Lecture 7: Orthogonality, Orthonormal Bases, and Gram–Schmidt

Question:

Are some bases better than others?

Orthogonal Vectors

Recall that the inner product of

$$x, y \in \mathbb{R}^n$$

is

$$\langle x, y \rangle = x^T y.$$

We say that x and y are **orthogonal** if

$$\boxed{x^T y = 0.}$$

Geometrically, they meet at a right angle.

Example:

$$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad y = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Then

$$x^T y = 1 - 1 = 0.$$

Therefore,

$$x \perp y.$$

More generally, vectors

$$q_1, \dots, q_r$$

are called an **orthogonal set** if

$$q_i^T q_j = 0 \qquad \text{whenever } i \neq j.$$

They are called **orthonormal** if, in addition,

$$\|q_i\|_2 = 1 \qquad \text{for every } i.$$

Equivalently,

$$\boxed{q_i^T q_j = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}}$$

Why Orthonormal Bases Are Convenient

Suppose

$$q_1, \dots, q_r$$

form an orthonormal basis of a subspace L.

If

$$x \in L,$$

then

$$x = \alpha_1 q_1 + \cdots + \alpha_r q_r.$$

How do we find α_i?

Take the inner product with q_i:

$$q_i^T x = q_i^T \left(\alpha_1 q_1 + \cdots + \alpha_r q_r \right).$$

Since the basis is orthonormal,

$$q_i^T q_j = 0 \qquad (i \neq j),$$

and

$$q_i^T q_i = 1.$$

Therefore,

$$q_i^T x = \alpha_i.$$

Hence

$$\boxed{\alpha_i = q_i^T x.}$$

So

$$\boxed{x = \sum_{i=1}^{r} (q_i^T x) q_i.}$$

Compare this with a general basis:

$$x = \alpha_1 v_1 + \cdots + \alpha_r v_r.$$

For a general basis, finding the coefficients requires solving a linear system.

For an orthonormal basis, the coefficients are obtained immediately by taking inner products.

This is why orthonormal bases are special.

Pythagorean Theorem

Suppose

$$q_1, \ldots, q_r$$

are orthonormal and

$$x = \sum_{i=1}^{r} \alpha_i q_i.$$

Then

$$\|x\|_2^2 = x^T x.$$

Expanding,

$$\begin{aligned} \|x\|_2^2 &= \left(\sum_i \alpha_i q_i \right)^T \left(\sum_j \alpha_j q_j \right) \\ &= \sum_i \sum_j \alpha_i \alpha_j q_i^T q_j. \end{aligned}$$

All cross terms vanish because

$$q_i^T q_j = 0 \qquad (i \neq j).$$

Therefore,

$$\boxed{\|x\|_2^2 = \sum_{i=1}^{r} \alpha_i^2.}$$

Since

$$\alpha_i = q_i^T x,$$

we also have

$$\boxed{\|x\|_2^2 = \sum_{i=1}^{r} (q_i^T x)^2, \qquad x \in L.}$$

This is the higher-dimensional version of

$$a^2 + b^2 = c^2.$$

Projection onto One Direction

Suppose

$$q \in \mathbb{R}^n, \qquad \|q\|_2 = 1.$$

For any vector x, the component of x in the direction q is

$$\boxed{(q^T x) q.}$$

Why?

We want a multiple

$$\alpha q$$

such that the remaining part

$$x - \alpha q$$

is orthogonal to q.

Thus,

$$q^T(x - \alpha q) = 0.$$

So

$$q^T x - \alpha q^T q = 0.$$

Since

$$q^T q = 1,$$

we obtain

$$\alpha = q^T x.$$

Therefore,

$$\boxed{\text{proj}_q(x) = (q^T x)q.}$$

If v is not normalized, then

$$q = \frac{v}{\|v\|_2},$$

and therefore

$$\text{proj}_v(x) = \frac{v^T x}{v^T v} v.$$

Gram–Schmidt: The Main Idea

Suppose we start with linearly independent vectors

$$a_1, a_2, \ldots, a_r.$$

We want to construct orthonormal vectors

$$q_1, q_2, \ldots, q_r$$

with the same span.

Start with

$$q_1 = \frac{a_1}{\|a_1\|_2}.$$

For the second vector, remove from a_2 the part pointing in the q_1 direction:

$$u_2 = a_2 - (q_1^T a_2)q_1.$$

Then

$$q_1^T u_2 = q_1^T a_2 - (q_1^T a_2)q_1^T q_1 = 0.$$

Thus u_2 is orthogonal to q_1.

Normalize:

$$q_2 = \frac{u_2}{\|u_2\|_2}.$$

For the third vector, remove both previous directions:

$$u_3 = a_3 - (q_1^T a_3)q_1 - (q_2^T a_3)q_2,$$

then set

$$q_3 = \frac{u_3}{\|u_3\|_2}.$$

In general,

$$\boxed{u_k = a_k - \sum_{j=1}^{k-1}(q_j^T a_k)q_j}$$

and

$$\boxed{q_k = \frac{u_k}{\|u_k\|_2}.}$$

This is the **Gram–Schmidt algorithm**.

Geometric Interpretation

At each step,

$$\sum_{j=1}^{k-1}(q_j^T a_k)q_j$$

is the part of a_k already explained by the previous directions.

Hence

$$u_k = a_k - \text{“part already explained”}.$$

So u_k contains only the new direction contributed by a_k.

Then we normalize this new direction.

This connects directly to our discussion of rank:

$$\boxed{\text{Gram–Schmidt keeps only genuinely new directions.}}$$

If

$$u_k = 0,$$

then

$$a_k \in \operatorname{span}\{a_1, \ldots, a_{k-1}\},$$

so a_k contributes no new direction.

Example

Apply Gram–Schmidt to

$$a_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad a_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

First,

$$\|a_1\|_2 = \sqrt{2},$$

so

$$q_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$

Next,

$$q_1^T a_2 = \frac{1}{\sqrt{2}}.$$

Therefore,

$$\begin{aligned} u_2 &= a_2 - (q_1^T a_2) q_1 \\ &= \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 1/2 \\ -1/2 \\ 1 \end{pmatrix}. \end{aligned}$$

Its norm is

$$\|u_2\|_2 = \sqrt{\frac{1}{4} + \frac{1}{4} + 1} = \sqrt{\frac{3}{2}}.$$

Hence

$$q_2 = \frac{u_2}{\|u_2\|_2} = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

Check:

$$q_1^T q_2 = 0, \qquad \|q_1\|_2 = \|q_2\|_2 = 1.$$

Therefore,

$$\{q_1, q_2\}$$

is an orthonormal basis for

$$\mathrm{span}\{a_1, a_2\}.$$

Matrix View

Put the orthonormal vectors into a matrix:

$$Q = \begin{pmatrix} | & & | \\ q_1 & \cdots & q_r \\ | & & | \end{pmatrix}.$$

Because the columns are orthonormal,

$$(Q^T Q)_{ij} = q_i^T q_j.$$

Therefore,

$$\boxed{Q^T Q = I_r.}$$

This equation is extremely important.

If

$$x \in \mathrm{Col}(Q),$$

then

$$x = Q\alpha$$

for some $\alpha \in \mathbb{R}^r$.

Multiplying by Q^T,

$$Q^T x = Q^T Q\alpha = \alpha.$$

Thus

$$\boxed{\alpha = Q^T x.}$$

Again: orthonormal coordinates are easy to compute.

Where We Are Going

Suppose

$$A = \begin{pmatrix} | & & | \\ a_1 & \cdots & a_n \\ | & & | \end{pmatrix}.$$

Gram–Schmidt converts its columns into orthonormal directions

$$q_1, \dots, q_r.$$

Since every column a_j can be expressed using the q_i's, we should be able to write

$$A = QR.$$

Next lecture:

$$\boxed{\text{Gram–Schmidt} \quad \Longrightarrow \quad \text{QR decomposition.}}$$

Lecture 8: The QR Decomposition

Previously

Given linearly independent vectors

$$a_1, \dots, a_r,$$

Gram–Schmidt constructs orthonormal vectors

$$q_1, \dots, q_r$$

with

$$\operatorname{span}\{q_1, \dots, q_r\} = \operatorname{span}\{a_1, \dots, a_r\}.$$

The construction is

$$u_k = a_k - \sum_{j=1}^{k-1} (q_j^T a_k) q_j, \qquad q_k = \frac{u_k}{\|u_k\|_2}.$$

The key matrix identity is

$$\boxed{Q^T Q = I.}$$

Today we apply Gram–Schmidt to the columns of a matrix.

From Gram–Schmidt to Matrix Factorization

Let

$$A = \begin{pmatrix} | & | & & | \\ a_1 & a_2 & \cdots & a_n \\ | & | & & | \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

For now suppose the columns are linearly independent.

Thus

$$n \le m$$

and

$$\operatorname{rank}(A) = n.$$

Apply Gram–Schmidt to

$$a_1, \dots, a_n.$$

We obtain orthonormal vectors

$$q_1, \dots, q_n.$$

From the first step,

$$a_1 = \|a_1\|_2 q_1.$$

Write

$$r_{11} = \|a_1\|_2.$$

Then

$$a_1 = r_{11} q_1.$$

For the second column,

$$u_2 = a_2 - (q_1^T a_2) q_1.$$

Therefore,

$$a_2 = (q_1^T a_2) q_1 + u_2.$$

But

$$u_2 = \|u_2\|_2 q_2.$$

Hence

$$a_2 = r_{12} q_1 + r_{22} q_2,$$

where

$$r_{12} = q_1^T a_2, \qquad r_{22} = \|u_2\|_2.$$

Similarly,

$$a_3 = r_{13} q_1 + r_{23} q_2 + r_{33} q_3.$$

In general,

$$\boxed{a_j = \sum_{i=1}^{j} r_{ij} q_i.}$$

Notice something important:

$$a_j$$

only uses

$$q_1, \ldots, q_j.$$

There are no terms involving

$$q_{j+1}, q_{j+2}, \ldots.$$

The QR Decomposition

Define

$$Q = \begin{pmatrix} | & & | \\ q_1 & \cdots & q_n \\ | & & | \end{pmatrix} \in \mathbb{R}^{m \times n}.$$

Define

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & \cdots & r_{1n} \\ 0 & r_{22} & r_{23} & \cdots & r_{2n} \\ 0 & 0 & r_{33} & \cdots & r_{3n} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & r_{nn} \end{pmatrix}.$$

Then column by column,

$$A = QR.$$

Thus

$$\boxed{A = QR,}$$

where

$$\boxed{Q^T Q = I}$$

and

$$\boxed{R \text{ is upper triangular.}}$$

This is called the **QR decomposition**.

Why Is R Upper Triangular?

Look at the jth column.

At step j, Gram–Schmidt removes the components of a_j along

$$q_1, \ldots, q_{j-1}$$

and creates the new direction q_j.

Therefore,

$$a_j \in \operatorname{span}\{q_1, \ldots, q_j\}.$$

So

$$a_j = r_{1j} q_1 + \cdots + r_{jj} q_j.$$

There are no coefficients corresponding to

$$q_{j+1}, \ldots, q_n.$$

Thus

$$r_{ij} = 0 \qquad \text{for } i > j,$$

which is exactly the definition of an upper triangular matrix.

How Do We Find the Entries of R?

Since

$$A = QR,$$

multiply on the left by Q^T:

$$Q^T A = Q^T Q R.$$

Because

$$Q^T Q = I,$$

we obtain

$$\boxed{R = Q^T A.}$$

Thus

$$\boxed{r_{ij} = q_i^T a_j.}$$

For $i > j$,

$$q_i^T a_j = 0,$$

which again explains why R is upper triangular.

Example

Consider

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Its columns are

$$a_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \qquad a_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}.$$

From last lecture,

$$q_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix},$$

and

$$q_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}.$$

Therefore,

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

Now calculate

$$R = Q^T A.$$

First,

$$r_{11} = q_1^T a_1 = \sqrt{2}.$$

Also,

$$r_{12} = q_1^T a_2 = \frac{1}{\sqrt{2}}.$$

Next,

$$r_{21} = q_2^T a_1 = 0,$$

and

$$r_{22} = q_2^T a_2 = \frac{3}{\sqrt{6}} = \sqrt{\frac{3}{2}}.$$

Hence

$$R = \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} \\ 0 & \sqrt{\frac{3}{2}} \end{pmatrix}.$$

Thus

$$\boxed{A = QR.}$$

What if A Is Rank Deficient?

Suppose

$$A \in \mathbb{R}^{m \times n}$$

has

$$\operatorname{rank}(A) = r < n.$$

Then some columns do not provide new directions.

During Gram–Schmidt, for such a column,

$$u_j = 0.$$

We simply do not create a new q_j from that column.

Eventually we obtain

$$r$$

orthonormal basis vectors for the column space:

$$q_1, \ldots, q_r.$$

Put them into

$$Q \in \mathbb{R}^{m \times r}.$$

Every column of A can be written as a combination of these r vectors, so

$$A = QR$$

for some

$$R \in \mathbb{R}^{r \times n}.$$

Therefore:

Theorem. If

$$A \in \mathbb{R}^{m \times n}$$

has rank r, then there exist

$$Q \in \mathbb{R}^{m \times r}, \qquad R \in \mathbb{R}^{r \times n},$$

such that

$$\boxed{A = QR,}$$

where the columns of Q are orthonormal and R is upper triangular in the appropriate rectangular sense.

QR and Rank

Compare this with our previous low-rank factorization

$$A = UV.$$

There,

$$U \in \mathbb{R}^{m \times r}$$

could be any basis of the column space.

QR gives us a much better structured version:

$$\boxed{A = QR,}$$

where

$$Q^T Q = I.$$

Thus QR is a low-rank factorization in which the basis of the column space is orthonormal.

This gives us two benefits:

1. the coordinates are easy to compute;
2. lengths and angles are much easier to analyze.

Orthonormal Matrices Preserve Length

Suppose

$$Q \in \mathbb{R}^{m \times r}$$

has orthonormal columns, so

$$Q^T Q = I_r.$$

For any

$$x \in \mathbb{R}^r,$$

we have

$$\begin{aligned} \|Qx\|_2^2 &= (Qx)^T (Qx) \\ &= x^T Q^T Qx \\ &= x^T x \\ &= \|x\|_2^2. \end{aligned}$$

Therefore,

$$\boxed{\|Qx\|_2 = \|x\|_2.}$$

So multiplying by Q changes the coordinate system but does not stretch vectors.

This is another major reason orthonormal bases are useful.

Back to Compression

Recall our original problem:

$$\min_{\operatorname{rank}(B)\le r} \|W - B\|_F.$$

Suppose we somehow find a good rank-r approximation

$$W_r.$$

Its QR decomposition gives

$$W_r = QR,$$

where

$$Q \in \mathbb{R}^{m\times r}, \qquad R \in \mathbb{R}^{r\times n}.$$

Instead of storing an

$$m \times n$$

matrix, we store an

$$m \times r$$

matrix and an

$$r \times n$$

matrix.

If

$$r \ll \min(m, n),$$

this can be much smaller.

But QR does *not* yet solve our main problem.

Given a full-rank matrix W, QR does not tell us which rank-r matrix is the best approximation.

We still need to solve

$$\boxed{\min_{\operatorname{rank}(B)\le r} \|W - B\|_F.}$$

That question will eventually lead us to the singular value decomposition.

One Numerical Warning

Classical Gram–Schmidt is conceptually simple and very useful for understanding QR.

However, straightforward classical Gram–Schmidt can be numerically unstable on a computer.

In practice, QR decompositions are usually computed using more stable methods such as Householder reflections.

For this course, however, Gram–Schmidt is the right starting point because it makes the geometry completely transparent.

Summary

Today:

$$\boxed{\text{Gram–Schmidt} \Longrightarrow \text{orthonormal basis} \Longrightarrow A = QR.}$$

For QR,

$$Q^T Q = I,$$

and

$$R$$

is upper triangular.

Also,

$$R = Q^T A.$$

Most importantly,

$$\boxed{\text{QR represents the column space of a matrix using orthonormal coordinates.}}$$

Next: orthogonal projection and least squares.