Linear Algebra

Orthogonality

Department of Computer Engineering
Sharif University of Technology

Hamid R. Rabiee rabiee@sharif.edu
Maryam Ramezani maryam.ramezani@sharif.edu

Table of contents

01

Orthogonality

Orthogonal Sets

Definition

- 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).

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

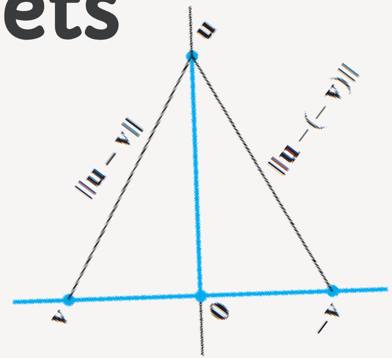
a) $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ for all $\mathbf{v} \neq \boldsymbol{w} \in B$, and (mutual orthogonality)

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

- Geometry
- Algebra



https://youtu.be/dqdSzqsm7bY

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

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 $S = \{a_1, \dots, a_k\}$ 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 $a_1, \ldots, a_k$, if and only if:

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

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

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

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

Orthonormal vectors

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

Orthonormal vectors

Theorem

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*

Example

- Standard unit n-vectors $e_1, .., e_n$
- The 3-vectors

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

- 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?

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

02

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.

03

Orthogonal Complements

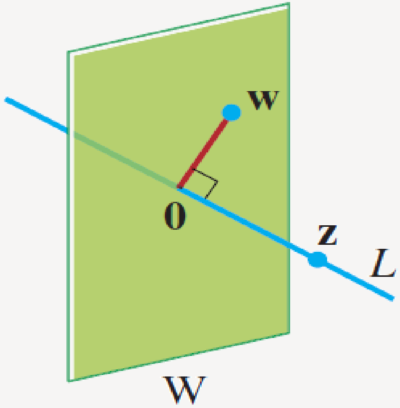
Definition

- 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.

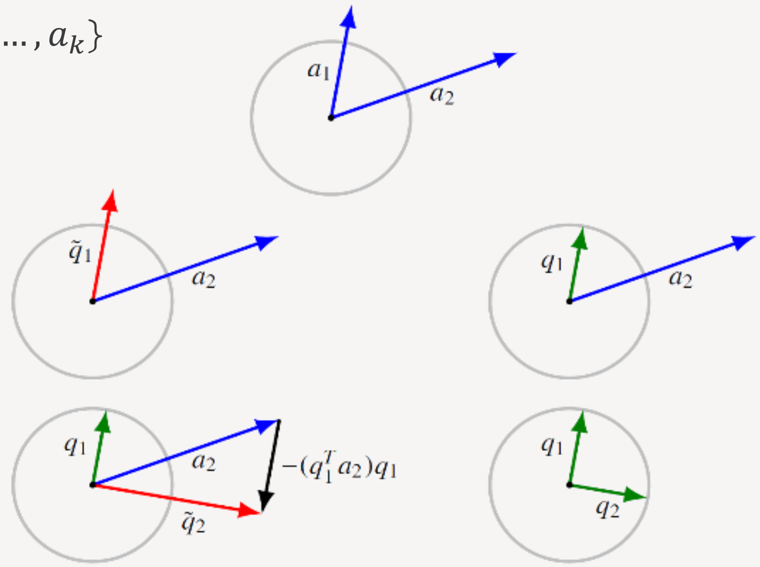
$W_1 = span((1,0,0))$ and $W_2 = span((0,1,0))$.

04

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 $a_1, \dots, a_k$ for $i = 1, \dots, k$

  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), $a_1, \dots, a_k$ are linearly independent.
- If G–S stops early in iteration $i = j$, then a_j is a linear combination of $a_1, \dots, a_{j-1}$ (so $a_1, \dots, a_k$ 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 $a_1, \dots, a_k$ for $i = 1, \dots, k$

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: $q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$

Complexity of the Gram–Schmidt algorithm. We now derive an operation count for the Gram–Schmidt algorithm. In the first step of iteration i of the algorithm, $i-1$ inner products

$$q_1^T a_i, \quad \ldots, \quad q_{i-1}^T a_i$$

between vectors of length n are computed. This takes $(i-1)(2n-1)$ flops. We then use these inner products as the coefficients in $i-1$ scalar multiplications with the vectors $q_1, \ldots, q_{i-1}$. This requires $n(i-1)$ flops. We then subtract the $i-1$ resulting vectors from a_i, which requires another $n(i-1)$ flops. The total flop count for step 1 is

$$(i-1)(2n-1) + n(i-1) + n(i-1) = (4n-1)(i-1)$$

flops. In step 3 we compute the norm of $\tilde{q}_i$, which takes approximately $2n$ flops. We then divide $\tilde{q}_i$ by its norm, which requires n scalar divisions. So the total flop count for the ith iteration is $(4n-1)(i-1) + 3n$ flops.

The total flop count for all k iterations of the algorithm is obtained by summing our counts for $i = 1, \ldots, k$:

$$\sum_{i=1}^{k}((4n-1)(i-1) + 3n) = (4n-1)\frac{k(k-1)}{2} + 3nk \approx 2nk^2,$$

where we use the fact that

$$\sum_{i=1}^{k}(i-1) = 1 + 2 + \cdots + (k-2) + (k-1) = \frac{k(k-1)}{2}, \tag{5.7}$$

which we justify below. The complexity of the Gram–Schmidt algorithm is $2nk^2$; its order is nk^2. We can guess that its running time grows linearly with the lengths

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

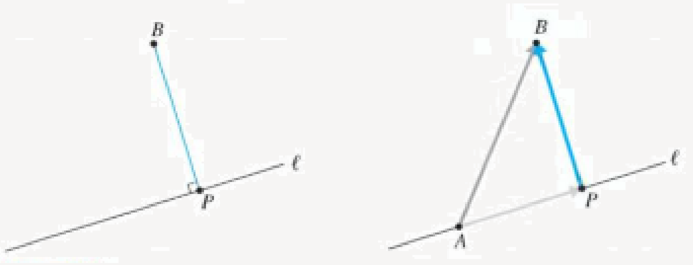
$$\langle f, g \rangle = \int_{-1}^{1} f(x)g(x)dx$$

05

Projection

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



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**

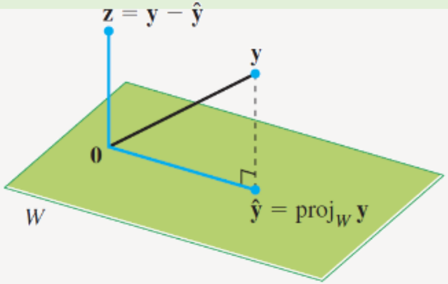
Orthogonal Decomposition Theorem

Theorem

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

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

Proof



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

06

Orthogonal Matrix

Orthogonal Matrix

Definition

- A matrix is orthogonal if its columns are:
  - orthogonal
  - has norm 1

Orthogonal Matrix

Theorem

A matrix is orthogonal if and only if it preserves length and angle.

Proof

Proof. Let us first show that an orthogonal transformation preserves length and angles. So, let us assume that $A^TA = 1$ first. Now, using the properties of the transpose as well as the definition $A^TA = 1$, we get $|Ax|^2 = Ax \cdot Ax = A^TAx \cdot x = 1x \cdot x = x \cdot x = |x|^2$ for all vectors x. Let α be the angle between x and y and let β denote the angle between Ax and Ay and α the angle between x and y. Using $Ax \cdot Ay = x \cdot y$ again, we get $|Ax||Ay|\cos(\beta) = Ax \cdot Ay = x \cdot y = |x||y|\cos(\alpha)$. Because $|Ax| = |x|, |Ay| = |y|$, this means $\cos(\alpha) = \cos(\beta)$. As we have defined the angle between two vectors to be a number in $[0, \pi]$ and cos is monotone on this interval, it follows that $\alpha = \beta$.
To the converse: if A preserves angles and length, then $v_1 = Ae_1, \cdots v_n = Ae_n$ form an orthonormal basis. By looking at $B = A^TA$ this shows off diagonal entries of B are 0 and diagonal entries of B are 1. The matrix A is orthogonal. □

Square Orthogonal Matrix

Note

- Columns of A are orthonormal $\leftrightarrow A^T A = I$
- Square matrix with orthonormal columns is an orthogonal matrix
  - Columns and rows are orthonormal vectors
  - $A^T A = AA^T = I$
  - Is necessarily invertible with inverse $A^T = A^{-1}$

Square Orthogonal Matrix

Example

- Identity matrix $I^T I = I$

- Rotation matrix

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

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

Square Orthogonal Matrix

Example

- Reflection matrix

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

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

Theorem

Every 2×2 orthogonal matrices can be expressed as Rotation or Reflection, or a composition of them.
proof?

Tall Orthogonal Matrix

- $A \in \mathbb{R}^{m\times n}, m > n$
- The inner products of the columns give the identity, so $\boldsymbol{A^T A = I_n}$

$$(A^T A)_{ij} = \langle a_i, a_j \rangle = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- AA^T is a **projection matrix** onto the column space of A. If $m>n$ then AA^T is not full rank. Therefore, $\boldsymbol{AA^T \neq I_m}$

$$\text{rank}(AA^T) \leq \min(\text{rank}(A), \text{rank}(A^T)) = \text{rank}(A) \leq n$$

But remember: $AA^T \in \mathbb{R}^{m\times m}$. So the maximum possible rank of AA^T is n, which is strictly less than m.

Wide Orthogonal Matrix

- $A \in \mathbb{R}^{m \times n}, m < n$
- The inner products of the rows give the identity, so $\boldsymbol{AA^T = I_m}$

$$(AA^T)_{ij} = \langle a_i, a_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

- $A^T A$ is a **projection matrix** onto the row space of A. If $m<n$ then $A^T A$ is not full rank. Therefore, $\boldsymbol{A^T A \neq I_n}$

$$\text{rank}(A^T A) \leq \min(\text{rank}(A), \text{rank}(A^T)) = \text{rank}(A) \leq m$$

But remember: $A^T A \in \mathbb{R}^{n \times n}$.

So the maximum possible rank of $A^T A$ **is** m**, which is strictly less than** n**.**

Properties of Orthogonal Matrix

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:

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

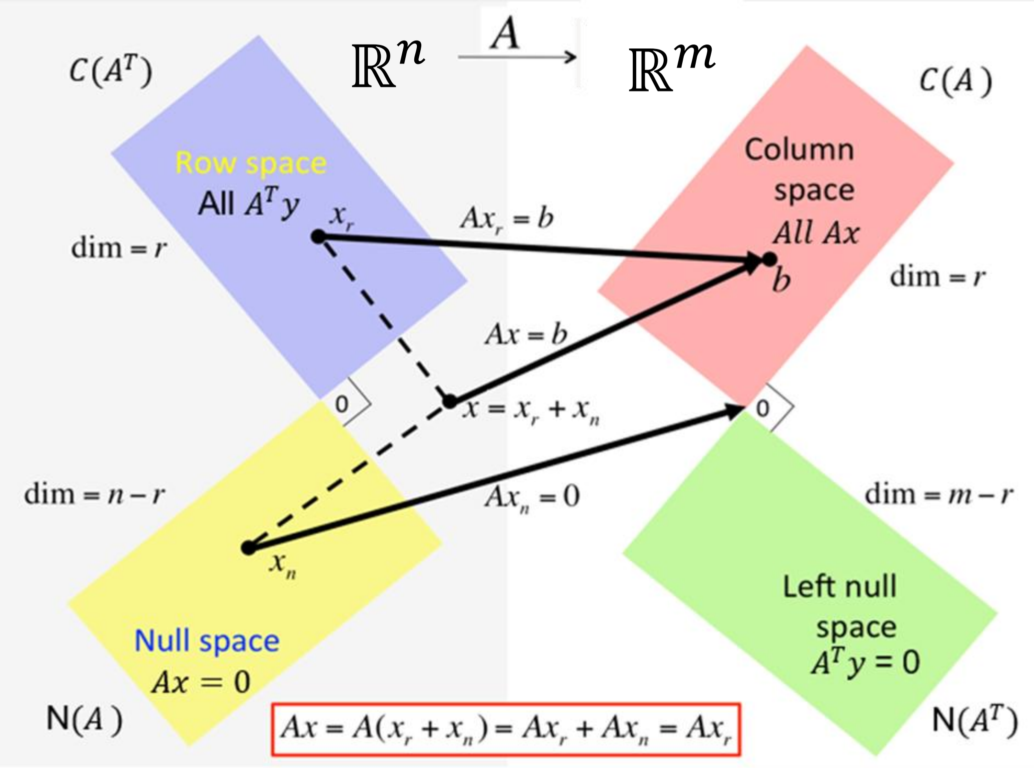
- Preserves angels:

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

This is a mapping with preserving properties of input

Four Fundamental Subspaces of Matrix Space

- Proof that:
  - $C(A^T) \perp N(A)$
  - $C(A) \perp N(A^T)$



07

QR Factorization (QR Decomposition)

Gram–Schmidt in Matrix Notation

Important

Run Gram-Schmidt on columns $a_1, \dots, a_k$ 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 ith 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

QR Factorization

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 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 Factorization

$$A_{4\times6} = [a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6]$$

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

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

$$A_{4\times6} = Q_{4\times3} \times R_{3\times6}$$

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