IIT Madras

ONLINE DEGREE

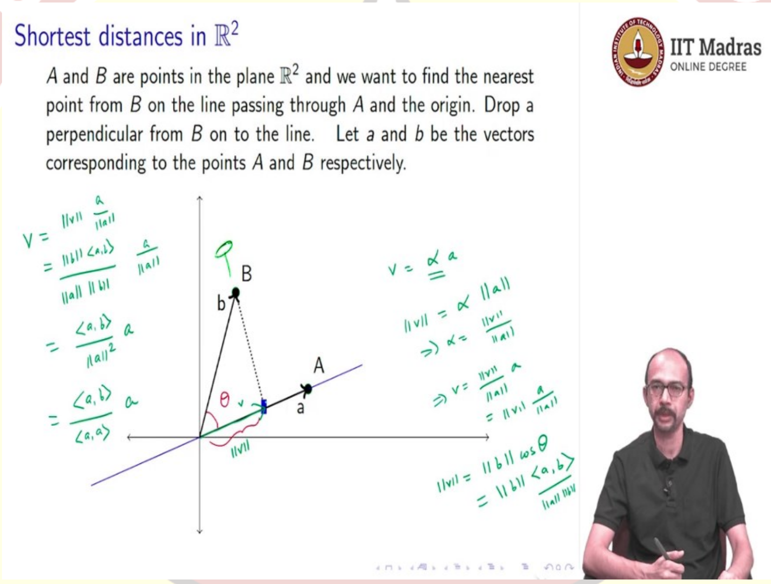
Mathematics for Data Science -2

Professor Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras

Projections using inner products

Hello and welcome to the Maths 2 component of the online BSc degree in Data Science and Programming. In this video we are going to talk about projections using inner products. So, we have seen before what is an inner product space. So, it is a vector space endowed with an inner product and we have seen that the dot product or the Rn is an example of an inner product. So, in this video we are going to describe the ideas of taking a projection using the inner product.

(Refer Slide Time: 0:47)



So, let us start with a basic example in $\mathbb{R}^2$. So, here is your plane $\mathbb{R}^2$. You have two points A and B and you draw the line joining A with the origin. So, that is the blue line in the picture and now we want to find the nearest point from B on the line passing through A and the origin. This is what we want to do. So, this is maybe something that you have done in high school meaning in your, I am not even sure seventh or eighth standard where we do geometry.

We actually do this using the ruler and compass where we take the compass, place it on B, then you draw an arc, it intersects the line in two places, then you take your arc again, you draw, your compass again, draw two arcs and then you join this and that sort of, that is the perpendicular passing through this line and that perpendicular is actually the shortest distance of B from this line. This is what, this is an, these are the idea that we have seen in high school geometry.

So, now I want to do the same thing using a little bit more sophisticated machinery. We do not actually want to draw this we are not interested in drawing it per say. What we want to do is to determine it, is there a way of writing down equations I order to get the shortest distance, is there way to write down equations to get the nearest point?

So, drop a perpendicular from B onto the line. You can use the procedure that I just mentioned in order to do that. So, the procedure that we have using the compass and so on that will tell you what the point is; meaning geometrically you can identify what the point is, but what we want is, we want to know for example, what are the coordinate vectors of this point? We are now, we have enhanced ourselves from Euclidian geometry to coordinate geometry and now in fact we are going towards vector geometry. That is what we are studying now.

So, let us see how to do that. So, now we introduced vectors. So, this is where vectors start coming in. So, draw the vectors corresponding to the points A and B. Let us call these vectors little a and little b respectively. And now let us use these vectors in order to determine what is this point. So, I want to determine what is this point. So, let us me highlight that point. It is this point here. That is the point that we want. So, let us see how to get that point.

So, suppose I know that this angle is θ or I do not know the angle is θ but I know the length of this line. If I know the length of this line, then I am done. Because the point A is given to me that is, so that mean I can get the vector a and then what I want to do is, I can, I want to scale this vector, because this is the vector corresponding to the point that I want. I will show that with green.

So, the vector that corresponding to the point I want, that is this vector here is some scaled version of A. So, the vector I want, let us call it v, so v =some α × a. This v is α × a. So, all I need to do is to determine what s α and then I can get what is this point. So, and what is α? A is exactly going to be determined/this length in red that we have denoted here.

So, how do I get α? So, what we know is, so suppose this length is given by, so this is essentially in norm of v; the distance of v from the origin. So, this is a norm of v. So, the norm of v is what we want. So, if we put it into this equation, this is mod of α, but we know that α is actually positive, so we do not need to put the mod, α × norm of a, which means α =norm of v divided/norm of a.

So, in other words, v =norm of v divided/norm of a × a or in other words, norm of v × a/norm of a. So, you can see a/norm of a is a unit vector, meaning it is a vector of length 1 and then you, when you multiply in the same direction as a and then we, when you multiplied/norm v, you will get exactly v. Fine.

So, now I know a because I know the point capital A and once I know A, I can determine what is norm of a. So, what I really need to know is what is norm of v? And for norm of v, well we know that norm of v =the norm of the vector b × cosine θ. This is something we know from our Euclidian geometry or this is the definition of cosine of an angle. It is the length of the adjacent side divided/the length of the hypotenuse.

So, I know this is norm of b × cosine of θ, but now I know what is cosine of θ, that is exactly where, so, so far I have only used the lengths, but I will use the inner product because the cosine of θ, remember θ is also the angle, it is not just the angle between b and v, it is also the angle between b and a. And so, I can write down cosine of θ as the inner product of a , b. I mean, in terms of the inner product of a , b. So, this is a , b divided/norm of a norm of b.

If I put all of that together, what we are going to get is that v =norm of v × a/norm a which =norm of b × inner product of a , b/norm of a norm of b, a × norm of a and these two norm of bs cancel, so this is inner product of a , b/norm of a square × a and I can write this, since I have a norm of a squared, I can write this as a , b, inner product of a , b divided/inner product of a , a × a. So, that is what we get.

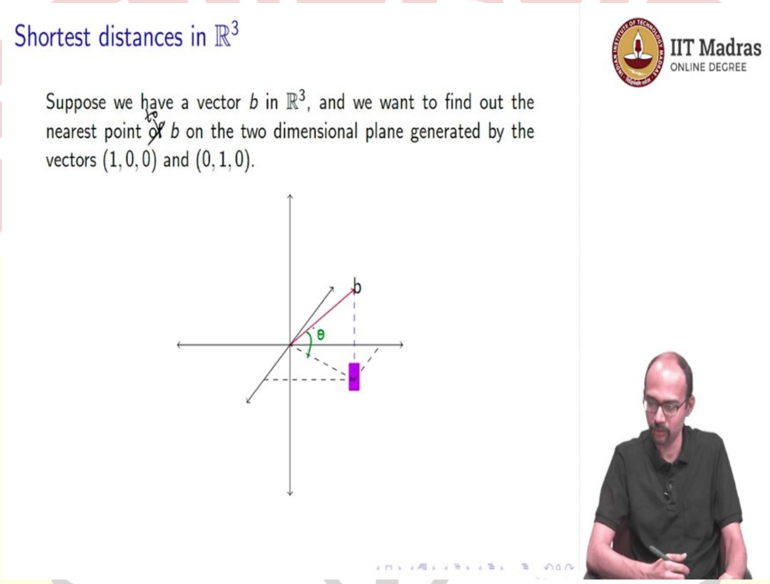
So, if you know a and b, well, I know what is this vector v which is the vector corresponding to the point that I wanted, that was the nearest point on this line. So, you can see how we have progressed from Euclidian geometry to coordinate geometry to vector geometry which helped us solve the problem in coordinate geometry. And the point here is that we used the inner product very crucially to get what we wanted.

So, in this case of course inner product is just the dot product. So, when I said inner product here, I mean the dot product. So, this is ab cosine θ. I mean, this is, so that is what I used here. So, a dot b =length of a × length of b × cosine θ. That is what I used here. So, this is a rather expansive discussing about the nearest point. So, let us also recall what the nearest point tells us in terms of shadows because that is really what projections are.

So, suppose there is a light source over here which is emitting light. It is right on top of B, on top in the sense of this perpendicular and if you have the light coming out like this, then the shadow of this vector is going to fall exactly where the green line is on v. So, this is what a projection does. It takes something and projects it down to something else. It is as if there is a light source behind.

And when we do a usual inner products and usual lengths and so on, then that means it is perpendicular, it is like the shadow is falling exactly perpendicular to what is below. But we will see things which are not like that also and there we will actually have to use inner products, the dot products would not be helpful. That is what this video is about.

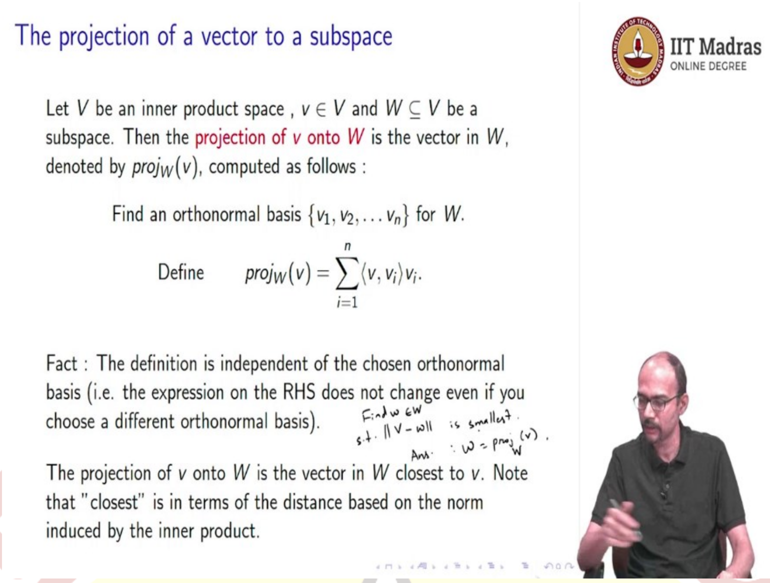
(Refer Slide Time: 10:02)



Let us quickly look at the example with the same idea. So, here we have a vector b in R3 that is the vector in red and we want to find the nearest point to b, this is not of, this should have been to, so the nearest point to b on the two dimensional plane generated/the vectors 1, 0, 0 and 0, 1, 0 that is the xy plane. So, what you do, you drop a perpendicular, again the same idea and then that point is exactly the point which is closest. So, this point here is the closest point.

And how would I determine this point now? So, to determine this point, again you would go through the same ideas that we did before. So, we would consider this angle, this angle is θ and then you would work out the same, work it out in the same way. Except here we do not have an a. So, instead of a would have to use the vectors 1, 0, 0 and 0, 1, 0 and we will see explicitly how to do that in a few minutes. The idea is going to be the same, we have to use the dot product and the lengths in a very effective way.

(Refer Slide Time: 11:26)



So, let us go ahead. What is a projection of a vector to a subspace? So, we have seen two examples of taking this idea of shadows and now let us do a general definition and we will see that the particular case of Rn with the usual the dot product and so on gives you the shadow that we just saw in the previous examples.

So, let V be an inner product space, you have a vector v, you have a subspace W. Then the projection of v onto W is the vector in W denoted/projection of v subscript w computed as follows and I want to underline that I am saying computed and not defined. Find an orthonormal basis $v_1, v_{2,} \ldots v_n$ for W. Define the projection as summation v , $v_i \times v_i$. So, now what happened here?

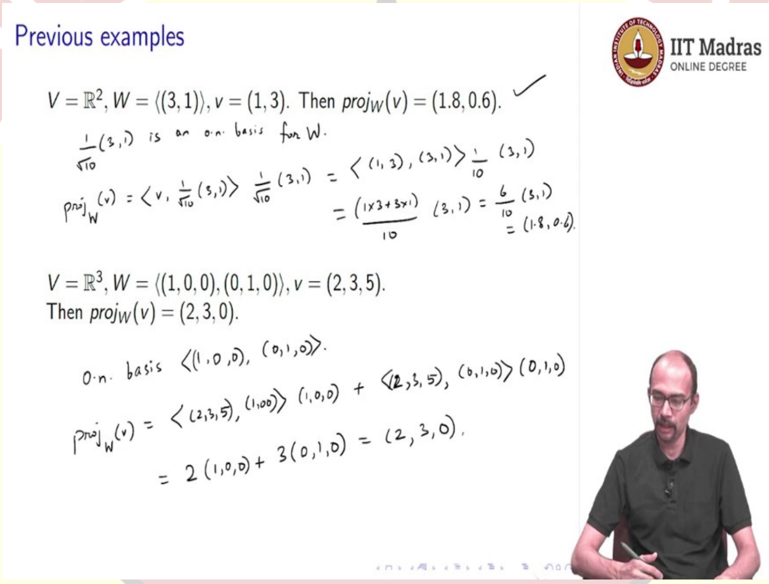
Well, we took an orthonormal basis which is why we did not have to divide/anything. In our first example, remember we had to divide/something, so the inner product of a with itself, so the norm of a squared, but since here our basis vectors are of length 1, we do not have to do that. Here, length means norm. So, here you just take the inner product of v with vi and multiply that number with vi and that is the projection of v onto W.

You might identify this with something we saw in our previous video. So, here is the main fact which is why I said computed, did not defined. The definition is independent or I should have said that expression is independent of the chosen orthonormal basis. What does that mean? The expression on the right hand side does not change. So, when I say the expression that means this expression here.

This expression does not change even if you chose a different orthonormal basis. So, if I choose $v_1, v_{2,} \ldots v_n$ and you choose $w_1, w_{2,} \ldots w_n$, then, even then, this will not change. So, the projection of v onto W is the vector in W closest to v. And what do we mean/closest? So, note that closest is in terms of the distance based on the norm induced/the inner product. So, what we mean is that if you take the smallest value of v-w so find w in W such that this norm is smallest, the answer is w is the projection of v onto W. That is what we mean/the smallest distance.

And here, what do we mean/distance, because distance is inner product space, there is no natural, I mean there is no distance in the same sense of length that we have done for Rn. Here distance means the norm, distance means norm. So, in terms of the norm, this is the smallest.

(Refer Slide Time: 14:52)



Let us do examples. The, so I believe the previous definition we have little intimidating. So, let us do some examples. So, V be $\mathbb{R}^2$ and let W be the subspace generated/3 , 1 and let v be 1 , 3, then the projection of v onto W is 1.8 , 0.6. Similarly, if v is $\mathbb{R}^3$ and W is the subspace generated/1, 0, 0 and 0, 1, 0 so that is the xy plane. And you take v to be 2, 3, 5, then the projection of v onto W is 2, 3, 0.

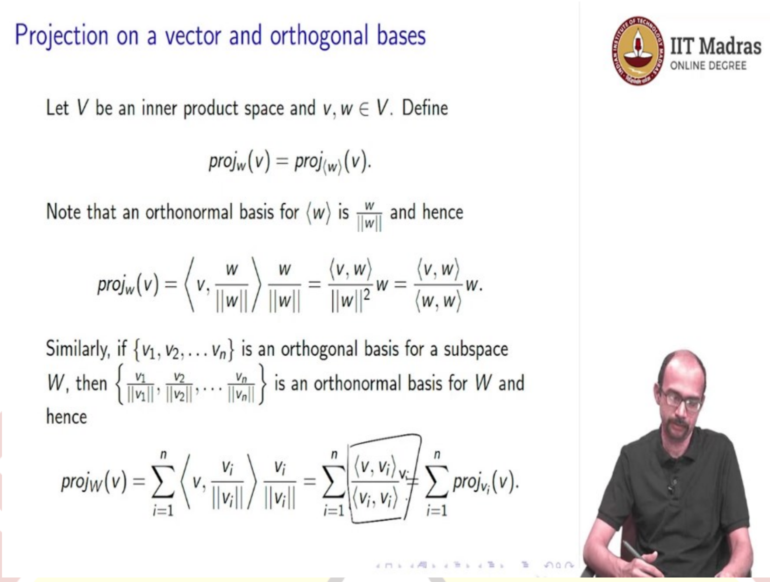
Let us work out these examples and it will make it clearer what we mean. First of all you have to identify an orthonormal basis for the vector space at w at hand. So, here if you take 3 , 1 and divide it by√10, so this is an orthonormal basis for W. Now, let us take v and then/definition you have to take v , this 1 by$\sqrt{10}$, 3 , 1 and multiply it/the basis vector. This whatever value you get, you multiply it to the basis vector. This is 3 , 1 and this is the projection.

So, let us compute what that is. So, if you take this inner product, you get, so this is inner product of 1 , 3, I can take that 1 by√10 out and I have 1 by√10 again here, so I get 1/10 3 , 1 and this inner product is 1×3 + 3 × 1/10 3 , 1. So, this is 6/10 3 , 1 and now if you multiply you get 1.8 , 0.6. So, this is exactly what we add over here.

Similarly, here, where in this case the, you have an orthonormal basis given. So, the, this is the standard vectors, standard basis vectors, so this is our orthonormal basis. So, now what is the projection? So, this is the inner product of 1, 0, 0 or let us write that on the other side 2, 3, 5 , 1, 0, 0 multiply it to 1, 0, 0 + 2, 3, 5 , 0, 1, 0 multiply it to 0, 1, 0. So, if you take this inner product, you get 2. So, this is 2 × 1 , 0 , 0 and if you take this inner product, you get 3. So, this is 3 , 0 , 0. So, you get 2 , 3 , 0.

So, this is exactly what you expect. If you have a vector like this and you take its projection onto the xy plane, that vector will have Z coordinate 0, but the x and y coordinate do not change and that is exactly what you are getting here. That is exactly the picture that we saw in our second example. So, the projection indeed does maintain, does do what we expect it to do. I mean, this definition that would have does indeed do what we expect it to do for the usual Euclidian geometry.

(Refer Slide Time: 18:46)



Let us now ask what happens if instead of an orthonormal basis you have orthogonal bases. We already saw in the previous example that we had I think it was 3 , 1, so we converted it into an orthonormal bases and then we applied the definition and that is what you can do here as well.

So, first of all let us define what is projection of a vector. So, let V be an inner product space and v and w be vectors in V. Define the projection of v on W. So, now this w remember is a vector is projection of the subspace generated by w, so you take projection of v onto the subspace generated/w. So, how do we find this? Well, you take your little w and your little w is a basis vector for the subspace it generates and you divide it by its norm, so that will give you have orthonormal basis.

There is only basis vector here, so there is no question of orthogonality. So, the only thing you have to ensure is that the norm is 1 which you can do/dividing/its norm. Now, I should point out that if it so happens that w is 0, then you do not have to do all of this. The projection is just defined to be 0 because there is nothing else you can do. So, assuming it is non-zero you can divide/its norm.

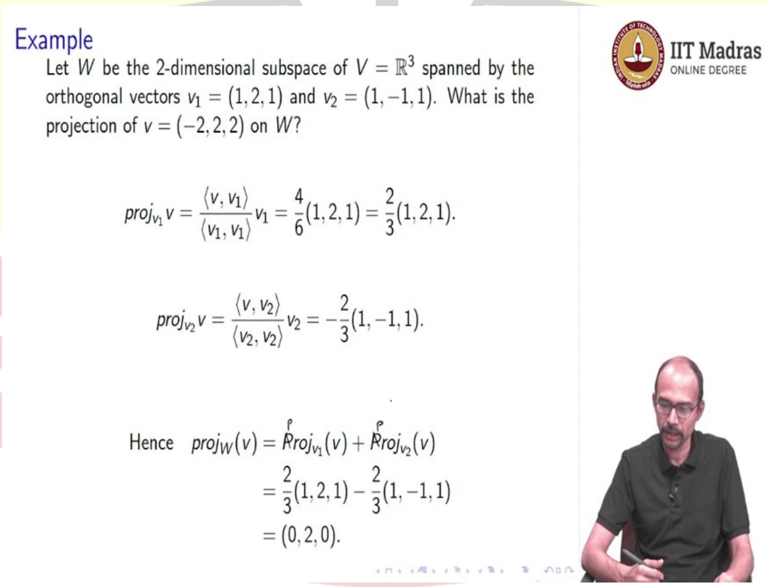
So, once you divide/its norm, you get w/norm of w which is an orthonormal basis for the subspace generated/w. And now you can take the projection. So, the projection of v on W is/definition v , inner product of v , w/norm w × w/norm w, which if you work out, this is what we did in our previous example or as also in the first example with R2; the beginning.

So, here we get the inner product of v , w, you can take the norm out and you get a square of the norm in the denominator from the two norms and then you can write it as v, inner product of v , w divide it/norm of w , w × w. So, this is the expression we got right at the beginning in the first slide of this video.

So, now we can expand this, if you have v1, v2, vn as an orthonormal, orthogonal basis for a subspace W, then you can divide each of them/their norms and that will give you an orthonormal basis, this is what we have seen in the previous video and that is why if you take the projection, that means you are taking the projection with each of those vi/norm vi and then multi, the inner product with each vi , norm vi, multiplying/vi/norm vi and taking the sum.

And if you work it out that means you get summation v ,, inner product of v , vi × inner product, divide it/v inner product of v_i , v_i and here is a typo, this should be × vi and then this thing here is exactly what we call the projection of v on vi. So, you can write it as the sum of projection of v on vi. So, the, so if you have an orthogonal basis, projection of v on W you can just write it as a sum of the projections on each of the basis vectors. So, that is the upshot of this argument here.

(Refer Slide Time: 22:22)

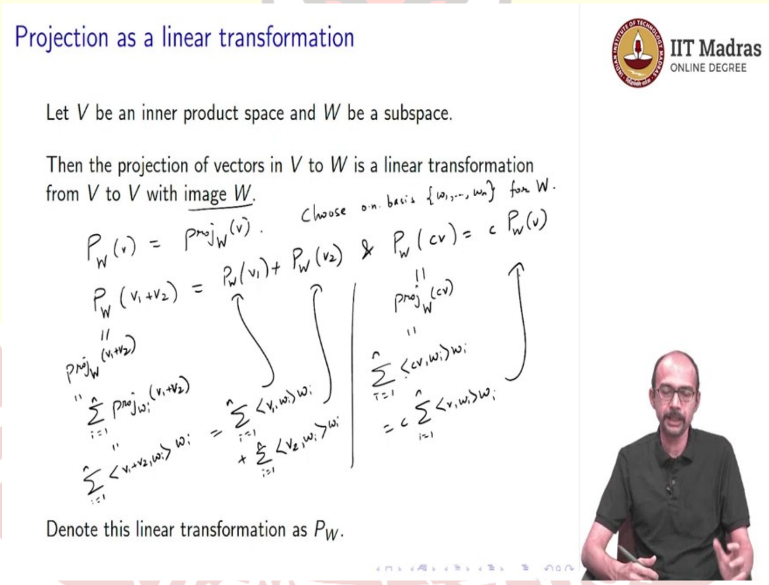


So, let us do an example. So, let W be the two-dimensional subspace of V is $\mathbb{R}^3$ spanned byt orthogonal vectors 1, 2, 1 and 1,-1, 1./the way you do not have to take my word that these are orthogonal. You can take the inner product, I should also again remind you that if no inner product is specified, then/default it is the standard inner product on for $\mathbb{R}^n$. So, here that means it is a dot product. So, if you take the dot product, you get 1,-2, + 1 which is 0. So, these are orthogonal.

So, what is the projection of v is-2, 2, 2 on W? We find out each of these individual projections. So, the projection of v on v1 is/definition as we saw, inner product v , v1 divided/v, inner product v1 v1 × v1 which if you work out gives you 2/3s 1, 2, 1. So, the numerator is-2, + 4, + 2. So,-2, + 4, + 2 is just 4 and the denominator is 1 + 4 + 1, so that is 6, so 4/6 × 1, 2, 1. So, that is what we want.

Similarly, if you do it for v2, the numerator is-2,-2 + 2. So, that is-2, and the denominator is 1 + 1 + 1 so that is 3, so you get $\frac{-2}{3}$ × 1-1, 1 and now the projection of v on W meaning the subspace is where you add these projections up, so that is given by, so this should not be a capital, this was a small, so that is 2/3 1, 2, 1-2/3 1,-1, 1 which gives you 2/3 × 2 + 1 in the second coordinate and 0s in the other coordinates, so that gives you 0, 2, 0. So, that is the projection of v on this subspace W.

(Refer Slide Time: 24:35)



Fine. So, at the end let us talk about projection as a linear transformation. So, we have studied the notion of a linear transformation. So, let V be a inner product space and W be a subspace, then the projection of vectors in V to W is a linear transformation from V to V with image W. So, what do we mean/this? Let us study that in a second. Before that let us remark that we denote this linear transformation as Pw. What is Pw? Let us see what is Pw?

So, Pw of v we are saying is defined as projection of v on W. This is Pw of v. And what do we need for a linear transformation to work, for this to be linear transformation we need that if you take two vectors v_1 , v_2, this is Pw of v_1 + Pw of v_2 and we need if you take Pw of some constant × v, you get c × Pw of v. If we know these two, then this is the definition of Pw being a linear transformation.

So, given a v, we get a new vector in v which is the projection vector and the actually this vector is in W, so it has image W that is what this is saying. So, let us maybe quickly run through these proofs. So, the idea here is the following, we chose an orthogonal basis, so choose orthogonal basis, let us say w1 through w_n for W, then/definition this thing on the left is summation, we just saw this.

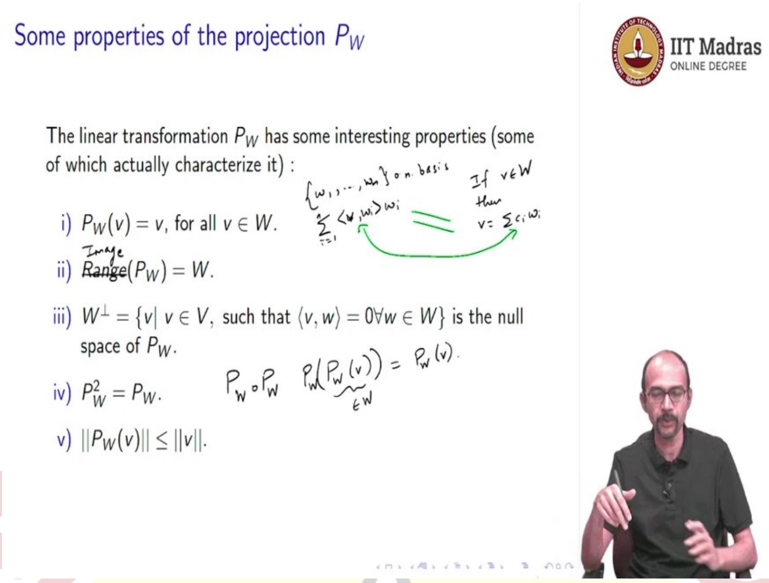
This is summation of, well, first of all/definition this is projection of v_1 , $v_2$on W which we just saw was the projection, the sum of the projections on each w_i, but what is this?/definition this is the inner product of $v_1 + v_2$, $c_i \times w_i$ and then sum 1 through n, but then you can make this into two different terms v_1 , $w_i \times w_i + v_2$, wi, this is v_1 , w_i so × w_i and this is exactly this term, this is exactly this term.

So, that is why the projection of $v_1 + v_2$ is the projection of v_1 + projection of v2 and how about the other the other part? So, Pw of cv/definition is projection of cv on W which again/definition is summation i is 1 through n, so I will directly write this point this is cv , w_i × wi, you can take the c out and then you can remove it all the way out of the sum as well. This c × summation v , $w_i \times w_i$, what which is exactly this.

So, that is the proof that it is indeed a linear transformation. So, what are we saying really, I mean if you think carefully of what we are saying, it is the following: the projection is some kind of taking shadows, so this is saying if the length of your vector is larger, then the shadow is equivalently larger. This is, this makes perfect sense. That is what this second part is saying.

And the other part is saying if you take the sums of two vectors, then the resultant vector, you take its shadow, instead if you take the shadows first and then take the sum, you will get the shadow of the result vector. This again, you think a little about the parallelogram law, this makes perfect sense. So, anyway I have proved it, but this is not I mean even geometrically this is very natural.

(Refer Slide Time: 29:08)



Let us quickly look at a few properties of this projection Pw. So, this is a linear transformation and some of these properties actually characterise it, but I will not expand on what that means. So, first of all if your v belongs to W, then Pw of v is v. Why is that? That is exactly the formula we saw in the previous video, I will come back to this point at the end of this slide.

Range of Pw is W that is what we said the image of Pw is W maybe I should, this is range maybe was not the word we use, we use the word image. That follows from the first point. So, let us define this set called W perp. So, it is the set of those vectors such that when you take the inner product of such a vector with any element of any vector in W you get 0.

So, for example, if you have the usual inner product and on $\mathbb{R}^2$ and you take the x axis, then the perp is the y axis. So, then, so this is true we can define this for any inner product space, so w perp is the set such so defined. And it turns out that this is the null space of Pw and it if you look at this definition carefully, it will take you a minute to prove that. Pw squared is Pw, what does this mean? I will come back to this at the end of the slide and finally norm of Pw of v is less than or = the norm of the v.

Very intuitive because if you take something like this and you are taking its shadow, you are saying that the shadow has smaller length. And now you have to understand here that the length is if you take a projection in the usual sense in the perpendicular sense, then this is clear. But if you take your light source elsewhere, then you remember that your norm is also different.

So, that is why even if you take it here, you could have a very long shadow, but you are computing length in a different way, so this will still be satisfied. This was a slightly tricky part, I will not expand on this because we probably do not need it in this course, but those of you who feel like this is interesting should think of it. Let me come back to this part about the image.

So, first of all this is called the image. Why is Pw of v = v? Well, what is the definition, choose an orthonormal basis, so this is an orthonormal basis and then look at v , $w_i \times w_i$ and sum over i. Now, if v is in W, so if v is in W, then v is summation $c_i w_i$ wi because wis form a basis. But then what is the c_i? We saw in the previous video that these cis are exactly these inner products, so the ci is exactly the inner product v , w_i. So, this =v. That means v =the projection of v on W if v is already a element of W.

Now this makes perfect sense if you already have a vector which is in your vector space, the nearest vector to this will be itself because it is at distance 0, so this is perfectly logical and indeed we have proved it. So, because of this the image is W, as I said the third point you should sit down and check, it will take you a minute.

What is the meaning of Pw squared is Pw? This is not actually Pw squared, what would mean/Pw squared is Pw composed Pw. What that means is if you take Pw and apply it to v, and then this is again a vector of v and then you apply it again to this. Well, but this you know Pw of v is already inside W, so Pw of anything inside W is already itself. So, that is why it is Pw of v. So, that explains the fourth one. Fifth one I will, I have already given some intuitive explanations, but maybe I will not expand on this. Thank you.