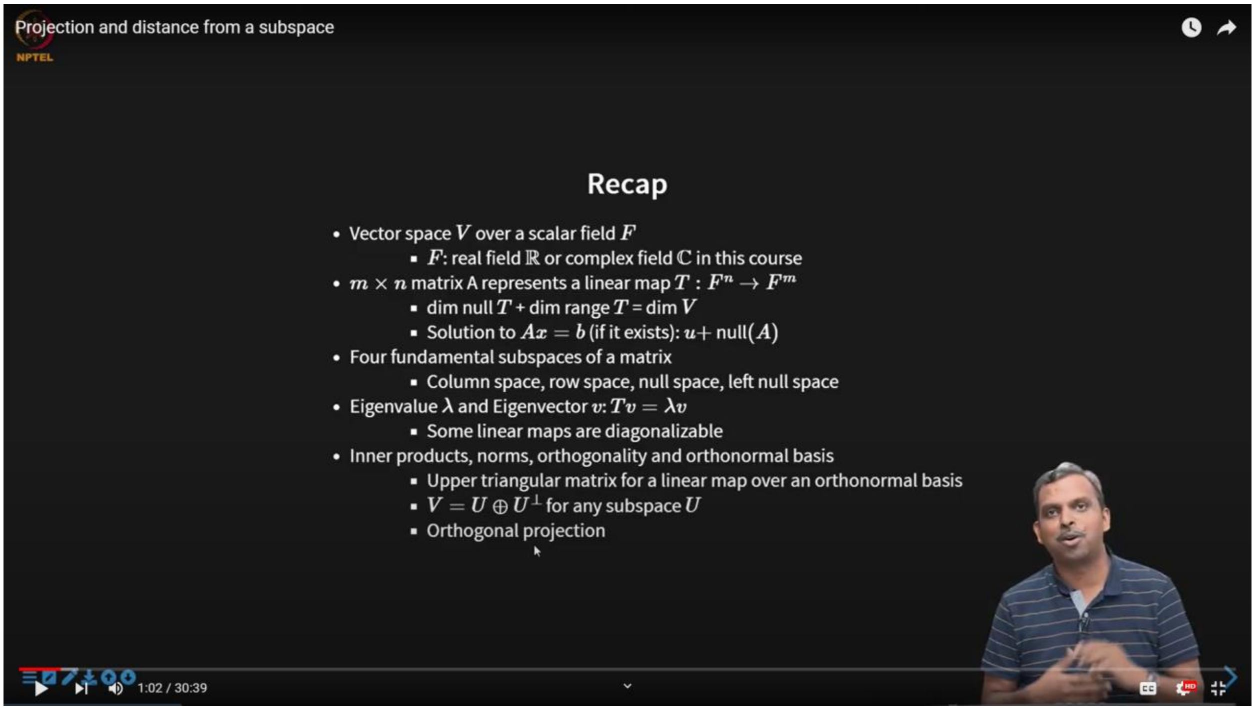
Applied Linear Algebra
Prof. Andrew Thangaraj
Department of Electrical Engineering
Indian Institute of Technology, Madras

Week 08
Projection and distance from a subspace

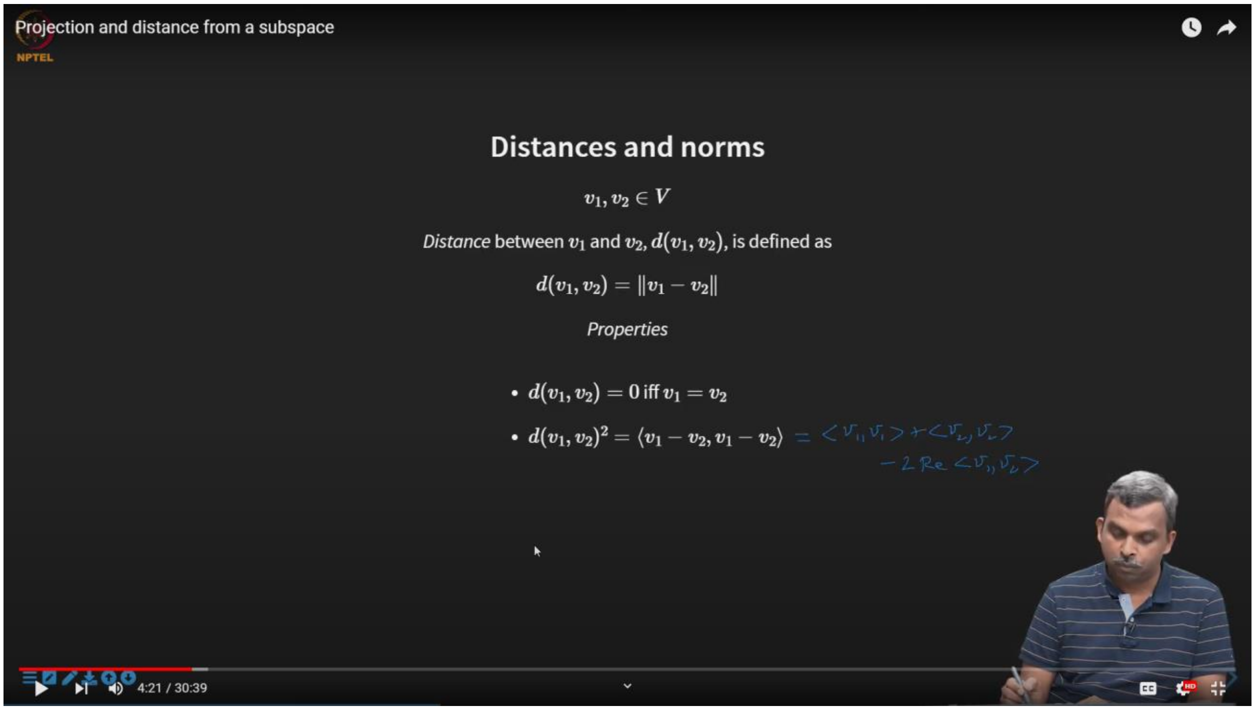
Hello and welcome. We saw in the previous lecture the definition of the orthogonal projection operator, how it is a, you know, linear operator that maps you into the subspace, into a given subspace in a specific way. And there was a very easy way to define that operator and find it using orthonormal basis extension. So that's very nice. But then how do you use it? Does it have any use? It has lots of uses and this lecture will show you one very important use for this orthogonal projection. So it's sort of intuitive. It extends our notion of what we know in two dimensions, two arbitrary dimensions in linear algebra in some sense, okay? So let's get started.

(Refer Slide Time: 01:02)



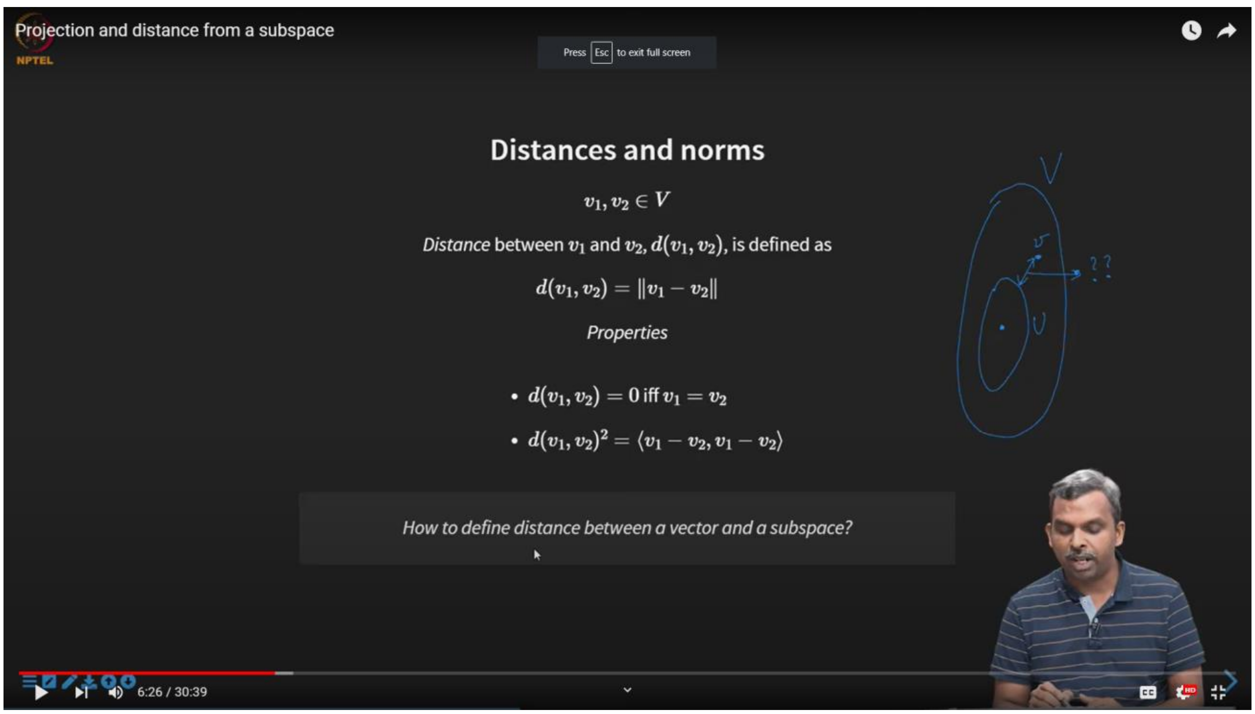
A quick recap as usual. We've been, I'll skip the first four points and jump to the last one, we've looked at inner products and norms and distances and angles and in particular this orthogonality. 90° or whatever, we think of it in two dimensions. Extends to the inner product being zero in higher dimensions. And that seems to have a lot of interesting effects in terms of linear independence, in terms of, you know, giving you this orthogonal complement which, you know, gives a direct sum decomposition of a vector space. That's all nice. Now we have this orthogonal projection of projecting into a subspace using this orthogonal basis, okay? So we will see now that it gives you very interesting properties. I didn't quite emphasize some of the very important properties. But you'll see that the orthogonality gives you a certain minimization which you can otherwise not have, okay? So let's get started.

(Refer Slide Time: 04:25)



Okay. So let us quickly review this notion of connection of distances and norms. I have been always talking about norms. So if you have two vectors v_1, v_2, or two points, sometimes these are addressed as points, the distance between them, how far apart are they, a good measure of that is to use the norm to define. And how do we do that? You do $v_1 - v_2$ and take the norm of that, okay? So this is a good measure of how far away v_1 is from v_2 in space in terms of whatever the vector space may be. If there is a norm, then this $||v_1 - v_2||$ is a good measure of how far away they are. So it's, quite often when you have long vectors, you want to get a sense of whether they are close or not and this is a good way to measure it, okay? So just one number which measures how far away two long vectors possibly are, right? So that is a good thing to have. A quick property based on the properties of norms and the connection to inner product. By the way, we're only looking at norms from inner products in this course. Like I mentioned in one of the lectures before, you may have norms without inner products but that's a different story, okay? So we're not talking really about those kinds of things here, okay? So we will assume the norms come from an inner product. All of that I will sort of assume from now on without saying too much. So I'll assume an inner product space V and assume there is a norm from that inner product that I am considering, okay? So one quick property you can see just based on this definition. If the distance is 0, it happens only when $v_1 = v_2$. It doesn't happen otherwise. That's a property of the norm. That's very nice. And there is this nice connection to the inner product. And in fact you can write this out in some more detail. If you use the, you know, distributive property you get $< v_1, v_1 > + < v_2, v_2 >$, right? And I am going to combine those two. So you will get $< -v_1, v_2 >$ and $< -v_2, v_1 >$. But $< v_2, v_1 >$ is conjugate of $< v_1, v_2 >$. So in fact, you get $-2Re < v_1, v_2 >$, right? So this is, well, I put real part here but, you know, we are assuming its real space and all that. So this is, I mean, whatever it may be, you can do this. So this, real or complex the scalar may be, so this result sort of works out very nicely. So this is the square of the distance. So this will have to be positive. You can check that this will be true. All of that will work out. So this is a nice little way to look at, you know, distance through the inner product. So you can see the norm of v_1. So this is norm of v_1, right? $||v_1||^2$. This is $||v_2||^2$. And this is $< v_1, v_2 >$. So you can see how the inner product also sort of controls the distance between v_1 and v_2 in an interesting way, right?

(Refer Slide Time: 06:27)



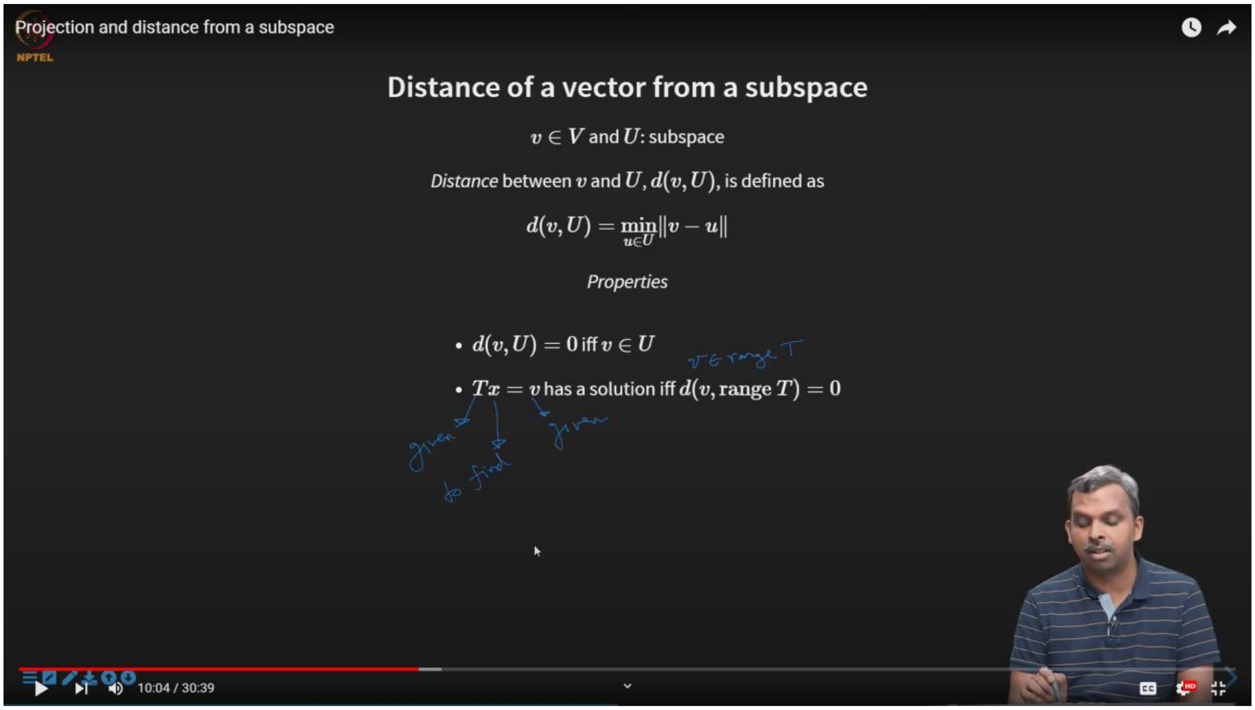
Okay. So this is very well. It's nice. Norm itself directly gives you a very easy way to just define distance between two vectors. One vector to vector another vector. What is the distance? Norm of that difference is very easy to define. Now what about a vector and a subspace, okay? So it turns out this is an interesting question to ask. It may not be apparent immediately. Towards the end of this week's lectures, I'll give you a couple of good applications of this idea. How do you measure distance from a vector to a subspace, right? So once again you sort of visualize how this works. So we are thinking of these vector spaces, right? And this is our typical picture, this egg picture of a vector space. And then you have a subspace U. And then let us say you have a vector v, okay?

What is the meaning of distance between these two guys? So is there a sense in thinking of a distance? Is there a meaningful way to define distance between, how far away is v from U? Given a norm, given an inner product space, is that a reasonable question to ask? So there are some easy rules you can think of, right? So if v is inside U, okay, then the distance should be 0, no? It's already inside, I mean there is no need to think of a non-zero distance. But if v is not in U, then maybe there is, it's possible to have a distance. And why should you worry about it? So it turns out many phenomena are sort of modeled as, you know, subspaces. Linearly closed, right? So you may want to think... I gave some example vaguely a while back about how you make these rank assumptions about the nature that's out there, you know? Choices that people make and all that. So you collect these long vectors together and you assume they are from some low rank space, right? And given a new vector, you are interested in finding out how far away it is from that low rank space. So this is a very natural problem that comes up with data all the time. So this is something that is very interesting. And like I said, I will give you two concrete applications the rest of this week after this lecture. But this is an interesting question to ask. Can you define a reasonable distance between a vector and a subspace, okay? So let us see how one can go about that.

So it turns out it's possible and this is the way to do it. And a very reasonable way to do it is this, okay? You have a point, a vector in V, in the inner product space V. And then U is a subspace. And I want to think of a distance between v and U. I am going to define it in this fashion, okay? So it is the $||v-u||$ and I will minimize it over all u in the subspace, okay? So this is a good picture to keep in mind. You have a subspace which is a set of many points, right? Many vectors are there in the subspace. And you have a vector. You go through every point in the subspace and try to find that point which is closest to v in some sense, right? And then you take the distance between that as the distance from the point to the subspace. It is sort of, this is a minimization problem, right? Clearly it's some sort of an optimization or minimization problem. Over all points u in the subspace, I want to find that point which will give me, which will take me closest to v in some sense and that distance is the minimum distance, okay? So it's a reasonable problem. But this problem looks a bit complicated, isn't it? So, I mean, this U itself is a big set, right? It's usually infinite, has so many vectors, how do you go about doing this? Is there something reasonable? Is there some intuition from small examples? We'll see that there'll be one nice intuition from small examples and we'll try to use that to generalize. And it turns out this problem when U is a subspace and v is a vector can be very easily solved. And projection plays an important role here, okay? So that will be the sort of the climax of this lecture. But still let us proceed slowly and unravel this a little bit one step at a time, okay?

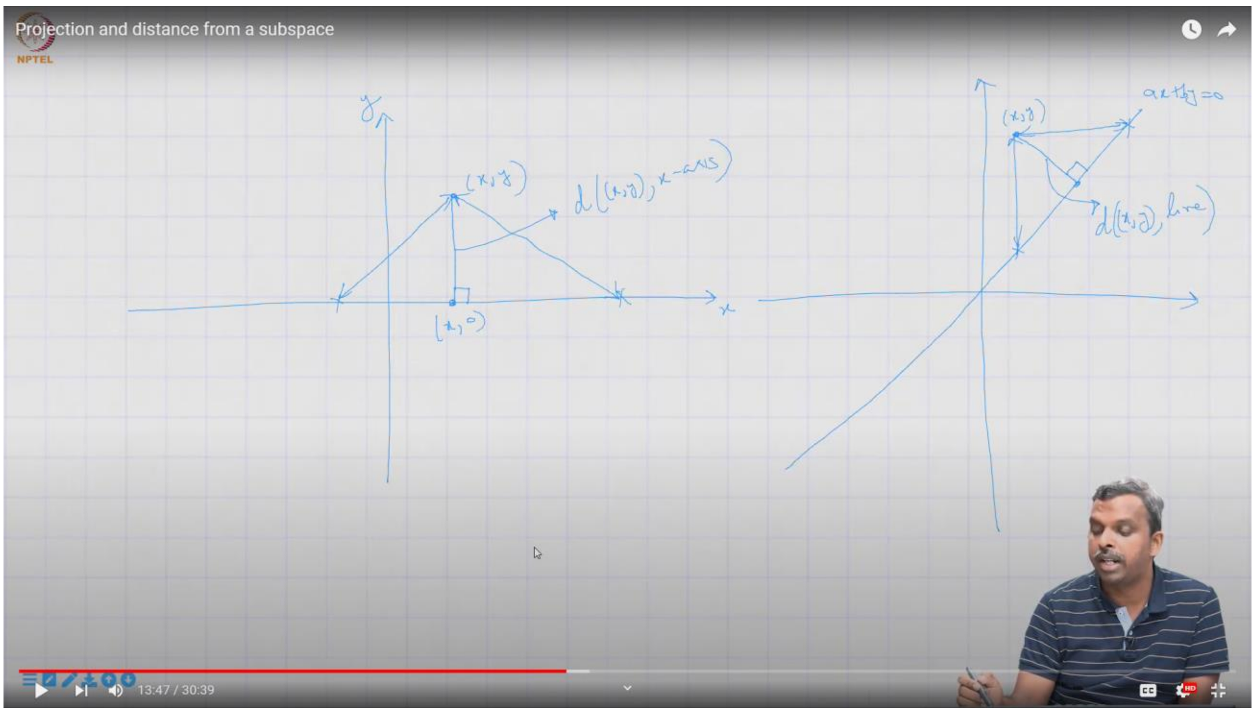
Quick properties, right? Even before we, you know, work out how to compute the distance we can have some properties for this distance. So nice, interesting properties come up. Like for instance, look at the first one. When will the distance be 0? That will happen only when the vector itself is in the subspace. So the distance is sort of like, whether 0 or not determines whether or not the vector is in the subspace, okay? So that's a nice thing to know. So this distance already is proving to be a little bit useful. And this property we can use when trying to solve linear equations. So supposing somebody gives you an operator T, a general operator T and then you have to solve for $Tx = v$. v is also given. v is given, T is given, you have to solve for x, okay? Supposing that's your problem, okay? So these two are given. This is given, you have to find this, to be found, okay? So this has a solution, okay, only if v is in the range, right? Isn't it? If and only if v belongs to range of T, okay? And that v belonging to range of T is the same as distance between v and range of T equal to zero, okay? So, I mean, I'm not saying anything new here, but you will see this we will use soon enough. It seems like a sort of a redundant property, but I am just showing you how, you know, distance being 0 and linear equations are connected in one very nice way, okay? So there is a solution or maybe infinitely many solutions, if and only if the distance between the v and the range of T is zero, okay? So keep this in mind, we will come back to it soon enough. So this definition even before we can compute what it is or see if this minimization is reasonable and can be done very quickly and effectively, already has some good properties for us, okay? So these properties will put them to use soon enough, okay? But anyway it is also doing a minimization, right? So that is always good. When you can do any optimization problem very nicely it's always a good thing, right? So that's a nice thing to do.

(Refer Slide Time: 10:04)



Okay. So let us look at some small examples and build up some intuition, okay? So I am going to go over to... So these are the three examples I'm going to do. A distance of a point from the x-axis, distance of a point from the $x = y$ line and distance of a point from the $ax + by = 0$ line, okay? So in $\mathbb{R}^2$ the only interesting subspaces are lines, right? Other things are not clearly interesting. So only interesting subspaces are the line and let's see what happens when we look at distance from, you know, a point to the line. I mean line through the origin, right? So that is what I am going to see. So let us go to this additional thing and then start looking at some examples. So here is $\mathbb{R}^2$, okay? And then I have a point (x, y) and here is the x-axis, right? What would you say is a reasonable distance or minimum distance from the point (x, y) to the x-axis, right? Which point in the x-axis will be closest to (x, y) and what will be the distance, the minimum distance here? If you pick any point here, this is the distance, right? The distance between these two points and then pick some up some other point here. So that would be the distance. But all these distances are going to be longer. And it will be the shortest, I mean, based on your geometric intuition you can see very quickly that the shortest distance... So this will be $d(x, y)$ from the x-axis, isn't it?

(Refer Slide Time: 13:47)

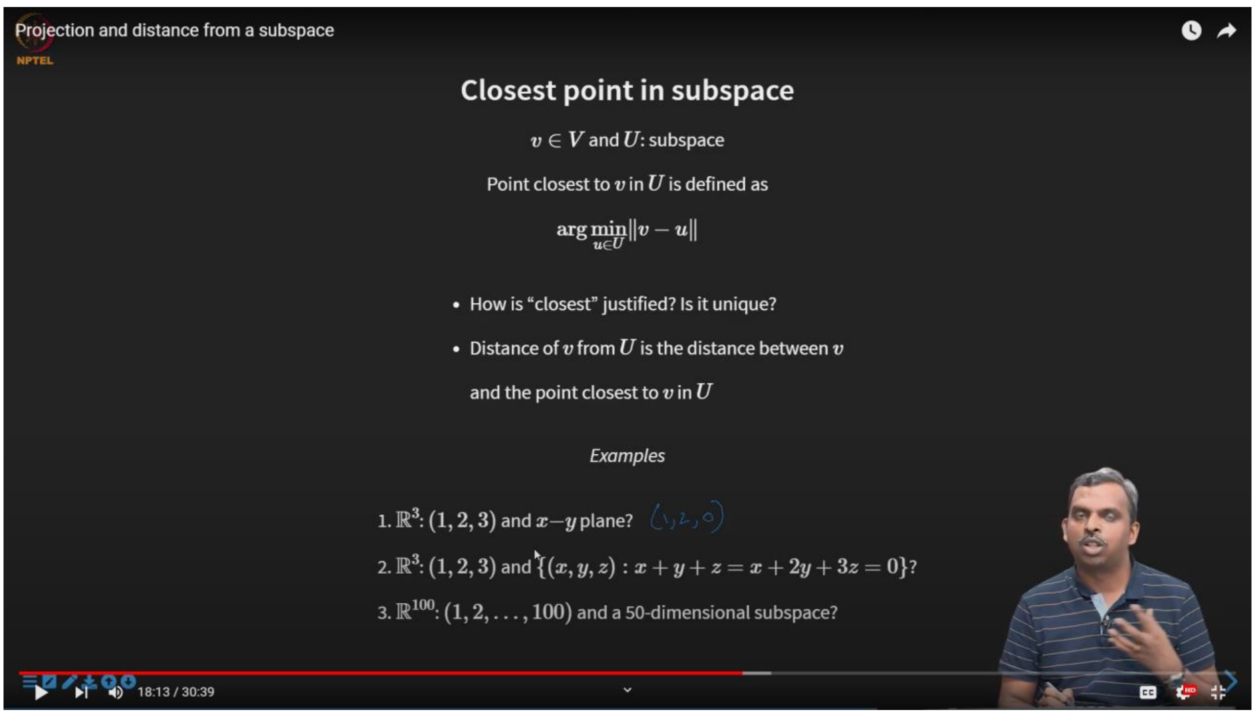


From your intuition from geometry, you can also very quickly prove it, that any other point which is not orthogonal, right, which is not perpendicular will end up having a longer distance, right? So hypotenuse is greater than one of the sides, right? So something like that. Very simple result. So you see $(x, 0)$ is going to be the smallest point and every other example I gave you, right? If you take the next example which was some line through the origin and then you are looking at a point here, right, I just keep it more generic, it was some, you know, let's say this is $ax + by = 0$, okay? And then again you can use the same idea, right? So how would you find the distance from this point? You have to keep doing all these distances, okay? And find that distance or that point which is smallest, you know, closest to this point and gives you the smallest distance and that is clearly... Again it will happen when you have this $90°$, isn't it? So this distance will be $d(x, y)$ from the line, isn't it? It's quite easy to see that that should be the distance. So you see the orthogonality plays a very central role here and that is like exactly the orthogonal projection isn't it? This point is the projection. Here again you see this is the orthogonal projection and that is easy to see and justify in the two dimensional case.

Can we extend it to arbitrary dimensions, right? So that is the natural next question. In two dimensions, finding the distance of a point to a line seems very trivial and orthogonality is playing a very, very central role. One can see very clearly orthogonality plays a central role. So is there an extension of this idea? Finding the closest point in a subspace from a point outside of it? Is orthogonality going to play a role there in the general case is the next natural question and we will explore that soon enough, okay? So we saw all these can be solved. I mean, later on I will give you precise solutions for this. It's easy to work out these things once you know that you have to do an orthogonal projection. Okay. So this, we already saw that this notion of minimum distance or distance between the, minimum distance between a point and the subspace, right? So the distance between the point and the subspace is closely related to the point in the subspace which is closest to the point in question. So supposing you have a v in the, point v, a vector v and a subspace U. I want to find the distance from v to U. Really, I need to find that point in U which is closest to v, okay? And it looks like from the example there is only one point, right? So we can, maybe there are, in general, if you want to look at a distance from a point to a set, maybe there are multiple points in the set which are at the same distance, but it looks like, you know, if you do orthogonality and all that there should be only one point, right? So it looks like that. So let's just make the definition. And again I've already put out how is closest justified. Is it unique? It will turn out to be unique. So that's why this definition is justified. So we'll define point closest to v in U as the argument of the minimum. So even if it is not, you know, unique, maybe you can have a set of points, it’s okay. So this definition makes sense. But we will argue that this closest is in fact a unique point when U is a subspace, okay? So when U is the subspace, all of that works out. So this is the point closest.

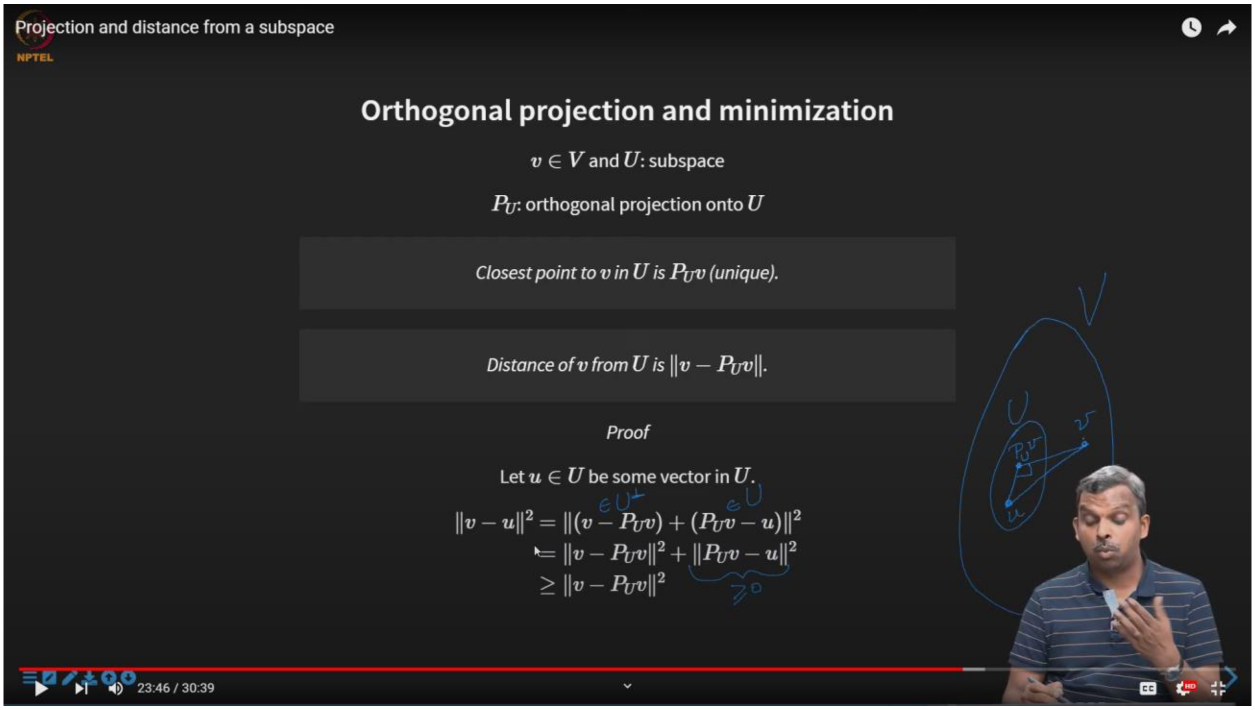
So in the previous example that we saw, we always had a point in the subspace which was closest to the point in question, right? So that point is important. Once you find the closest point, you are done, right? The minimization of the distance is done. Once you find the closest point in the subspace, you know the distance from the vector to that point is the distance that you want. So distance itself is not crucial to find. You really have to find the point which is closest to v in the subspace, right? And yeah, those are the points that I am making, okay? So let us look at a few examples and this time let us go to three dimensions, okay? So previously it was two dimensions, maybe it was too trivial. So let us look at three dimensions. So here is a point. $(1,2,3)$ and the $x-y$ plane, okay? So again I have done $(1,2,3)$ and a line in three dimensional plane, right? So that actually defines a line, okay? $x + y + z = x + 2y + 3z = 0$. Why does that define a line, okay? By now you should know this. This is, I'm defining this using the complement, right? So this whole definition gives you two vectors in the complementary space and they are clearly linearly independent, okay? And so the original subspace itself must have dimension one, right? You can quickly read that off from this definition, okay? So it defines a line. And what is the distance between those two? So these kind of things can be very naturally answered once we know the point that is closest. And how do we find that point that is closest? For $(1, 2, 3)$ and $x - y$ plane it is very easy, right? What will be the point? z-axis should be zero. So this answer seems very, very easy, okay? This answer seems very easy, right? So it's the orthogonal thing. And then you can easily see that that should be true. But what about this guy? How do you go about finding this point? I mean it's not very clear. Maybe there is some orthogonal projection going on. We have to iron that out very cleanly. And look at the third question, okay? So that is the range, the kind of problems you will see in practice, okay? So I put 150. That itself might be actually very less compared to the kind of problems you face today in these kinds of areas, right? You have long vectors and they are all related in some way. Maybe they all come from some small 50 dimensional space or something and then a new vector comes in and you want to find out how far away it is from all these vectors that I already have, okay? So this is a very typical problem today that people deal with, you know, in so many problems in learning and many other applications. So you need to find distance between a long vector and a big huge subspace of some other dimension. So how do you, so you really need a very effective and simple operation that can do this. And it turns out all that is true and the orthogonal projection plays an important role here. So let us come back to this problem a little bit later. Even the second one we'll come back to a little bit later. I will give you a precise solution.

(Refer Slide Time: 18:13)



But before that let's see what the general answer to this question is. Given a vector space V and a subspace U, what is the point in the subspace which is closest to v, okay? So can we have an explicit expression for it? It turns out the answer is yes. Let's see that, okay? So this is the connection between orthogonal projection and the minimization and point closest to a subspace that we've been talking about. If you have a point v in the inner product space V and if U is a subspace and you have P_U being the orthogonal projection operator onto U, okay? You know how to find it, right? If you do an orthonormal basis extension and simply do, it does not matter which basis you pick, you know, I showed you an example how two different bases give you the same answer, okay, any basis you can pick and then you do that, you know, projection using the part of the orthogonal basis which is part of U, right? The orthonormal basis for U, just take an orthonormal basis for U and you can define the projection operator, okay? You don't really need to do the extension, no? Extension gives you the basis for $U^{\perp}$. You don't need $U^{\perp}$, you only need U for doing the projection onto U, okay?

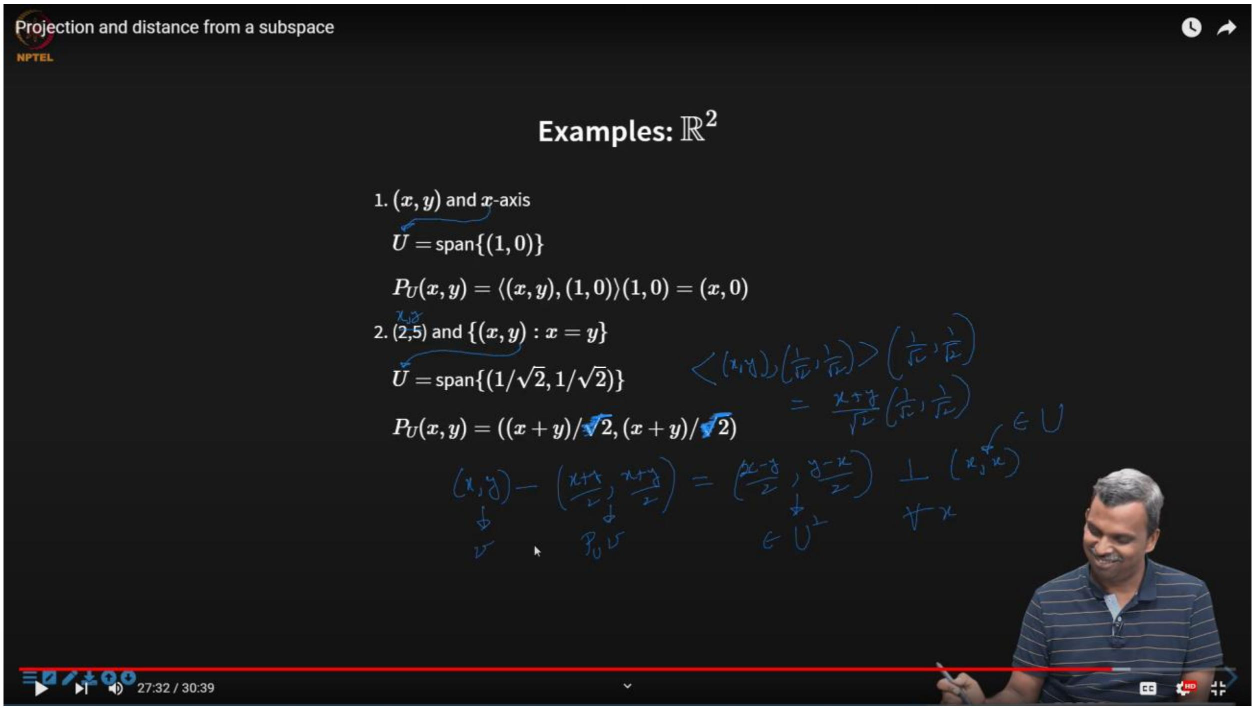
(Refer Slide Time: 23:47)



So it turns out, here is the result, the closest point to v in the subspace is $P_U(v)$, okay? And that is unique, okay? So you project v onto U using the orthogonal projector. Using the orthogonal projector. You cannot use some other operator to take it to U. You have to use that specific, you know, unique orthogonal projector that you have to U. And then you project v onto that, you get the point closest to v. So once you get the point closest to v, distance is trivial, right? So it's just the distance between $v - P_U(v)$. So you will see, I won't obsess too much about the distance, I will only worry about the point that is closest to v, okay? So it is sort of an extension of the two dimensional idea, right? In two dimensions, we saw there was a line and a point. How do you find the point closest? You simply drop a perpendicular line and then that point would be closest. The same thing extends to any dimension, okay? Any dimension you have a subspace, you have a point outside, you do an orthogonal projection, that point is the closest to this point, okay? It's a nice thing to know.

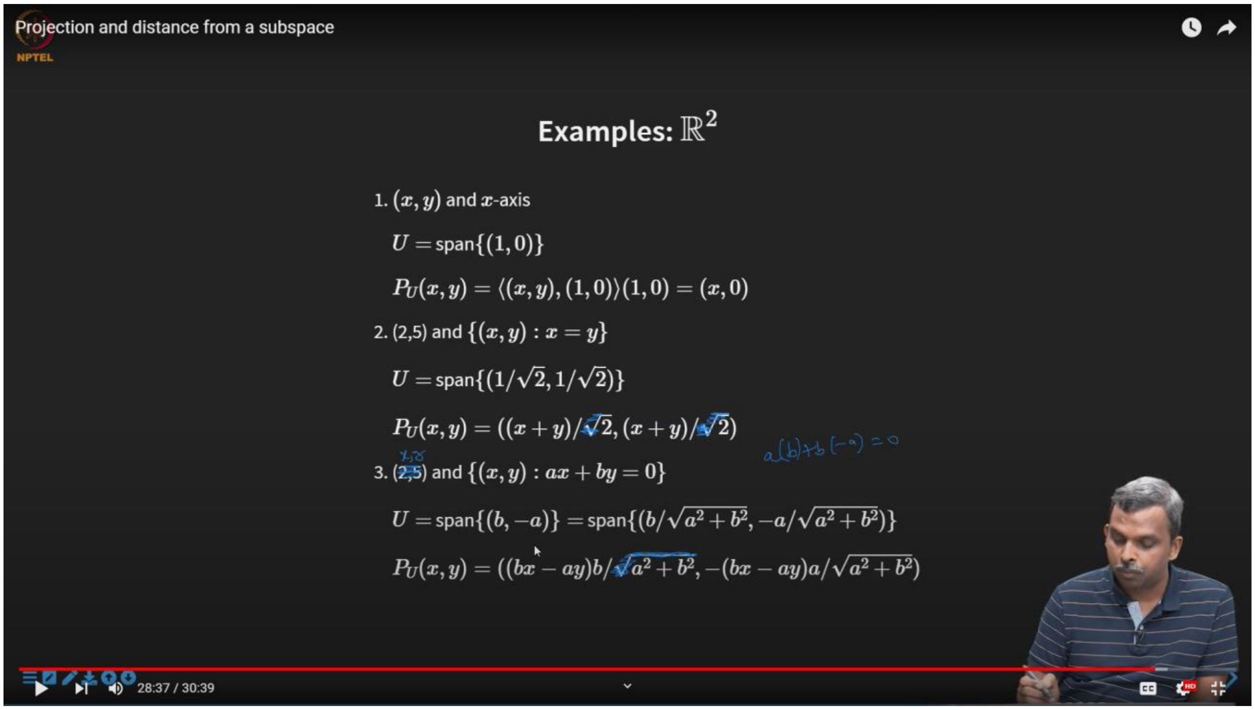
The proof is actually ridiculously simple, you know? I mean it's just four lines. So the proof goes like this. I'll just draw a picture here to sort of illustrate what the proof, how the proof works. So you have the vector space V and you have the subspace U and let us say you have the point v. And here is $P_U(v)$, okay? And this has some orthogonal behavior, right? So I am putting this orthogonal because I mean, I'll say what that means. So the proof involves taking another point u in U, okay? And then looking at this distance, okay? So if I can show the distance between u and v is always greater than or equal to the distance between $P_U(v)$ and v, I'm done, isn't it? So $P_U(v)$ is another point in U. v is a point outside of U. And if you take any other point u in the subspace U and you show the distance between u and v is strictly greater than or equal... Not strictly greater, greater than or equal to $P_U(v)$, right, for u being an arbitrary point, then this is the smallest, and you're done, okay? And it turns out that's very, very easy. So you look at the distance between v and u. That's $||v-u||^2$, okay? Squared distance between v and u. I will do this addition and subtraction of $P_U(v)$, okay? I will subtract $P_U(v)$ and add $P_U(v)$. I have not done anything. And then I do this grouping. When you do this grouping, something very, very interesting happens. This guy $v - P_U(v)$ we know belongs to $U^\perp$, isn't it? $v - P_U(v)$ belongs to $U^\perp$, right? So once you project, that belongs to U and what's remaining belongs to $U^\perp$, okay? So the fact that the remaining belongs to $U^\perp$ is very, very important. That's why the orthogonal projection is important. When you do the orthogonal projection, what remains belongs to $U^\perp$ which is orthogonal to the, you know, the original subspace on which you projected to. So that's very central here. And you see how it plays a role. What about this guy? $P_U(v) - u$? Where does it belong? Does it belong to any subspace that we know of here? U, $U^\perp$, V? Or can we not say? It turns out it belongs to U, right? That is very easy to see. u is in U, $P_U(v)$ is also in U. So subtract, subspace is closed under subtraction, so it better be inside U, okay? So now notice you have two vectors and you are looking at the distance or the norm of the, norm square of the sum of two vectors and they are orthogonal, okay? And you use your familiar and famous Pythagorean law, this will be equal to the norm of the first one squared plus the norm of the second one squared. So that is where the orthogonality comes in, okay? So you notice this orthogonality here. So $v - P_U(v)$ is going to be orthogonal to $P_U(v) - u$. So this distance becomes like a hypotenuse and that is going to be greater than any one of these guys. And that you can easily see. So this one is greater than or equal to zero. Norm is always non-negative. So this would be greater than or equal to this, okay? When will you have equality? You have equality only when u is $P_U(v)$. So that's unique in some sense. So the projection is the only thing which will give you the minimization in the distance, okay? So isn't that wonderful? So you have this nice little simple orthogonal projection operator and that maintains the orthogonality in any dimension, okay? So that is a nice extension of, we saw this simple two-dimensional picture and that extension is this picture, okay? So let's see some examples.

(Refer Slide Time: 27:33)



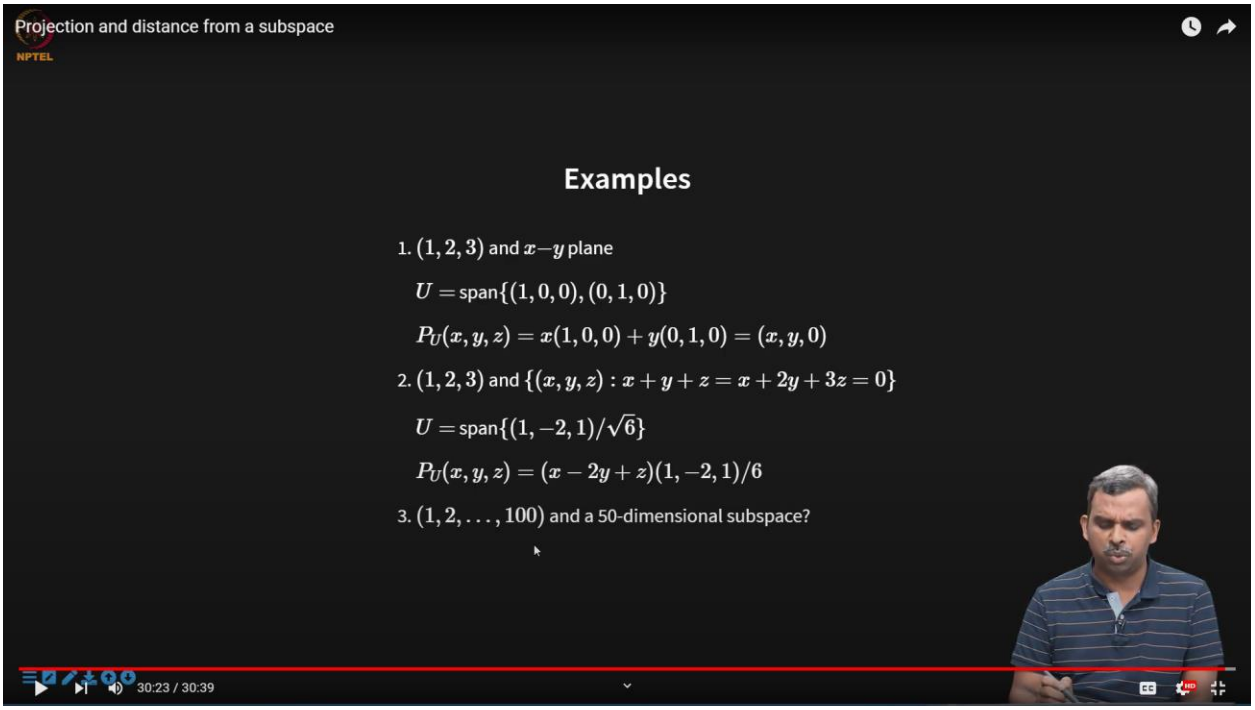
Go back to the $\mathbb{R}^2$ example. It's the, let's see the $x - y$ and x axis. x axis is U is the $span\{(1,0)\}$. $(1,0)$ is the orthonormal basis. Once you have the orthonormal basis, you know how to find the projection, right? For every basis vector, you do an inner product of the vector with that basis vector multiplied by the basis. And $< (x,y), (1,0) >$ is simply x. $x(1,0)$ is $(x,0)$, okay? So we knew this. This was just to show you how it works. Let's do a slightly more complicated example. $(2,5)$. So I put $(2,5)$ but I have actually done for (x,y), okay? So maybe I should write that down, it's (x,y) and the $x = y$ line, okay? Now the $x = y$ line, this is U. Even here this x axis is U, $x = y$ line is spanned by $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Why am I picking $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ instead of $(1,1)$ because $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ is orthonormal, okay? $(1,1)$ is not orthonormal, okay? So well there is only one vector. You just have to make sure that the norm is 1, right? So $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. So then what is the projection? It's simply, you know, $< (x,y), \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) > (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$ isn't it? Okay? All right? So does that work out okay? Did I make a mistake here? Okay, so you have $x + y$, this inner product will work out to $\frac{x+y}{\sqrt{2}}(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. So I have a, I have missed a $\sqrt{2}$ here, isn't it? So this root will not be there, okay? $\sqrt{2}$ will not be there. So it is just $(\frac{x+y}{2}, \frac{x+y}{2})$, okay? So that is the orthogonal projection, okay? So you can see why that works out. So you can check for instance, you know, (x,y), right? minus $(\frac{x+y}{2}, \frac{x+y}{2})$, right? What is this? This would be minus. Did I get that right? $\frac{x-y}{2}$ and it would be $\frac{y-x}{2}$, isn't it? Right? So this is (x,y) and the point, projection, okay? So v minus, this is v, minus, this is $P_U(v)$, okay? And U is what? The $x = y$ line. So this should be perpendicular to (x,x), isn't it, for all x, right? So this is a point in U, okay? This is a point in $U^{\perp}$, okay? You can check all these things. So these are nice to see. There is of course a by 2 here which is a typo, please note it down. So multiplication I forgot the square root, okay? So maybe this mistake will come in the next example also, we will see, okay? Anyway don't be so alarmed by these mistakes, I mean everybody makes these kinds of minor typos and it's good to make them also sometimes. As long as you find them, you're good, okay? Of course in the quiz if you make them, you lose marks, but that's okay, it's part of the process.

(Refer Slide Time: 28:40)



Okay. So let's see here. This is the general thing. And once again I have done it for (x, y). This is (x, y) and, okay, so all these mistakes still carry over. So let me at least erase this so that it looks quite okay. So now I am looking at the line $ax + by = 0$. I am writing $ax + by = 0$ as the span of $(b, -a)$, okay? Is that correct? $(b, -a)$ belongs to this line, right? If you put $ab + b(-a)$, you get 0, okay? So $(b, -a)$ belongs to this line and I have to make it orthonormal, so I will write it as $span\{\left(\frac{b}{\sqrt{(a^2+b^2)}}, -\frac{a}{\sqrt{(a^2+b^2)}}\right)\}$, okay? So when you do now P_U, I have to take a dot product of (x, y) with this guy. So I will get $\frac{bx-ay}{\sqrt{(a^2+b^2)}} b$ by this. So the square root will go away once again which, again I made a mistake here by keeping, okay? And you will get the projection, okay? So if you want a general formula for projection of a point to a line in $\mathbb{R}^2$, here is your general formula, okay? So if you are the type of person who likes to mug up formulae, so you get this general formula here, okay? So, but it is easy to derive, right? So once you find the spanning vector, find the orthonormal basis for the subspace, you can do it. So one dimension is very, very easy, that's the point I want to convey here, okay? Hopefully you got that.

(Refer Slide Time: 30:23)



So let us go to slightly bigger problems. Not too many, we'll go to three dimensions and then we apply the same thing. You see everything works out in the same way. For the $x - y$ plane is very easy, maybe we'll look at the slightly more complicated case. And here you have, you know, $x + y + z$ equals this. And it's spanned by this $\frac{(1,-2,1)}{\sqrt{6}}$. I mean, I divide by $\sqrt{6}$ to get once again, you know, orthonormality. And once you do that you take inner product with (x, y, z), you will get... I think it's not $(1, 2, 3)$ anymore, it is (x, y, z) right? I think, I mean any point is good enough. But you can see what I have done here. So the projection of (x, y, z) is simply $\frac{(x - 2y + z)(1,-2,1)}{6}$, okay? I got the by 6 correct, I didn't put a square root, okay, so the 6 comes out, so this would be the point, okay? So that is easy to do. But you see, when you really look at a big problem, right? So when you really look at a big problem, $(1, 2, \ldots, 100)$, okay, length hundred vector and a fifty dimensional subspace of $\mathbb{R}^{100}$, okay, what do you do, right? How do you, I mean, how do you go about doing this? Is there, like, a reasonable way to do it, and is there a simpler, more succinct way to introduce this? Are there connections to applications for problems like this? All of this we'll take up in the next lecture and maybe see one more application of this orthogonal projection in another lecture, okay? Thank you.