**Orthogonality**

Table of Contents

[Projections 4](#_Toc66216422)

[Least Squares Approximation 8](#_Toc66216423)

[Orthonormal Vectors 14](#_Toc66216424)

Vectors are said to be orthogonal if they are perpendicular to each other. For the vectors and , . Thus, the two vectors are orthogonal. We will also see that, in such cases, . and .

There will be a few rules here regarding transposes which may seem confusing since they were not encountered anywhere before.

We can prove that for vectors that are orthogonal using the Pythagoras theorem.

(There’s a brand-new transpose rule hidden there.)

(Another brand-new rule here.)

Now consider the orthogonality of subspaces. Take the floor and wall of a room. The line along which they meet contains vectors that are not perpendicular to each other. Thus, they are not orthogonal. However, the line along which two walls meet and the floor are always perpendicular to each other. Thus, these are orthogonal.

Formally, orthogonality cannot be achieved if . The dimension of a wall and a floor are 2 each, and the whole space is a 3-dimensional room, so these are not orthogonal. However, the line along which two walls meet is 1 dimensional, so the floor and that line can be orthogonal.

Now consider the null space of some matrix . For we know,

Thus, , . Similarly, , . Thus, every row is perpendicular to every vector in the null space. A row space and null space are orthogonal.

Also notice that has a rank of , while has a rank of , which means is completely covered. Similarly, has a rank of and has a rank of , which means these cover completely as well. Each of these pairs are thus complements of each other. They are called orthogonal complements.

## Projections

For the vector , the projection on the -axis is . To achieve this, we multiply it by a 3x3 matrix.

The projection of on the plane is . To achieve this,

Here, notice that and .

error

In the above diagram, the projection of on is . Since lies on the same line as , we can say . The difference between and is , the error. Why this is called the error will be explained shortly. The shortest possible distance between and for the projection is when is perpendicular to .

is essentially the resulting projection, given by multiplying the projection matrix with . Thus, . Note that the numerator and denominator do not cancel out.

The column space of is a line through . The rank of is 1.

Notice that (meaning it is symmetric) and (since applying the projection on top of itself will not move it). These two conditions are only true for projection matrices.

We need to perform projections in situations when the original equation has no valid solution. If has no solution, we know that is not in the same column space as . As a compromise, we solve the closest thing that is solvable, the projection. Thus, instead of solving for , we solve for instead, where is the best possible solution and is the projection of onto the column space. This is why the difference between and the projection , given by , is called the error.

Now let us consider a 3-dimensional projection. In this case, we will have a plane that is defined by 2 vectors, and . These two vectors create a column space, say , and a third vector is not in this column space. This means that , since if it were , then would be lying on the plane.

In the best case, is perpendicular to the column space of , . This also means that is in the left null space of , .

We have and our objective is to find . The key to the solution is that is perpendicular to the plane, and thus perpendicular to both and .

Thus,

Re-writing these equations as a matrix,

This is a general formula for -dimensional matrices.

From this, we get

Notice that we could not simply divide anymore, but had to take the inverse. This is because we do not know how many dimensions there are.

There is a good chance that a bad mistake will be made here. If we try to open up the brackets and do the inversions, we will get

This is obviously wrong since it would result in no projection at all. The reason this is wrong is because we can only perform inversions like this when is a square, invertible matrix. If were a square invertible matrix, then its column space would be , which would mean had to be in the column space of regardless of its value. In that scenario, would be correct. However, since is not a square invertible matrix, this is incorrect.

In real life, projections are useful in situations where we essentially have to find the line of best fit. Say we have 3 points, , and . The line of best fit is given by the equation . Thus, the three equations we get are:

From these, we can form this matrix:

This is the equation that cannot be solved.

## Least Squares Approximation

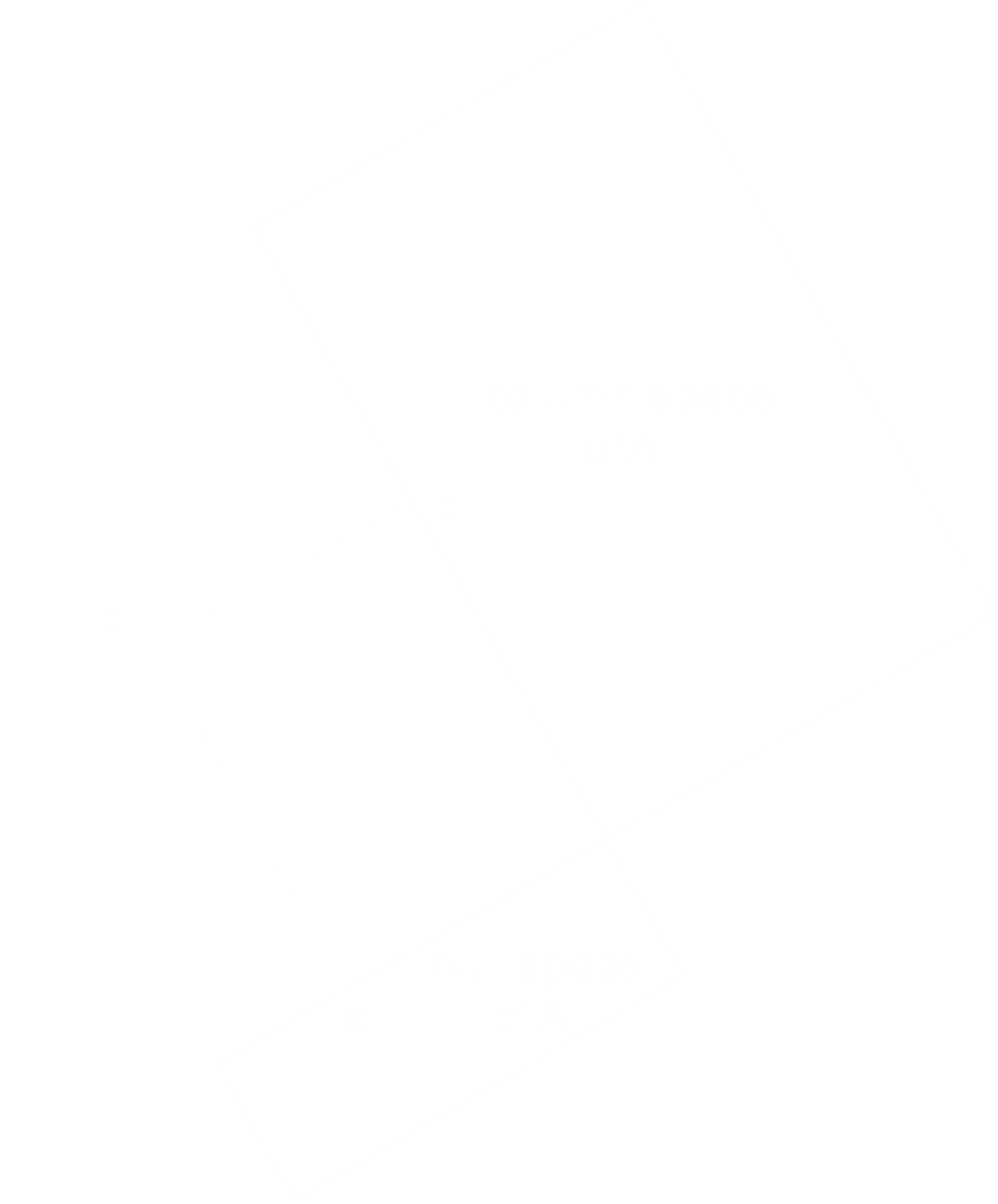
For the projection matrix ,

* If is already in the column space,
* if is perpendicular to the column space,

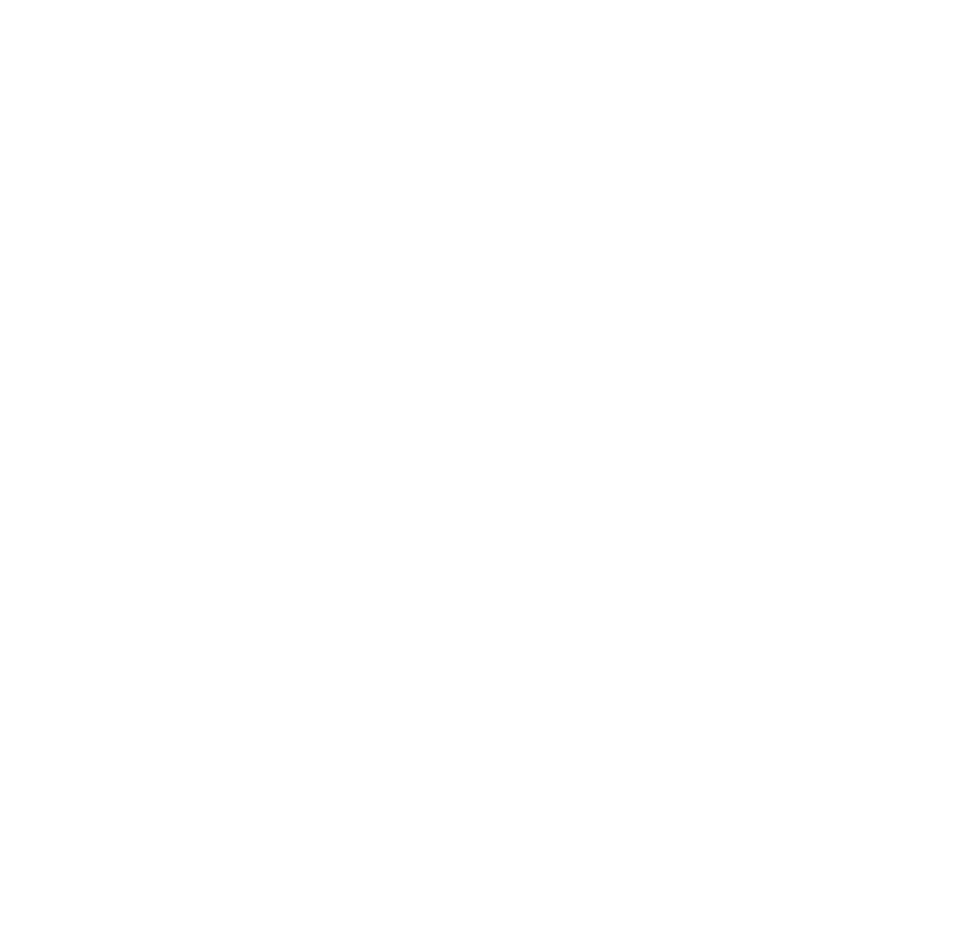
We have seen how the projection matrix would be an identity matrix if were in the column space of in the last lecture.

If is perpendicular to the column space of , it is in the null space of . This makes it obvious that since the last part of the formula for has .

The approach we will be covering here is called the least squares approximation. It helps us minimize the error between the actual vector , and the projected vector . There are two ways of looking at this problem, using vectors and using a plotted graph.



From this image, one could argue that is the projection of onto the left null space of . This would be the view from a vector’s perspective.



Now for the view from the graph’s perspective. Say we have three points here, , and , and the best-case solution would be .

For the first point the best solution would be .

For the second point the best solution would be .

For the third point the best solution would be .

If we put these three equations into the overall equation for the system ,

The best possible solution to this unsolvable system would be the one that gives the least total error. There will be a small error between each point and the line of best fit (regression line).

What we will be doing is squaring each of the errors and adding them and we want to find the minimum possible value. Thus, this is called the least squares approximation.

The least squares method however, overcompensates for outliers. If there are any anomalous points in our data, the least squares method will not give us the best fit line since we take the square of the error with that point which throws the entire line off. For now, let us assume we are not working with any anomalous points.

The error vector is . Since we are concerned about minimizing things, we need to be considering the length. At the same time, it is convenient to square this. Thus, simply put, we are trying to minimize the length of the distance between and .

In the diagram for the points, we can see three points, , and . These points are the ones that, if used instead of , and , the actual data points, would make the system solvable since they all lie on the line i.e. they are in the column space of .

We know that . In the example we are using, and . Thus, . This matrix is symmetric and invertible. We will prove that it is invertible later on. and .

Thus, the two equations we get are:

We can easily solve these for and .

We could also have solved these using calculus.

We can take the partial derivatives with respect to and to get the same two equations.

Thus, the line of best fit is .

From this, we get

Similarly

Note that the signs are not important. Doing the subtraction the other way around is acceptable.

Now go back to the original equation .

We know that and are perpendicular to each other. lies on the column space of , and is perpendicular to it. This means is also perpendicular to any vector in the column space of .

Proving is Invertible

If has independent columns, is invertible.

Suppose

We know that a matrix is invertible when its null space is only the vector. This means in order to prove that is invertible, we need to prove that has to be here.

An easy way to do this is to just multiply both sides by .

Since is the same thing as ,

We know has independent columns. must be in its null space, and the only thing in the null space of a matrix with independent columns is the vector. Thus, must be .

There is one case in which the columns are guaranteed to be independent, and that is if they are perpendicular unit vectors (orthonormal vectors). We will be discussing orthonormal vectors next.

## Orthonormal Vectors

We will be denoting vectors we work with here with to indicate that these are orthonormal vectors and not normal ones. This also means that all vectors denoted this way are orthogonal to each other and have a length of 1 unit.

For orthogonal vectors, if and if .

Take a matrix of orthonormal vectors . Thus,

This is a matrix with orthonormal columns. We could even venture to call it an orthonormal matrix. Sadly, we cannot call it an orthogonal matrix, since that term is only used for orthonormal matrices that are square.

If is square, also tells us that . Consider a permutation matrix . If we do the multiplication, we will find that . Thus, is an orthonormal matrix. The examples do not necessarily need to be this simple, since is also orthonormal. The two columns are perpendicular to each other, and since the length of each column is , the matrix is multiplied by to make the columns unit vectors. Of course, the matrix can be rectangular as well, as .

Now to look at why orthonormal matrices are useful. Consider a projection matrix . Since the column space has orthonormal basis, . If is square, . For the equation , orthonomal matrices cause the equation to become , which is just . This basically means .

If we are not given orthonormal vectors to work with, we can turn normal vectors into orthonormal ones and then work on them. This method is called the Gram-Schmidt process. Say we have two vectors and that are independent. Lets say the orthogonal vectors we want are and respectively. We can then make them orthonormal by dividing by length.

We essentially have to turn the vector into something that is orthogonal to the vector . We do not need to touch to do this, so . On the other hand, we have already seen how to project a vector onto a vector , giving us a projected vector . Here, we just need to work with the error , since was orthogonal to . Thus, is just what we previously called . Thus, .

We can even check that we got the correct formula since, if and are orthogonal, .

Things get a little more complicated if we say we have a third independent vector that we also need to make orthonormal. To get , we need to subtract from its components in the and directions.

From all of these, we can find , and .

Consider a numerical example now.

,

The formula for this entire topic is just . This is the expression for Gram-Schmidt.

is always upper triangular, i.e. .

Thus, the connection between a matrix with independent columns and the orthonormal version of that matrix, , is a triangular matrix .