Orthogonality

Linear Algebra

Department of Computer Engineering

Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu

Maryam Ramezani maryam.ramezani@sharif.edu

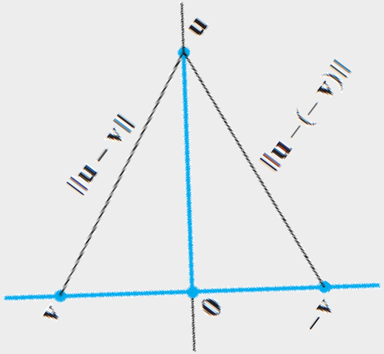
Overview

- Orthogonality
- Gram–Schmidt Algorithm
- Orthogonal Complements
- Projection

Orthogonality

Orthogonal vectors

- Geometry



https://youtu.be/dqdSzqsm7bY

- Algebra

Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = \mathbf{0}$.

Suppose V is an inner product space.
Two vectors $\mathbf{v}, \mathbf{w} \in V$ are called **orthogonal** if $\langle \mathbf{v}, \mathbf{w} \rangle = 0$.

The Pythagorean Theorem

Two vectors $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if and only if $||\mathbf{u} + \mathbf{v}||^2 = ||\mathbf{u}||^2 + ||\mathbf{v}||^2$

Orthogonal Sets

- A set of vectors $\{a_1, \ldots, a_k\}$ in R^n is orthogonal set if each pair of distinct vectors is orthogonal (mutually orthogonal vectors).

Definition

A basis B of an inner product space V is called an **orthonormal basis** of V if

a) $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{v} \neq \boldsymbol{w} \in B$, and (mutual orthogonality)

b) $\|\mathbf{v}\| = 1$ for all $\mathbf{v} \in B$. (normalization)

- set of n-vectors $a_1, \ldots, a_k$ are *(mutually) orthogonal* if $a_i \perp a_j$ for $i \neq j$
- They are *normalized* if $\|a_i\| = 1$ for $i = 1, \ldots, k$
- They are *orthonormal* if both hold
- Can be expressed using inner products as

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

Orthogonal Sets

Example

- ❑ Zero vector is orthogonal to every vector in vector space V
- ❑ The standard basis of $\mathbb{R}^n$ or $\mathbb{C}^n$ is an orthogonal set with respect to the standard inner product.

Orthogonal Sets

Theorem

If $S = \{a_1, \ldots, a_k\}$ is an orthogonal set of nonzero vectors in R^n, then S is linearly independent and is a basis for the subspace spanned by S.

Proof

If k = n, then prove that S is a basis for R^n

Linear combinations of orthonormal vectors

Corollary

- A simple way to check if an n−vector y is a linear combination of the orthonormal vectors $a_1, \dots, a_k$, if and only if:

$$y = (a_1^T y)a_1 + \dots + (a_k^T y)a_k$$

- For orthogonal vectors $a_1, \dots, a_k$:

$$y = c_1 a_1 + \cdots + c_k a_k$$

$$c_j = \frac{y . a_j}{a_j . a_j}$$

Orthonormal vectors

Independence-dimension inequality

If the n-vectors $a_1, \dots, a_k$ are linearly independent, then $k \leq n$.

- Orthonormal sets of vectors are linearly independent
- By independence–dimension inequality, must have $k \leq n$
- When $k = n, a_1, \dots, a_n$ are an *orthonormal basis*

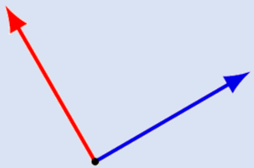
Orthonormal bases

Example

- Standard unit n-vectors $e_1, \ldots, e_n$
- The 3-vectors

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

- The 2-vectors shown below



- The standard basis in $P_n(x)\ [-1,1]$ (be the set of real-valued polynomials of degree at most n.)

Linear combinations of orthonormal vectors

Example

Write x as a linear combination of a_1, a_2, a_3?

$$x = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad a_1 = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}, a_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, a_3 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

Orthogonal Supspaces

Orthogonal Subspaces

Definition

- Two subspaces W_1 and W_2 of the same space V are orthogonal, denoted by $W_1 \perp W_2$, if and only if each vector $w_1 \in W_1$ is orthogonal to each vector $w_2 \in W_2$ for all w_1, w_2 in W_1, W_2 respectively:

$$< w_1, w_2 > = 0$$

Example

If the bases of two subspaces are orthogonal, it implies that the subspaces themselves are orthogonal.

Orthogonal Complements

Orthogonal Complements

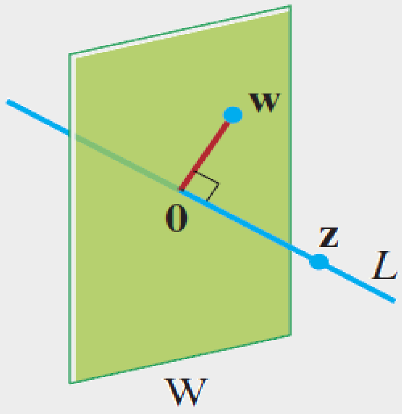
Definition

- If a vector z is orthogonal to every vector in a subspace W of R^n, then z is said to be orthogonal to W.
- **The set of all vectors z that are orthogonal** to W is called the orthogonal complement of W and is denoted by $W^{\perp}$

Example

W be a plane through the origin in $\mathbb{R}^3$.

$L = W^{\perp}$and $W = L^{\perp}$



Orthogonal Complements

Theorem

$W^{\perp}$ is a subspace of $\mathbb{R}^n$.

Theorem

$W^{\perp} \cap W = \{\mathbf{0}\}$.

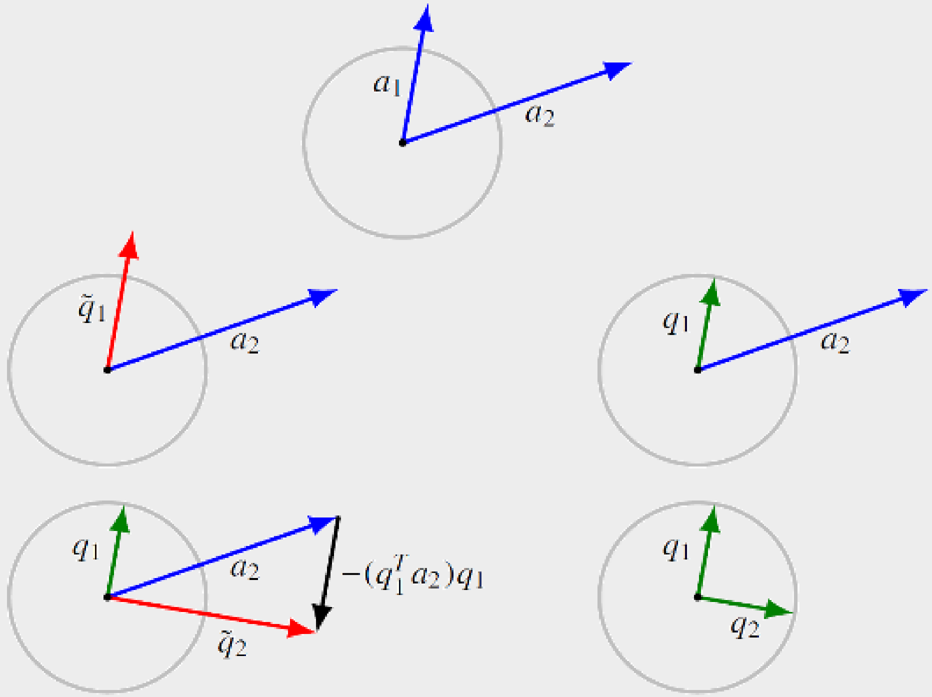
Important

We emphasize that W_1 and W_2 can be orthogonal without being complements.
$W_1 = span((1,0,0))$ and $W_2 = span((0,1,0))$.

Gram-Schmidt Algorithm

Gram–Schmidt (orthogonalization) algorithm

- Find orthonormal basis for span $\{a_1, a_2, \dots, a_k\}$
- Geometry:



Gram–Schmidt (orthogonalization) algorithm

- Find orthonormal basis for span $\{a_1, a_2, \dots, a_k\}$
- Algebra:

1) $q1 = \frac{a_1}{\|a_1\|}$

2) $\widetilde{q_2} = a_2 - (q_1^T a_2)q_1 \rightarrow q_2 = \frac{\widetilde{q_2}}{\|\widetilde{q_2}\|}$

3) $\widetilde{q_3} = a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2 \rightarrow q_3 = \frac{\widetilde{q_3}}{\|\widetilde{q_3}\|}$

.

.

.

k) $\widetilde{q_k} = a_k - (q_1^T a_k)q_1 - \cdots - \left(q_{k-1}^T a_k\right)q_{k-1} \rightarrow q_k = \frac{\widetilde{q_k}}{\|\widetilde{q_k}\|}$

Gram–Schmidt (orthogonalization) algorithm

Example

Find orthogonal set for $a = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, b = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathrm{c} = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$

Gram–Schmidt (orthogonalization) algorithm

- Why $\{q_1, q_2, \dots, q_k\}$ is a orthonormal basis for span $\{a_1, a_2, \dots, a_k\}$?

  - $\{q_1, q_2, \dots, q_k\}$ are normalized.
  - $\{q_1, q_2, \dots, q_k\}$ is a orthogonal set
  - a_i is a linear combination of $\{q_1, q_2, \dots, q_i\}$

  ⬇

  span$\{q_1, q_2, \dots, q_k\}$ = span$\{a_1, a_2, \dots, a_k\}$

- q_i is a linear combination of $\{a_1, a_2, \dots, a_i\}$

Gram–Schmidt (orthogonalization) algorithm

- Given n–vectors $a_1, \ldots, a_k$

for $i = 1, \ldots, k$

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

Note

- If G–S does not stop early (in step 2), $a_1, \ldots, a_k$ are linearly independent.
- If G–S stops early in iteration $i = j$, then a_j is a linear combination of $a_1, \ldots, a_{j-1}$ (so $a_1, \ldots, a_k$ are linearly dependent)

$$a_j = (q_1^T a_j)q_1 + \cdots + (q_{j-1}^T a_j)q_{j-1}$$

Complexity of Gram–Schmidt algorithm

- Gram–Schmidt algorithm gives us an explicit method for determining if a list of vectors is linearly dependent or independent.

- What is complexity and number of flops for this algorithm?
  - $O(nk^2)$ why?
- Given n–vectors $a_1, \ldots, a_k$ for $i = 1, \ldots, k$
  1. Orthogonalization: $\tilde{q}_i = a_i - (q_1^T a_i)q_1 - \cdots - (q_{i-1}^T a_i)q_{i-1}$
  2. Test for linear dependence: if $\tilde{q}_i = 0$, quit
  3. Normalization: $q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$

Orthonormal basis

Corollary

Every finite-dimensional inner product space has an orthonormal basis.

Conclusion

Existence of Orthonormal Bases

- Every finite-dimensional inner product space has an orthonormal basis.
- Since finite-dimensional inner product spaces (by definition) have a basis consisting of finitely many vectors, and the Gram-Schmidt process tells us how to convert that basis into an orthonormal basis, we now know that every finite-dimensional inner product space has an orthonormal basis.

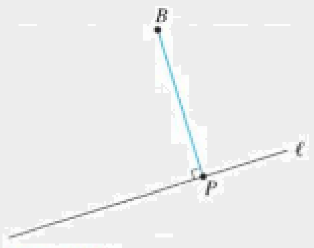
Example

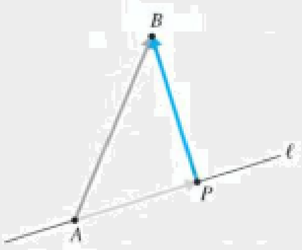
Find an orthonormal basis for $P_2(x)\ in\ [-1,1]$ with respect to the inner product

$$\langle f,g\rangle = \int_{-1}^{1} f(x)g(x)dx$$

Projection

- Finding the distance from a point B to line l = Finding the length of line segment BP
- AP: projection of AB onto the line l

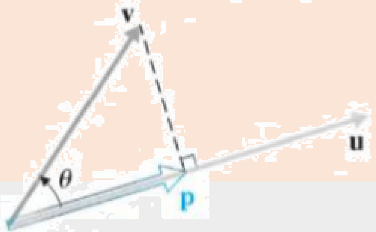




Definition

If **u** and **v** are vectors in $\mathbb{R}^n$ and $\mathbf{u} \neq \mathbf{0}$, then the **projection of v onto u** is the vector $proj_{\mathbf{u}}(\mathbf{v})$ defined by

$$proj_{\mathbf{u}}(\mathbf{v}) = \left(\frac{\mathbf{u} \cdot \mathbf{v}}{\mathbf{u} \cdot \mathbf{u}}\right)\mathbf{u}$$



The projection of **v** onto **u**

The Orthogonal Decomposition Theorem

Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written **uniquely** in the form:

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad (1)$$

$\hat{\mathbf{y}} \rightarrow \text{proj}_W \, \mathbf{y}$

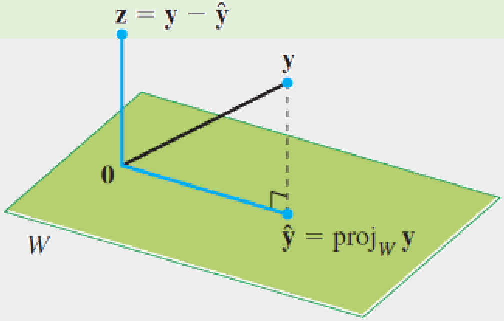
where $\hat{\mathbf{y}}$ is in W and **$\mathbf{z}$ is in $W^\perp$**. In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y} \cdot \mathbf{u}_1}{\mathbf{u}_1 . \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y} \cdot \mathbf{u}_p}{\mathbf{u}_p \cdot \mathbf{u}_p}\mathbf{u}_p \quad (2)$$

and $z = \mathbf{y} - \hat{\mathbf{y}}$

Important

The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W and not on the particular basis used in (2).



The orthogonal projection of $\mathbf{y}$ onto W.

Orthogonal Projection of y onto W

Theorem

Let W be a subspace of V. Then each **u** in V can be written in the form:

$$\mathbf{u} = \hat{\mathbf{y}} + \boldsymbol{y}$$

$\hat{\mathbf{y}}$ **projection u on W**

Proof

Orthogonal Decomposition Theorem

Theorem

Let W be a subspace of V. Then for any vector v in V, there exists a unique vector w in W, and a unique vector z in $W^{\perp}$, such that v = w + z. The vector w is called the orthogonal projection of v onto W.

Proof

Orthogonal Matrix

Note

- Columns of A are orthonormal $\leftrightarrow A^T A = I$

- Square matrix with orthonormal columns is a **orthogonal matrix**
  - Columns and rows are orthonormal vectors
  - $A^T A = AA^T = I$
  - is necessarily invertible with inverse $A^T = A^{-1}$

Orthogonal Matrix

Example

- Identity matrix $I^T I = I$

- Rotation matrix

$$R^T R = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} =$$

$$\begin{bmatrix} cos^2\theta + sin^2\theta & -cos\theta sin\theta + sin\theta cos\theta \\ -sin\theta cos\theta + cos\theta sin\theta & sin^2\theta + cos^2\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Orthogonal Matrix

Example

- Reflection matrix

$$\begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix}^T \begin{bmatrix} \cos(2\theta) & \sin(2\theta) \\ \sin(2\theta) & -\cos(2\theta) \end{bmatrix} =$$

$$\begin{bmatrix} \cos^2(2\theta) + \sin^2(2\theta) & cos(2\theta)\sin(2\theta) - \sin(2\theta)\cos(2\theta) \\ \sin(2\theta)\cos(2\theta) - \cos(2\theta)\sin(2\theta) & \sin^2(2\theta) + \cos^2(2\theta) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

Lemma

All orthogonal matrices can be expressed as Rotation or Reflection

Orthonormal Columns Properties

Note

If $A \in \mathbb{R}^{m\times n}$ has orthonormal columns, then the linear function $f(x) = Ax$

- Preserves inner product:

$$(Ax)^T(Ay) = x^T y$$

- Preserves norm:

$$\|Ax\| = \|x\|$$

- Preserves distances:

$$\|Ax - Ay\| = \|x - y\|$$

- Preserves angels:

$$\angle(Ax, Ay) = \arccos\left(\frac{(Ax)^T(Ay)}{\|Ax\|\|Ay\|}\right) = \arccos\left(\frac{x^T y}{\|x\|\|y\|}\right) = \angle(x, y)$$

This is a mapping with preserving properties of input

Gram–Schmidt in matrix notation

Important

Run Gram–Schmidt on columns $a_1, \ldots, a_k$ of $n \times k$ matrix A:

$$\tilde{q}_1 = a_1, \qquad q_1 = \frac{\tilde{q}_1}{\|\tilde{q}_1\|}$$

$$\Longrightarrow a_1 = \|\tilde{q}_1\| q_1$$

$$\tilde{q}_2 = a_2 - (q_1^T a_2) q_1, \qquad q_2 = \frac{\tilde{q}_2}{\|\tilde{q}_2\|}$$

$$\Rightarrow a_2 = (q_1^T a_2) q_1 + \|\tilde{q}_2\| q_2$$

$$\vdots$$

$$\tilde{q}_i = a_i - (q_1^T a_i) q_1 - \cdots - \left(q_{i-1}^T a_i\right) q_{i-1}, \qquad q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$$

$$a_i = (q_1^T a_i) q_1 + \cdots + \left(q_{i-1}^T a_i\right) q_{i-1} + \|\tilde{q}_i\| q_i$$

Review

Matrix–Matrix Multiplication

As a set of matrix–vector products.

$$C = AB = A\begin{bmatrix} | & | & & | \\ b_1 & b_2 & \cdots & b_p \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ Ab_1 & Ab_2 & \cdots & Ab_p \\ | & | & & | \end{bmatrix}$$

Here the *i*th column of C is given by the matrix–vector product with the vector on the right, $c_i = Ab_i$. These matrix–vector products can in turn be interpreted using both viewpoints given in the previous subsection.

Matrix–Vector Multiplication

If we write A by columns, then we have:

$$y = Ax = \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = [a_1]x_1 + [a_2]x_2 + \cdots + [a_n]x_n \,.$$

- y is a linear combination of the columns A.

Gram–Schmidt in matrix notation

Important

$$a_1 = \|\tilde{q}_1\| q_1$$

$$a_2 = (q_1^T a_2) q_1 + \|\tilde{q}_2\| q_2$$

$$\vdots$$

$$a_k = (q_1^T a_k) q_1 + \cdots + \left(q_{k-1}^T a_k\right) q_{k-1} + \|\tilde{q}_k\| q_k$$

$$[a_1 \quad a_2 \quad \dots \quad a_k] = [q_1 \quad q_2 \quad \dots \quad q_k] \begin{bmatrix} \|\tilde{q}_1\| & q_1^T a_2 & \cdots & q_1^T a_k \\ 0 & \|\tilde{q}_2\| & \cdots & q_2^T a_k \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & q_{k-1}^T a_k \\ 0 & 0 & \cdots & \|\tilde{q}_k\| \end{bmatrix}$$

$$A_{n\times k} = Q_{n\times k} \times R_{k\times k}$$

Gram–Schmidt in matrix notation

Important

1. Run Gram–Schmidt on columns $a_1, \ldots, a_k$ of $\mathrm{n} \times k$ matrix A
2. If columns are linearly independent, get orthonormal $q_1, \ldots, q_k$
3. Define $\mathrm{n} \times k$ matrix Q with columns $q_1, \ldots, q_k$
4. $Q^T Q = I$
5. From Gram–Schmidt algorithm

$$\begin{aligned} a_i &= (q_1^T a_i) q_1 + \cdots + \left(q_{i-1}^T a_i\right) q_{i-1} + + \|\tilde{q}_i\| q_i \\ &= R_{1i} q_1 + \cdots + R_{ii} q_i \end{aligned}$$

   With $R_{1j} = q_i^T a_j$ for $i < j$ and $R_{ii} = \|\tilde{q}_i\|$

6. Defining $R_{ij} = 0$ for $i > j$ we have $A = QR$
7. R is upper triangular, with positive diagonal entries

QR factorization

Definition

A factorization of a matrix A as $A = QR$ where Factors satisfy $Q^T Q = I$, R upper triangular with positive diagonal entries, is called a **QR factorization** of A.

Suppose A is a square matrix with linearly independent columns. Then there exist unique matrices Q and R such that Q is unitary, R is upper triangular with only positive numbers on its diagonal, and

$$A = QR.$$

$R_{jk} = < a_k, q_j >$

Note

The QR factorization of a matrix :

- ❑ Can be computed using Gram–Schmidt algorithm (or some variations)
- ❑ Has a huge number of uses, which we'll see soon

Important

To find QR decomposition:

- Q: Use Gram-Schmidt to find orthonormal basis for column space of A
- Let $R = Q^T A$
- OR: $R_{jk} =< a_k, q_j >$

- If A is a square matrix, then Q is square with orthonormal columns (orthogonal matrix)

QR Decomposition (QU) (Factorization)

Theorem

if $A \epsilon \mathbb{R}^{m \times n}$ has linearly independent columns then it can be factored as

$$A = QR$$

Q-factor

- ❑ Q is $m \times n$ with orthonormal columns $(Q^T Q = I)$
- ❑ If A is square $(m = n)$, then Q is orthogonal $(Q^T Q = QQ^T = I)$

R-factor

- ❑ R is n× n, upper triangular, with nonzero diagonal elements
- ❑ R is nonsingular (diagonal elements are nonzero)

QR Decomposition

Example

$$A = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix}$$

$$q_1 = \frac{1}{2}\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, q_2 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, q_3 = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}, \|\tilde{q}_1\| = 2, \|\tilde{q}_2\| = 2, \|\tilde{q}_3\| = 4$$

- QR :

$$\begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 8 \\ 0 & 0 & 4 \end{bmatrix}$$

Generalization of QR Decompose

$$A_{4\times 6} = [a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6]$$

Linear Independent

$$\begin{cases} a_1 = a_{11}q_1 \\ a_2 = a_{21}q_1 + a_{22}q_2 \\ a_3 = a_{31}q_1 + a_{32}q_2 \\ a_4 = a_{41}q_1 + a_{42}q_2 + a_{43}q_3 \\ a_5 = a_{51}q_1 + a_{52}q_2 + a_{53}q_3 \\ a_6 = a_{61}q_1 + a_{62}q_2 + a_{63}q_3 \end{cases}$$

Block upper triangular matrix

$$[a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6] = [q_1 \quad q_2 \quad q_3]\begin{bmatrix} a_{11} & a_{21} & a_{31} & a_{41} & a_{51} & a_{61} \\ 0 & a_{22} & a_{32} & a_{42} & a_{52} & a_{62} \\ 0 & 0 & 0 & a_{43} & a_{53} & a_{63} \end{bmatrix}$$

$$A_{4\times 6} = Q_{4\times 3} \times R_{3\times 6}$$

References

- Chapter 1: Advanced Linear and Matrix Algebra, Nathaniel Johnston
- Chapter 6: Linear Algebra David Cherney
- Linear Algebra and Optimization for Machine Learning
- Introduction to Applied Linear Algebra Vectors, Matrices, and Least Squares