Linear Algebra
Week 9

G-07

21 XI 2024

1 Orthonormality

We talked about orthogonal vectors, now we additionally want them to have norm 1, unit norm, to be orthonormal. We will see the properties of such vectors and matrices that have orthonormal columns. Having an orthonormal basis to a subspace comes with a lot of computational benefits and makes your life easier. We are additionally going to look at these fantastic bases and how to find them.

1.1 Definitions and Properties

Definition 5.4.1. *(Orthonormal Vectors)* Vectors $q_1, \ldots, q_n \in \mathbb{R}^m$ are orthonormal if they are orthogonal and have norm 1. In other words, for all $i, j \in \{1, \ldots, n\}$

$$q_i^\top q_j = \delta_{ij}$$

where δ_{ij} is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

You should get familiar with the Kronecker delta. This definition applied on two vectors q_1, q_2 gives you $q_1^\top q_2 = q_2^\top q_1 = 0$ *i.e.* they are orthogonal, and $q_1^\top q_1 = \|q_1\|^2 = 1$ and $q_2^\top q_2 = \|q_2\|^2 = 1$ meaning they both have norm 1.

You can pack this set of vectors $q_1, \dots, q_n \in \mathbb{R}^m$ in a matrix Q where the q_i's are the columns of Q. Then the condition in the definition is satisfied when $Q^\top Q = I$.

A quick example to see $QQ^\top$ is not necessarily the identity matrix when Q is not square:

$$QQ^\top = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \neq I$$

The columns are orthogonal to each other and they all have norm 1. So the columns are orthonormal. But the result is not the identity matrix I. However this is the case if the matrix with orthonormal columns is additionally a square matrix. We then call it an orthogonal matrix:

Definition 5.4.3. *(Orthogonal Matrix)* A **square** matrix $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix when $Q^\top Q = I$. In this case, $QQ^\top = I$, and $Q^{-1} = Q^\top$, and the columns of Q **form an orthonormal basis** for $\mathbb{R}^n$.

The inverse is the transpose! Be careful, even though the name says orthogonal, orthogonal matrices have orthonormal columns. Otherwise $Q^\top Q$ and $QQ^\top$ would not be equal to I.

1.2 Examples of Orthonormal Vectors and Orthogonal Matrices

- **(Ex. 5.4.2.)** A classic example of an orthonormal set of vectors is the *canonical basis* $e^1, \dots, e^n \in \mathbb{R}^n$ where e^i is the i^{th} unit vector. (Has 1 in the i^{th} element otherwise 0's).

If you have to show that a matrix A is orthogonal, then all you should verify is that A is square and $A^\top A = I$. The equality $AA^\top = I$ is then implied.

- **(Ex. 5.4.4.)** The rotation matrix R_θ which correspond to rotating the plane counterclockwise by θ is obviously square:

$$R_\theta = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

  is an orthogonal matrix. If we rotate everything counterclockwise by θ and then multiply with $R_\theta^\top$ to rotate everything counterclockwise by $-\theta$ then we end up where we start. $R_\theta^\top R_\theta = I$.

- **(Ex. 5.4.5.)** Permutation matrices are orthogonal matrices. Permutations are defined as maps and the permutation matrix P associated with a permutation π has $A_{ij} = 1$ if $\pi(i) = j$. The transpose has however $A_{ij} = 1$ if $\pi(j) = i$. Hence $P^\top P = I$. If you permutate once and then permutate things in reverse then you get back to the order you started.

- **Reflection Matrices** Take $A = I - vv^\top$. Then Ax corresponds to reflecting along the hyperplane defined by $\mathbf{v}$. See solution of Assignment 8 Exercise 3 for further explanation. If the vector v has norm $\sqrt{2}$, then the reflection matrix A is orthogonal. See by direct computation $A^\top A = (I - vv^\top)^\top(I - vv^\top) = I - 2vv^\top + (vv^\top)(vv^\top) = I - 2vv^\top + v(v^\top v)v^\top = I - 2vv^\top + v(\|v\|^2)v^\top = I - 2vv^\top + (\|v\|^2)vv^\top$. We want this to be equal to I for all v. This is the case if $\|v\|^2 = 2$.

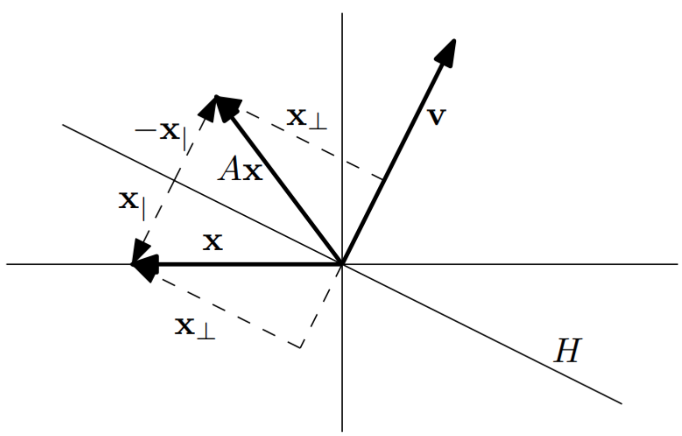


Figure 1: A sketch of the transformation.

Reflection, visualized, taken from solution of assignment 8

1.3 Norm Preserving

A very important property of orthogonal matrices is that they preserve the norm. Think of an orthogonal matrix Q as a linear transformation. Applying Q to your space leaves the lengths and angles as they are. The proof is by direct computation.

Proposition 5.4.6. Orthogonal matrices preserve norm and inner product of vectors. In other words, if $Q \in \mathbb{R}^{n\times n}$ is orthogonal, then for all $x, y \in \mathbb{R}^n$

$$\|Qx\| = \|x\| \text{ and } (Qx)^\top(Qy) = x^\top y$$

Let's prove this: For $x, y \in \mathbb{R}^n$ we have $(Qx)^\top(Qy) = x^\top Q^\top Qy \overset{Q^\top Q=I}{=} x^\top y$. This proves the second equality. The first one actually follows from this. Since a norm as is never negative as it represents the length, we can prove that Q preserves the norm-squared which would then imply the first equation. $\|Qx\|^2 = (Qx)^\top(Qx) \overset{\text{2. equation}}{=} x^\top x = \|x\|^2$.

□

This is actually an if and only if situation. We know a norm preserving matrix must be orthogonal as well. We prove one direction here. The other one is left for you to prove as the bonus of this week. No spoilers! (If you are reading this in January then you can check the solution.)

Preserving Angles: Thinking of how we calculate the angles -remember the formula for $\cos(\alpha_{ab})$ for the angle α between two vectors a and b - it should not be surprising that orthogonal matrices also preserve the angles knowing they preserve inner product and norm.

2 Projections with Orthonormal Basis

Remember the equations from the previous chapter, for projections and linear squares? They had a lot of $A^\top A$'s in them. Now if A is orthogonal all $A^\top A$'s disappear since $A^\top A = I$. Let's see this as a proposition:

Proposition 5.4.7. Let S be a subspace of $\mathbb{R}^m$ and $q_1, \ldots, q_n$ be an orthonormal basis for S. Let Q be the $m \times n$ matrix whose columns are the q_i's: $Q = \begin{bmatrix} q_1 & , \cdots , & q_n \end{bmatrix}$. Then the **Projection Matrix** that projects to S is given by $QQ^\top$, and the **Least Squares solution** to $Qx = b$ is given by $\hat{x} = Q^\top b$.

Remark 5.4.8. (Informal) Let Q be square. Then the projection matrix $QQ^\top$ is the identity matrix since the columns of Q is already a basis for the whole vector space and if you project some vector from within the vector space onto the vector space itself, the vector stays the same. If you unroll your matrix vector multiplication for the projection $Q(Q^\top x)$ it will look like this

$$x = q_1 \left(q_1^\top x\right) + q_2 \left(q_2^\top x\right) + \cdots + q_n \left(q_n^\top x\right).$$

See that $q_i^\top x$'s are scalars. So this is the linear combination of q_i's that get you the vector x. This is called a *change of basis*, as we are going to see.

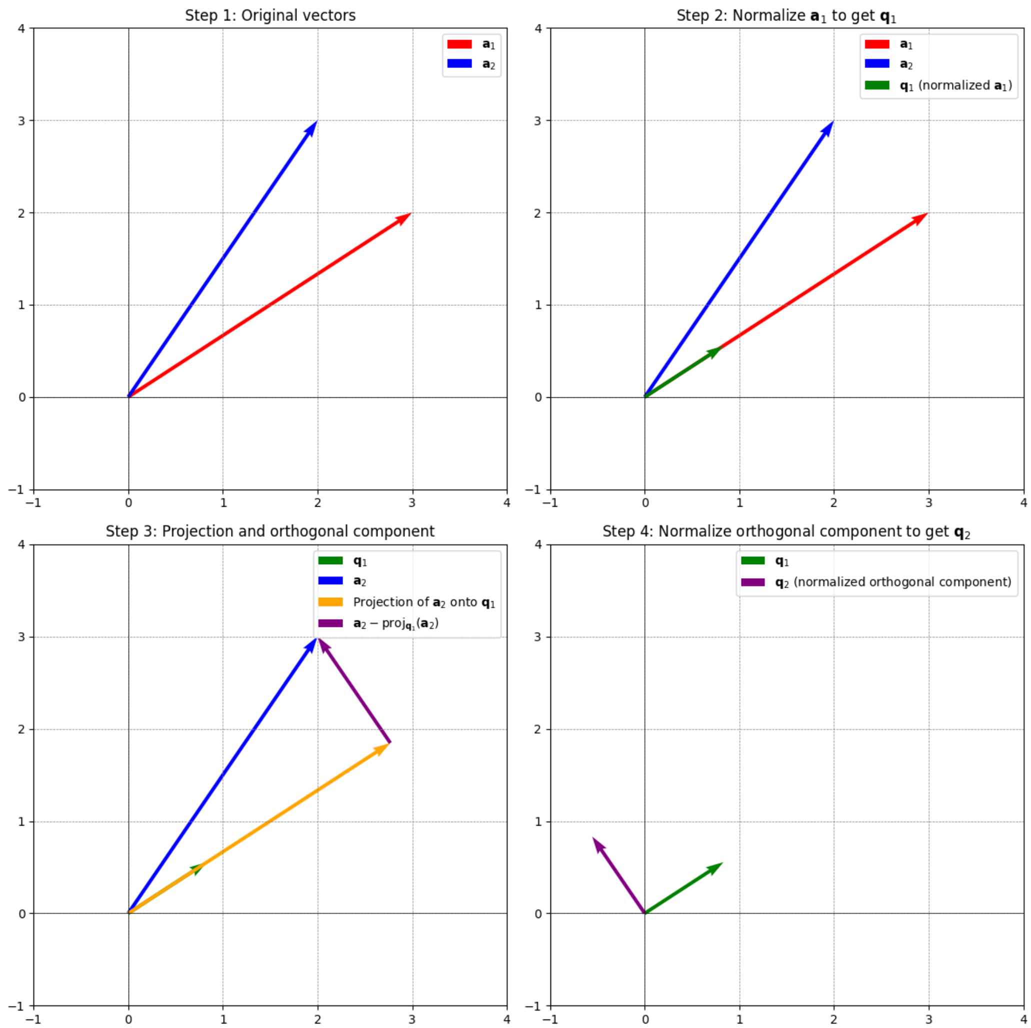
3 Gram-Schmidt Process

Orthonormal bases and orthogonal matrices are great! But if all we have is just a lame non-orthonormal basis for our subspace S, we can easily calculate an orthonormal basis out of the provided basis vectors.

Let's see this in 2 vectors a_1 and a_2. The algorithm is as follows:

1. Normalize a_1 to get $q_1 = \frac{a_1}{\|a_1\|}$
2. Project a_2 onto q_1. Remember the formula $proj_{q_1}(a_2) = \frac{q_1 q_1^\top}{q_1^\top q_1} a_2$. Since q_1 is normalized and has norm 1 we have $q_1^\top q_1 = \|q_1\|^2 = 1^2 = 1$. This is why $proj_{q_1}(a_2) = q_1 q_1^\top a_2 = (a_2^\top q_1) q_1$.[1]
3. Subtract the projection of a_2 on q_1 you calculated in the previous step from a_2 to get $q_2' = a_2 - (a_2^\top q_1) q_1$.
4. Last but not least, normalize q_2' to get $q_2 = \frac{q_2'}{\|q_2\|}$

[1] See the notes from week 8 page 3 on my website if you didn't understand why this last equation holds. It is not exactly the same thing but it helps. Also remember $a^\top b = b^\top a$.



Visualization of the Gram-Schmidt Process on two vectors. You can find the source code under additional material for week 9 on my website.

What happens if you have more than two vectors? The principle is exactly the same but now in step 2 instead of projecting your next a_k on q_1 you project the next vector a_k on the subspace spanned by all the q_i's with you have constructed so far. Then you subtract the projection from a_k to get your q'_k and lastly you normalize it. Here is the algorithm, taken from the lecture notes:

Algorithm 5.4.9. *[Gram-Schmidt Process] Given n linearly independent vectors* $a_1, \ldots, a_n$ *that span a subspace S, the Gram-Schmidt process constructs* $q_1, \ldots q_n$ *in the following way:*

- $q_1 = \frac{a_1}{\|a_1\|}$.
- *For* $k = 2, \ldots, n$ *set*

  $q'_k = a_k - \sum_{i=1}^{k-1} (a_k^\top q_i) q_i$

  $q_k = \frac{q'_k}{\|q'_k\|}$.

Theorem 5.4.10 (Correctness of Gram-Schmidt) Given n linearly independent vectors $a_1, \ldots, a_n$ the Gram-Schmidt process returns an orthonormal basis for the span of $a_1, \ldots, a_n$.

The proof is by induction. I would recommend proving this by yourself first before you read the proof in the lecture notes because this case is a beautiful example of a proof by induction. If you are reading the proof, the calculation part is just some vector arithmetic. Get a pen and try to understand it.

Linearly Dependent Case: The definitions in the lecture notes are done with sets of *linearly independent* vectors. If you have a *linearly dependent* set of vectors $a_1, \ldots, a_n$, in the step where you subtract the projection from the vector, you would get $\mathbf{0}$. Since the vector is linearly dependent to previous vectors its projection onto the span of the orthogonalized previous vectors is equal to itself. This means the orthogonalized q_k would be the vector $\mathbf{0}$ for the linearly dependent vector. In the end, if you exclude the potentially multiple $\mathbf{0}$'s among your q_i, then you still get an orthonormal basis for the span of $a_1, \ldots, a_n$

4 QR-Decomposition

We unlock a new matrix factorization, namely $A = QR$ decomposition.

Definition 5.4.11 *(QR decomposition)*. Let A be an $m \times n$ matrix with linearly independent columns. The QR decomposition is given by

$$A = QR$$

where Q is an $m \times n$ matrix with orthonormal columns (they are the output of Gram-Schmidt: *Algorithm 5.4.9*, on the columns of A) and R is an upper triangular matrix given by $R = Q^\top A$.

Let $a_1, \ldots, a_n \in \mathbb{R}^m$ be the linearly independent columns of $A \in \mathbb{R}^{m\times n}$ and $q_1, \ldots, q_n \in \mathbb{R}^m$ be the orthogonalized columns of A, using the Gram Schmidt Algorithm. Notice that $\mathbf{Span}(a_1, \ldots, a_n) = \mathbf{Span}(q_1, \ldots, q_n)$ because Gram-Schmidt finds a basis for the span of the vectors a_i. Having said that, consider the projection matrix $QQ^\top$ on the span of the q_i's. Since A is already in the span of the q_i's, applying the projection to A doesn't change the matrix. So we have $QQ^\top A = A$. Let $R = Q^\top A$. If you get these two equations together you have $QQ^\top A = QR = A$.

Lemma 5.4.12. The matrix R defined in *Definition 5.4.11* is upper triangular and invertible. Moreover, $QQ^\top A = A$ and hence, $A = QR$ is well defined.

You should read the proof of this lemma to understand the matrix R. It uses the fact that the spans of a's and q's are the same and the projection $QQ^\top$ doesn't change the vectors a. You can write the matrices in $Q^\top A$ with column and row vectors and multiply them to see your R is actually upper triangular. Let $Q = [\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_m]$ and $A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$ where $\mathbf{q}_i$'s and $\mathbf{a}_i$'s are vectors. The matrix multiplication $Q^\top A$ is given by:

$$R = Q^\top A = \begin{bmatrix} \mathbf{q}_1^\top \mathbf{a}_1 & \mathbf{q}_1^\top \mathbf{a}_2 & \ldots & \mathbf{q}_1^\top \mathbf{a}_n \\ \mathbf{q}_2^\top \mathbf{a}_1 & \mathbf{q}_2^\top \mathbf{a}_2 & \ldots & \mathbf{q}_2^\top \mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_m^\top \mathbf{a}_1 & \mathbf{q}_m^\top \mathbf{a}_2 & \ldots & \mathbf{q}_m^\top \mathbf{a}_n \end{bmatrix}$$

a_j is orthogonal to q_i for $j < i$ by construction of the orthonormal vectors by Gram-Schmidt. We have $A_{ij} = q_i^\top a_j$. Therefore for $j < i$ we have $A_{ij} = 0$. So R is indeed upper triangular. You should read the proof in the lecture notes to see why R is invertible and well defined.

4.1 Projections and Least Squares with $A = QR$

From the lecture notes:

Fact 5.4.13. *The QR decomposition greatly simplifies calculations involving Projections and Least Squares.*

- *Since* $\mathrm{C}(A) = \mathrm{C}(Q)$ *then projections on* $\mathrm{C}(A)$ *can be done with* Q *which means they are given by* $\mathrm{proj}_{\mathrm{C}(A)}(b) = QQ^\top b$.
- *The least squares solution to* $Ax = b$ *denoted by* $\hat{x}$ *is defined as a solution of the normal equations (recall* (3)*)*

  $$A^\top A\hat{x} = A^\top b.$$

  Furthermore, $A^\top A = (QR)^\top (QR) = R^\top Q^\top QR = R^\top R$, *and so we can write*

  $$R^\top R\hat{x} = R^\top Q^\top b. \tag{8}$$

  Since R *has independent columns (is full column rank) then* $\mathrm{N}(R) = \{0\}$ *and so we can simplify* (8) *to*

  $$R\hat{x} = Q^\top b, \tag{9}$$

  which can be efficiently solved by back-substitution since R *is a triangular matrix.*

We can do the simplification from (8) to (9) because $\mathbf{N}(R) = \{\mathbf{0}\}$ implies that R and in particular $R^\top$ is invertible as shown in *Lemma 5.4.12*.

5 Hints

1. Solved in class.

2. Bonus. No hints!

3. Remember the formula for 2×2 inverse: $A^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

4. Yes you should write a lot of $\sum$'s. There is no easy way. But this is actually direct calculation done with variables.

5. Think of a *radius* function. No matter which point $\mathbf{p}$ is provided as input we want the radius r to be the same.

6. You first prove the multiplication of permutation matrices is again a permutation matrix -not an easy task-. Then comes the argument of existence of a k. There are finitely many permutation matrices of size $n \times n$. So after some exponents you should go into a cycle. You can use the logic of groups and subgroups from discrete maths as well.

mkilic

♠