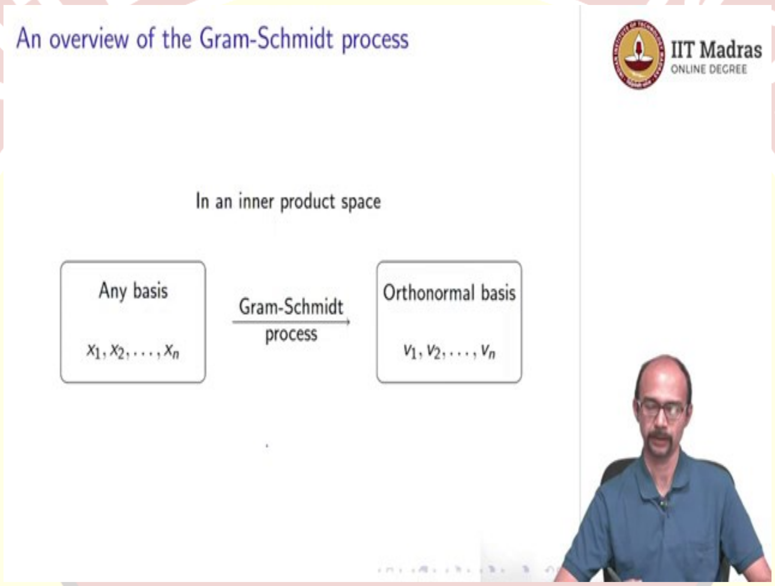
IIT Madras
ONLINE DEGREE

Mathematics for Data Science 2
Professor. Sarang Sane
Department of Mathematics
Indian Institute of Management, Madras
The Gram-Schmidt Process

In this video we are going to talk about the Gram-Schmidt process. So, what is the Gram-Schmidt process? Let us look at this process in a nutshell.

(Refer Slide Time: 0:27)



So, what the Gram-Schmidt process does is that in an inner product space it takes any basis $x_1, x_2, \ldots x_n$ and it produces, we use the Gram-Schmidt process and it produces an orthonormal basis $v_1, v_2, \ldots v_n$. So, let us recall before we start going ahead what is a basis. So, a basis is a linearly independent set which is also spanning.

So, in other words every element of your vector space or in this case for inner product space can be written as a unique linear combination of these vectors $x_1, x_2, \ldots x_n$. So, this is true for any vector space that we can produce a basis and we have seen how to do that, but suppose now that you have an inner product space. So, you have not only a vector space but on that vector space you also have an inner product.

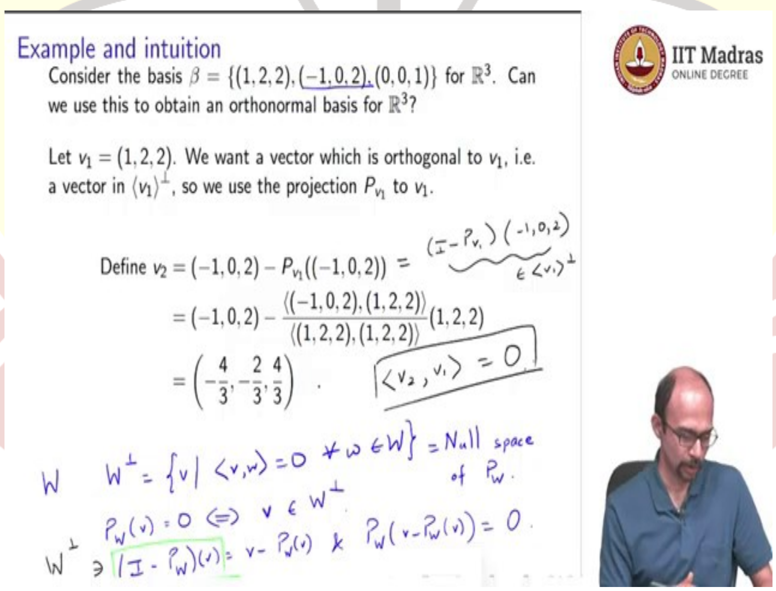
Then you can use the Gram-Schmidt process to get from these $x_1, x_2, \ldots x_n$, which is a basis, an orthonormal basis, which is, just to recall that each of these vectors are mutually orthogonal, that means the inner product of v_i, v_j is 0 if i is not = to j, and each of these vectors have norm, that means the inner product of vi with itself so vi , vi inner product is 1. So, that is an orthonormal basis.

And in the previous videos we have seen why orthonormal basis are useful, because you can express various phenomenon in very nice ways. In particular you can write down the linear combination of any vector. The constants that appear in the linear combination or the coefficients that appear in the linear combination very nicely, in terms of the inner product, and further they represent something called the projection which we saw in the previous video.

So, the project can be written or is the inner product of the given vector with some of these vectors depending on where you want to project and that is one of the, I mean you have to divide/something and so on. So, there was a formula that we found out last time, but that formula was particularly easy if you are, first of all there is a formula if have an inner product, if you have orthogonal basis and if it is orthonormal the formula is particularly, you do not even have to divide/anything.

So, even in this video we are going to use this idea of projection and that is the main sort of key tool in the Gram-Schmidt process.

(Refer Slide Time: 3:05)



So, let us do an example and the intuition behind that, and study the intuition behind the Gram-Schmidt process. So, let us look at this basis for $\mathbb{R}^3$, 1, 2, 2; - 1, 0, 2 and 0, 0, 1. So, first of all you have to convince yourself that this is a basis. So, recall that we know how to get basis for $\mathbb{R}^n$Rn by, if you have given n vectors how to check if it is a basis, you just take those n vectors, put them into a metrics as the columns or it is okay even as rows, and you compute the determinant. So, if the determinant is non 0, then it will be a basis and you can check that indeed this is a basis. So, I will leave that checking to you.

So, the question is can we use this to obtain an orthonormal basis for $\mathbb{R}^3$? So, now we want to convert from this basis to another basis which is an orthonormal basis. Now, as we have mentioned earlier in the convention here is that if no inner product is specifically mentioned then that means you are looking at the standard inner product, the dot product of $\mathbb{R}^3$. So, this is an orthonormal basis with respect to the dot product. So, that is what we want to obtain, an orthonormal basis with respect to the dot product.

So, we start/taking 1, 2, 2, well if you want to get something, so let us start with the first vector. Now we want a vector which is orthogonal to v1, so let us first try to get an orthogonal basis. So, that means we want to get v_1, v_2, v_3., so that they are mutually orthogonal. Then once we have an orthogonal basis we know how to go from there to orthonormal,/dividing/each vector/its norm.

So, let v_1 be 1, 2, 2; we want a vector which is orthogonal v_1. So, last time we saw that vectors which are orthogonal to v_1 are exactly those which are in, what we call v_1 perp or in general if you have a subspace w, then vectors which are orthogonal to all vectors in w are exactly, that set which happens to be a subspace is exactly the set or the subset w perp. So, we want a vector which is in v_1 perp or span of v_1 perp. So, we use the projection P P_{v_1} to v_1. How do we use this?

Well, here is how we use it. define v_2 to be, you take the second basis vector - 1, 0, 2, so you take this vector - 1, 0, 2 and you subtract the projection of - 1, 0, 2 onto v_1 from - 1, 0, 2, that is what you do, and this is a standard computation which is something we have done in the previous video namely P1 of - 1, 0, 2 is some scalar multiple of v1 and what is that scalar multiple? It is the inner product of - 1, 0, 2 with 1, 2, 2 divided/the norm square of 1, 2, 2.

So, if you compute this you get, the numerator is - 1, + 4 so 3 and the denominator is 1 + 4 + 4, so 9. So, 3/9 is 1/3. So I get - 1, 0, 2; - 1/3 × 1, 2, 2 and then if you do the subtraction this is what you get. - 4 /3, - 2 /3, 4 /3. So, first of all why did we do this? So, this was because remember last time that we saw that W perp, so we have any subspace W, then you have this other subspace called W perp which is all those vectors v, such that v ,.

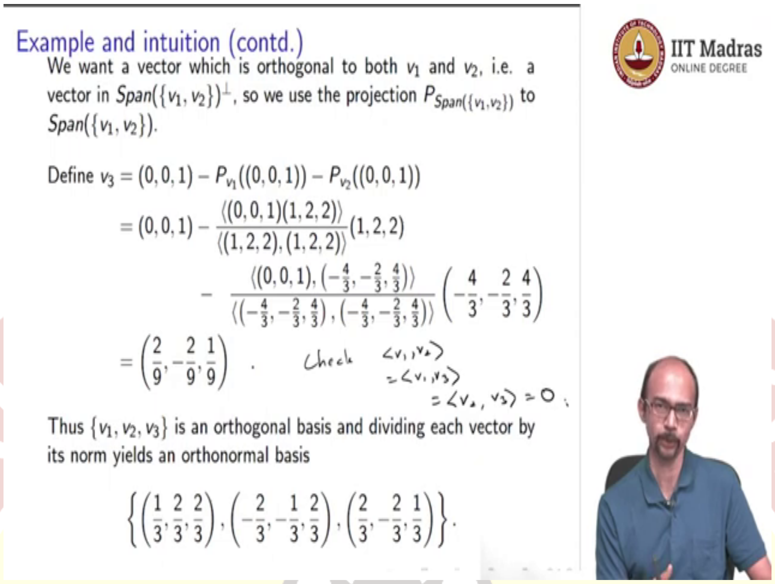
Let me write this carefully, v, such that inner product of v and W is 0 for all w in W. It was this set and then we saw it is a subspace because we saw that this is the null space of Pw. So, what are we doing here? To obtain something in W perp, what we should do is we should get something so that when we evaluate Pw on that then it gives us 0. So, Pw of v = 0 if an only if v belongs to W perp, this was what we saw.

And the point is if you look at the identity - Pw so this is, you evaluate this on v, if you evaluate this on v so you get v - Pw of v and if you again take Pw of this, so now you have already projected, after you apply Pw you have already projected in W, so remember that Pw composed Pw was just Pw because the projection does not change elements in W and that gives us that this is 0. This was something we studied.

Which means that if you take any element of this firm over here, this is inside W perp. So, this is what we have seen and that is exactly what we are using here, we are using that if you do - 1, 0, 2; - P v_1 of - 1, 0, 2 so you can rewrite this as 1 - P v_1, identity - P v_1 of - 1, 0, 2 and this vector will be in v_1 perp.

This is what the idea is for this computation, and that is why we do this computation, we get v_2 and this explanation below is a general method to see that v_2 is in v_1 perp, but you can actually explicitly do this/taking this vector v_2 and computing v_2 , v_1, I will encourage you to check that this is actually 0. So, let us move on.

(Refer Slide Time: 9:40)



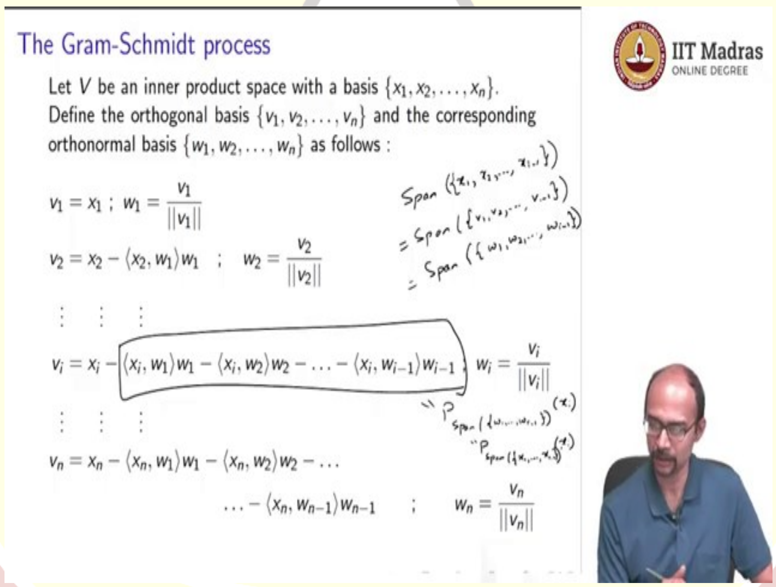
So, now we have v_1, v_2,which are orthogonal to each other, so now we want a vector which is orthogonal to both v_1, v_2, because remember we want to get an orthogonal basis. So in other words we want this vector let us call it v_3 to be in the span of v_1, v_2 perp. That is what we want, and how do we do this? When we again use the projection and this is exactly the idea that we had in the previous slide. You take this span of v_1, v_2 as W and you apply, 1 -, I - Pw on some vector. You apply this on some vector and then you are going to get a vector which is perpendicular to v_1, v_2The only thing is you have to be careful that this vector that you have should not be already in the span of v_1, v_2 and that is where this basis comes in. So, the by 3 vector is not in the span of v_1, v_2; how do I know that? Because the span of $v_1, v_2$is the same as the span of the previous vectors that we had, because you see this 1, 2, 2 is just $v_1$and - 1, 0, 2 and v_1, the span, is the same as the span of v_1, v_2. So, that is another thing you should check because they are linear combinations of each other.

So, we know that v_3 or the /3 basis vector is not in the span of $v_1, v_2$and that is how we are going to use it. So, define v_3 to be 0, 0, 1; - P v_1 of 0, 0, 1; - P v_2 of 0, 0, 1. Now P W of 0, 0, 1 is exactly P v_1 of 0, 0, 1 + P v_2 of 0, 0, 1. This is something we saw in the previous video and now you can evaluate this. I will encourage you to do that. It is a little bit of a computation and if you do it correctly, and I hope I have done it correctly then I have got 2/9 , - 2/9 , 1/3. I will encourage you to check that.

And the point is, this is an orthogonal basis in orthogonal basis. And I do not even need to check this because of what I did in the previous slide. I mean I gave you a general theory for how to do this. So, this is an orthogonal basis. Now if you have not understood why it is an orthogonal basis, I will encourage you again to check that this is an orthogonal basis. So, check that v_1, v_2 is $v_1, v_3$is $v_2, v_3$is 0. If you have not completely followed what I said in the previous slide.

So, in that case you will get that v_1, v_2, v_3 is an orthogonal basis and now you divide each vector/its norm. I will encourage you to check what is the norm and then if you do that you get the orthonormal basis 1 /3, 2 /3, 2 /3; - 2 /3, - 1 /3, 2 /3 and 2 /3, - 2 /3, 1 /3. You may remember this basis from a previous video.

(Refer Slide Time: 12:50)



So, now let us do the Gram-Schmidt process. The intuition we have seen already, you start with a basis and then you keep modifying the vectors in that basis. So, as to get an orthogonal basis and further you will be dividing/the norm. So, as to make it orthonormal. Now the Gram-Schmidt process kind of takes this orthogonal basis and divide/the norm in one shot meaning it does not wait till the end to divide. It starts dividing right from the start. The reason being it helps for efficiency, computational efficiency.

So, let us start with the basis $x_1, x_2, \ldots x_n$ and define the orthogonal basis $v_1, v_2, \ldots v_n$ and the corresponding orthonormal basis $w_1, w_2, \ldots w_n$ as follows. So, we v_1 is x_1, w_1 is v_1/norm of v_1. This is what we saw in the previous example. v_2 is x_2 - x_2 , w 1 inner product, × w_1. So now why did not I, in the previous slides, in previous example I took v_2, x_2 - $-x_2$, v1,inner product and then divided/norm of v_1, into v_1.

So, the thing is if you already convert your basis vectors into the orthonormal form then you need not divide/their norm because you know that the norm is 1, and this is what I meant/saying that there is some computational efficiency. So, you do not have to keep dividing/its norm. And then w_2 is v_2/norm of v_2, so you have, v_1 and v_2 is an orthogonal basis with span the same as x_1, x_2 and w_1 and w_2 is orthonormal basis which, I should be careful, it is not a basis, so I take back the word basis.

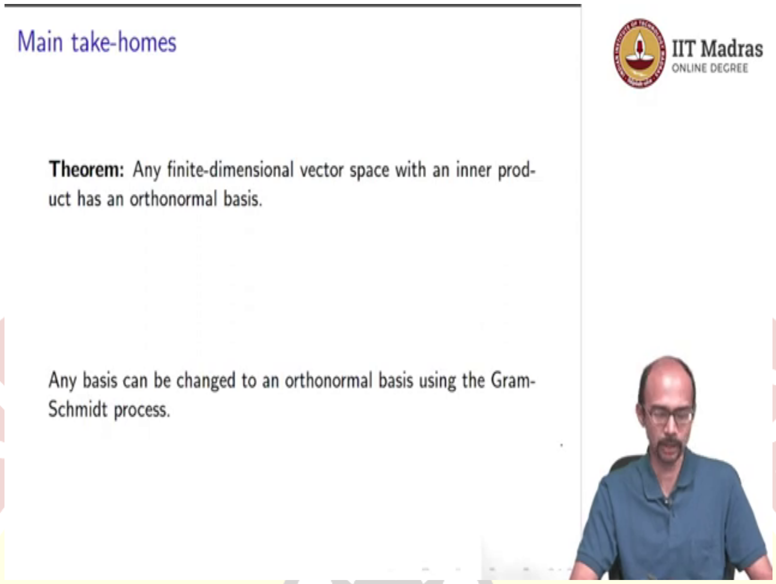
These are linearly independent vectors, but which are orthogonal and w1 and w2 are orthonormal vectors. So, they are automatically linearly independent. So, we will keep going. What is the general term? So, the i th term will be given there, suppose you have reached $i - 1$ at stage, if you have v_{i-1} and w_{i-1}, all the vectors are still there. So, vi is x_i - inner product of x_i , $w_1 \times w_1$ - inner product of x_i , $2 \times w_2$ - all the way up to inner product of x_i , $w_{i-1} \times w_{i-1}$, and then w_i is you normalize this vector.

Which means if you take vi and divide/its norm, and you keep going till the end and in v_n you have all the previous vectors defined till n - 1 and then v_n is x_n - inner product x_n , $w_1 \times w_1$ - inner product x_n , w 2 × w 2, all the way up to inner product x_2 , $w_{n-1} \times w_{n-1}$, and then w n is , $v_n/$norm of v_n.

So, now the main point here is that if you take this basis w_1, w_2, w_n it is an orthonormal basis. v_1, v_2, v_n is an orthogonal basis and the reason this works, this is the main point, is that span of the set x_1, x_2, x_{i-1} = the span of the set v_1, v_2, v_{i-1} and which is again = to the span of the set, w_1, w_2, w_{i-1}. And because these spans are the same we know that x i does not belong to this span and we can carry on.

So, the main take-homes in this, maybe before I come here I will point out also that this part here is exactly = to P span of w_1 through w_{i-1} of x_i. This is the projection of x_i, on to the subspace, spanned$/w_1, w_2, w_{i-1}$, and this comment appear, tells you that this is as a result the same as the projection of x_i on the span of x_1, x_2, x_{i-1}. So, really this entire thing is about projecting carefully and as we know projections give you things which are orthogonal. This was the main point of the previous projection's video.

(Refer Slide Time: 18:05)



So, with that brief comment I will go to the next and final slide which is what the main take-homes of this entire video. So, the main take-homes are that if you take any finite-dimensional vector space with an inner product, so in other words a finite-dimensional inner product space, then it has an orthonormal basis, this was not something obvious from the previous videos that you can always get an orthonormal basis, but now not only can we get an orthonormal basis we in fact have an explicit process. So, any basis can be changed to an orthonormal basis using the Gram-Schmidt process.

So, how do you get an orthonormal basis? You start with any basis, so the assumption here is that you can get a basis and indeed we have seen in previous videos how to get a basis and, when I say previous videos, this is a few, maybe a week or so back or even before that and then now once you get a basis you can change it to orthonormal basis/using this Gram-Schmidt process.

So, the idea often is that when you have a, I mean when you are doing, using this kind of mathematics, typically you happen to know a basis, yeah, you will somehow be able to construct some basis and then it is some algorithm, namely the Gram-Schmidt process, that will allow you to actually get an orthonormal basis, and once you get an orthonormal basis expressing any vector in terms of this orthogonal basis becomes very easy. That is what you have seen in the previous videos. So, with this I guess I will end this video. Thank you.