perp = N(A^\top) \oplus R(A). \blacksquare \end{aligned}$$

Remark: The subspaces $R(A), N(A), R(A^\top)$ and $N(A^\top)$ are called four fundamental subspaces of an $m \times n$ matrix A.

Orthogonalization

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be linearly independent vectors in $\mathbb{C}^n$ such that $\mathbf{v}_2^*\mathbf{v}_1 \neq 0$. We wish to construct orthonormal vectcors $\mathbf{u}_1$ and $\mathbf{u}_2$ such that

$$\mathrm{span}(\mathbf{v}_1) = \mathrm{span}(\mathbf{u}_1) \text{ and } \mathrm{span}(\mathbf{v}_1, \mathbf{v}_2) = \mathrm{span}(\mathbf{u}_1, \mathbf{u}_2).$$

Set $\mathbf{u}_1 := \dfrac{\mathbf{v}_1}{\|\mathbf{v}_1\|}$. Then $\mathrm{span}(\mathbf{v}_1) = \mathrm{span}(\mathbf{u}_1)$. Next, choose $\alpha \in \mathbb{C}$ such that $(\mathbf{v}_2 - \alpha\mathbf{u}_1) \perp \mathbf{u}_1$. This gives $\langle \mathbf{v}_2 - \alpha\mathbf{u}_1, \mathbf{u}_1 \rangle = 0 \Longrightarrow \langle \mathbf{v}_2, \mathbf{u}_1 \rangle = \alpha$. Thus $(\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1) \perp \mathbf{u}_1$.

Define $\mathbf{u}_2 := \dfrac{\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1}{\|\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1\|}$. Then $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthonormal and $\mathbf{u}_1, \mathbf{u}_2 \in \mathrm{span}(\mathbf{v}_1, \mathbf{v}_2)$. Now

$$\mathbf{v}_2 = (\|\mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1\|)\mathbf{u}_2 + \langle \mathbf{v}_2, \mathbf{u}_1 \rangle \mathbf{u}_1 \in \mathrm{span}(\mathbf{u}_1, \mathbf{u}_2) \Longrightarrow \mathbf{v}_1, \mathbf{v}_2 \in \mathrm{span}(\mathbf{u}_1, \mathbf{u}_2).$$

This shows that $\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2) = \mathrm{span}(\mathbf{u}_1, \mathbf{u}_2)$.

If $\mathbf{v}_3$ is another vector then defne $\mathbf{u}_3 := \dfrac{\mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{v}_3, \mathbf{u}_2 \rangle \mathbf{u}_2}{\|\mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{v}_3, \mathbf{u}_2 \rangle \mathbf{u}_2\|}$. Then $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are orthonormal and $\mathrm{span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \mathrm{span}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$.

Gram-Schmidt Orthogonalization

Let $\mathbf{v}_1, \dots, \mathbf{v}_m$ be linearly indepedent vectors in $\mathbb{C}^n$. Then there exist orthonormal vectors $\mathbf{u}_1, \dots, \mathbf{u}_m$ in $\mathbb{C}^n$ such that

$$\operatorname{span}(\mathbf{v}_1, \dots, \mathbf{v}_j) = \operatorname{span}(\mathbf{u}_1, \dots, \mathbf{u}_j) \ \text{ for } \ j = 1 : m.$$

The Gram-Schmidt process constructs orthonormal vectors $\mathbf{u}_1, \dots, \mathbf{u}_m$ as follows. Define

$$\begin{aligned}
\mathbf{u}_1 &:= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \\
\mathbf{u}_j &:= \frac{\mathbf{v}_j - \langle \mathbf{v}_j, \mathbf{u}_1 \rangle \mathbf{u}_1 - \cdots - \langle \mathbf{v}_j, \mathbf{u}_{j-1} \rangle \mathbf{u}_{j-1}}{\|\mathbf{v}_j - \langle \mathbf{v}_j, \mathbf{u}_1 \rangle \mathbf{u}_1 - \cdots - \langle \mathbf{v}_j, \mathbf{u}_{j-1} \rangle \mathbf{u}_{j-1}\|}, \ j = 2 : m.
\end{aligned}$$

Note that $\|\mathbf{v}_j - \langle \mathbf{v}_j, \mathbf{u}_1 \rangle \mathbf{u}_1 - \cdots - \langle \mathbf{v}_j, \mathbf{u}_{j-1} \rangle \mathbf{u}_{j-1}\| \neq 0 \iff$ the vectors $\mathbf{v}_1, \dots, \mathbf{v}_j$ are linearly independent. By induction $\operatorname{span}(\mathbf{v}_1, \dots, \mathbf{v}_j) = \operatorname{span}(\mathbf{u}_1, \dots, \mathbf{u}_j)$ for $j = 1 : n$.

QR factorization

Setting $r_{11} := \|\mathbf{v}_1\|$, $r_{jj} := \|\mathbf{v}_j - \langle \mathbf{v}_j, \mathbf{u}_1\rangle \mathbf{u}_1 - \cdots - \langle \mathbf{v}_j, \mathbf{u}_{j-1}\rangle \mathbf{u}_{j-1}\|$ and $r_{kj} := \langle \mathbf{v}_j, \mathbf{u}_k\rangle$, for $k = 1 : j-1$, we have

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{u}_1 r_{11}, \\ \mathbf{v}_j &= \langle \mathbf{v}_j, \mathbf{u}_1\rangle \mathbf{u}_1 + \cdots + \langle \mathbf{v}_j, \mathbf{u}_{j-1}\rangle \mathbf{u}_{j-1} + r_{jj}\mathbf{u}_j, \\ &= \mathbf{u}_1 r_{1j} + \cdots + \mathbf{u}_{j-1} r_{j-1,j} + r_{jj}\mathbf{u}_j, \quad j = 2 : m. \end{aligned}$$

Then, in matrix notation, we have

$$A := \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_m \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_m \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ & r_{22} & \cdots & r_{2m} \\ & & \ddots & \vdots \\ & & & r_{mm} \end{bmatrix} = QR.$$

Thus, if $A \in \mathbb{C}^{n\times m}$ and $\mathrm{rank}(A) = m$, then A has a QR factorization $A = QR$, where Q is an isometry and R is upper triangular and nonsingular.

Example

Consider $\mathbf{v}_1 := \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^\top$, $\mathbf{v}_2 := \begin{bmatrix} 2 & 1 & 0 \end{bmatrix}^\top$ and $\mathbf{v}_3 := \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}^\top$. Then by the Gram-Schmidt process, we have $r_{11} := \|\mathbf{v}_1\| = \sqrt{2}$ which gives

$$\mathbf{u}_1 := \frac{\mathbf{v}_1}{r_{11}} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}.$$

Next, we have $r_{12} := \mathbf{u}_1^\top \mathbf{v}_2 = \sqrt{2}$ and

$$\begin{aligned} \mathbf{q}_2 &:= \mathbf{v}_2 - (\mathbf{u}_1^\top \mathbf{v}_2)\mathbf{u}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix} - \frac{\sqrt{2}}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \\ \mathbf{u}_2 &:= \frac{\mathbf{q}_2}{r_{22}} = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \text{ where } r_{22} := \|\mathbf{q}_2\| = \sqrt{3}. \end{aligned}$$

Example

Finally, $r_{13} := \mathbf{u}_1^\top \mathbf{v}_3 = 1/\sqrt{2}$ and $r_{23} := \mathbf{u}_2^\top \mathbf{v}_3 = 0$. Hence we have

$$\begin{aligned}
\mathbf{q}_3 &:= \mathbf{v}_3 - (\mathbf{u}_1^\top \mathbf{v}_3)\mathbf{u}_1 - (\mathbf{u}_2^\top \mathbf{v}_3)\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \\
\mathbf{u}_3 &:= \frac{\mathbf{q}_3}{r_{33}} = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \quad \text{where } r_{33} := \|\mathbf{q}_3\| = \frac{\sqrt{6}}{2}.
\end{aligned}$$

Setting $A := \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}$ and $Q := \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$, we have the QR factorization of A

$$\underbrace{\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}}_{A} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{-1}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{6}} \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} \sqrt{2} & \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \frac{\sqrt{6}}{2} \end{bmatrix}}_{R}. \blacksquare$$