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

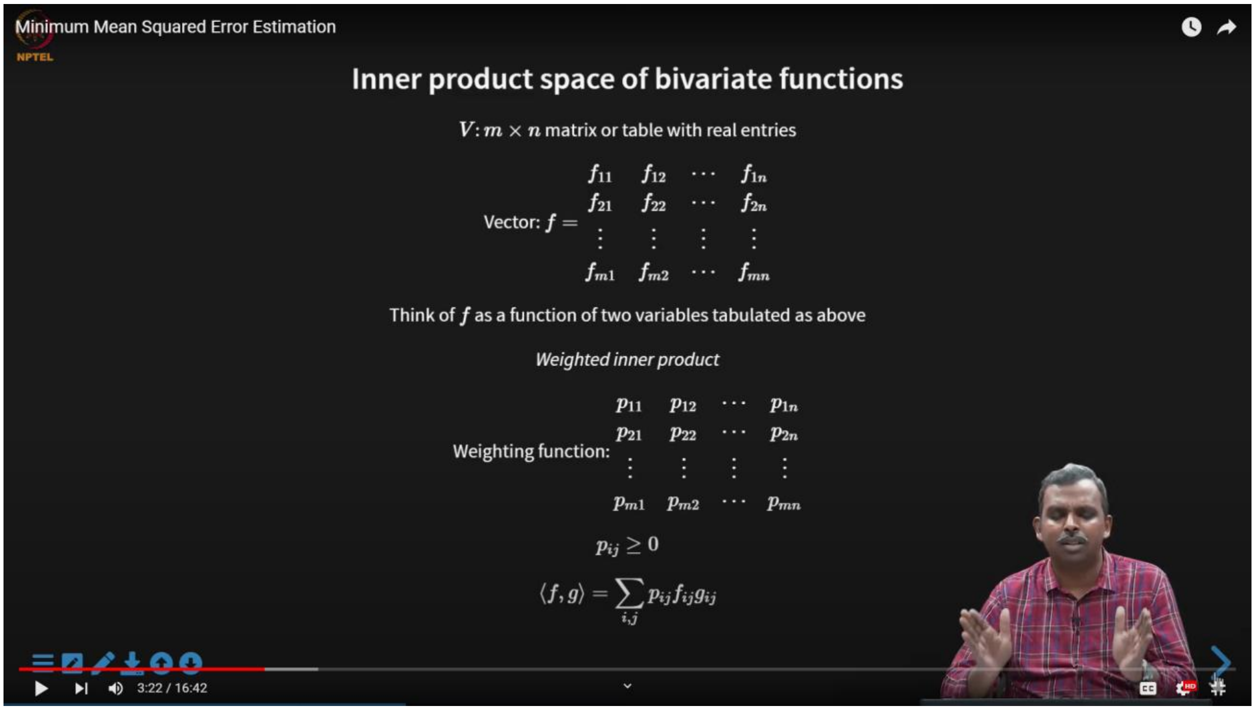
Week 08
Minimum Mean Squared Error Estimation

Hello and welcome. In this lecture we're going to see another application for orthogonal projection. Like I said, I mean, anytime you solve a minimization problem there's lots of applications and this time the application is called Minimum Mean Squared Error estimation. Now this is a big area by itself, there are books that are written on this topic. But at the essence, essentially the idea is around orthogonal projection and Linear Algebra, okay? So let us see how the setting goes. I will consider a particularly simple setting. There are more complicated settings in which you can do the same thing, okay?

So this is a quick recap. Let me not go into great detail with the recap. Okay. So we are going to consider a particular inner product space for this problem. This will be the inner product space of bivariate functions. What is bivariate? There are two variables x and y. And I am going to look at functions over those two variables. What can I do when I think of functions, when I have two variables? When I have one variable, I can write the function in one line. When I have two variables, I have to write it as a table, okay? So now my function I am going to write as a table. But usually when you think of two variables, you are thinking the two variables as taking real values. Infinite number of values for one, infinite number of values for the other. We are not going to do that in this example. It is a simple example. We are going to assume our variable takes finitely many values, okay? It will take finitely many values in one variable and finitely many values in the second variable, okay? In fact, and if you look at my example, I have said one variable takes only m possible values another variable takes only n possible values. So I can now tabulate my function as an $m \times n$ matrix or table, right? So I can put f_{11}, f_{12}, ... So f_{11} is basically that function evaluated at (x_1, y_1), f_{12} is function evaluated at (x_1, y_2), okay? There are only n different values of y, m different values of x. It's a simple sort of function, okay? So you can tabulate it and look at it as a function. All right? So this is the vector space and clearly this is a vector space, no? I can have any number of functions and functions can be added. I will get another function. Or you can multiply the function with a scalar, I will get another function. All of this can be quickly written. It's like the matrix space in some sense, okay?

So there is an inner product that we will define on this space, okay? My inner product is going to be slightly different. Normally so far we have not looked at these weighted inner products too much, right? We have always looked at, you know, weight being one. So here I will look at a weighted inner product, okay? So I will have a weighting function which will again be a table, okay? p_{ij}. And I'll keep the weighting function as non-negative so that my, you know, inner product is valid. I will define my inner product of two functions f and g, in this fashion, okay? $\sum_{i,j}(p_{ij}f_{ij}g_{ij})$, okay? So the weights enter the picture and the values of the two functions, okay? So very simple inner product space. I am visualizing it as the inner product space of bivariate functions. If you want, you can think of it as just matrices and the weighting function is something that's there, okay? So just the small added complexity.

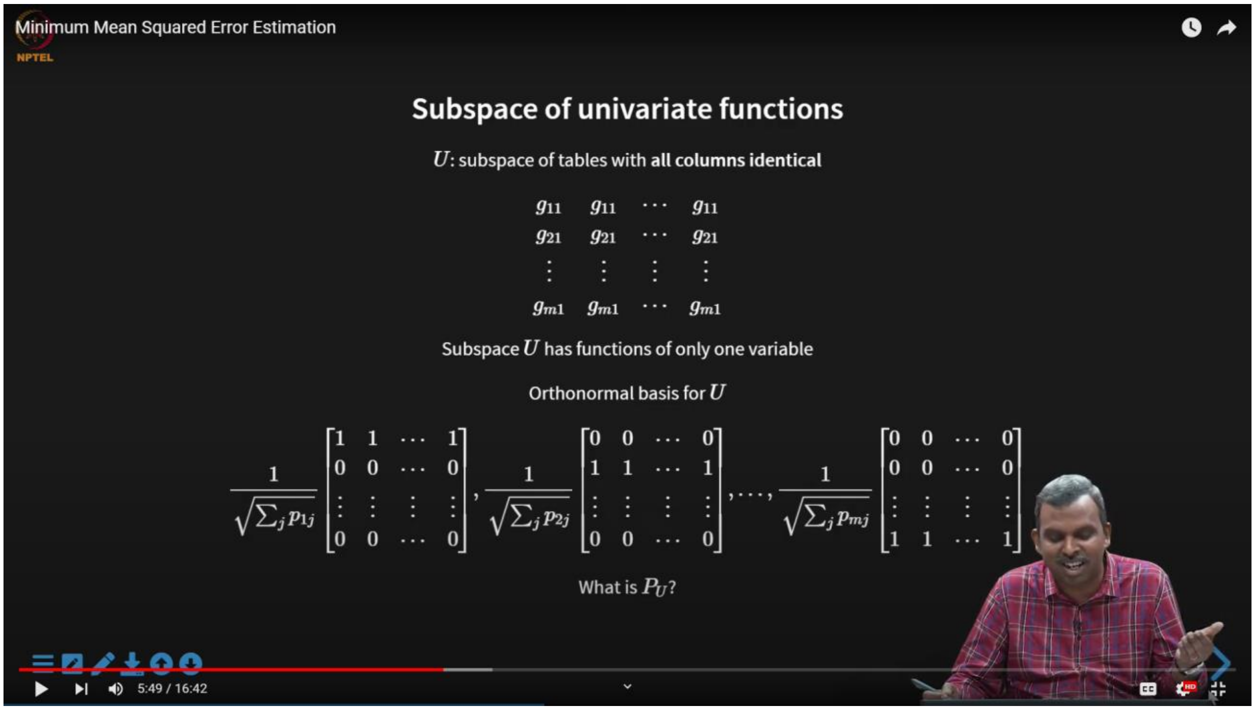
(Refer Slide Time: 03:22)



What's interesting is the subspace that I'm going to choose for projection, okay? So I'm going to choose the subspace of univariate functions, okay? So I have a function of two variables. Supposing now I look at a function which depends only on x, okay? It does not depend on y at all, okay? I can still think of it as a two variable function if I want, okay? So for every value of y, what will I do? I will simply repeat the same x, okay? So that way this univariate function, function of only x becomes a subspace of the vector space of bivariate functions, okay? So here is an example for you to think of, okay? You just make all the columns identical, okay, then you have a univariate function subspace of this bivariate function space, okay? So these are all just terminologies. I am just using them for you to visualize what's happening, okay? So this is definitely a valid subspace of the previous vector space that I defined. There's nothing wrong in this, okay? So this is the univariate subspace, okay?

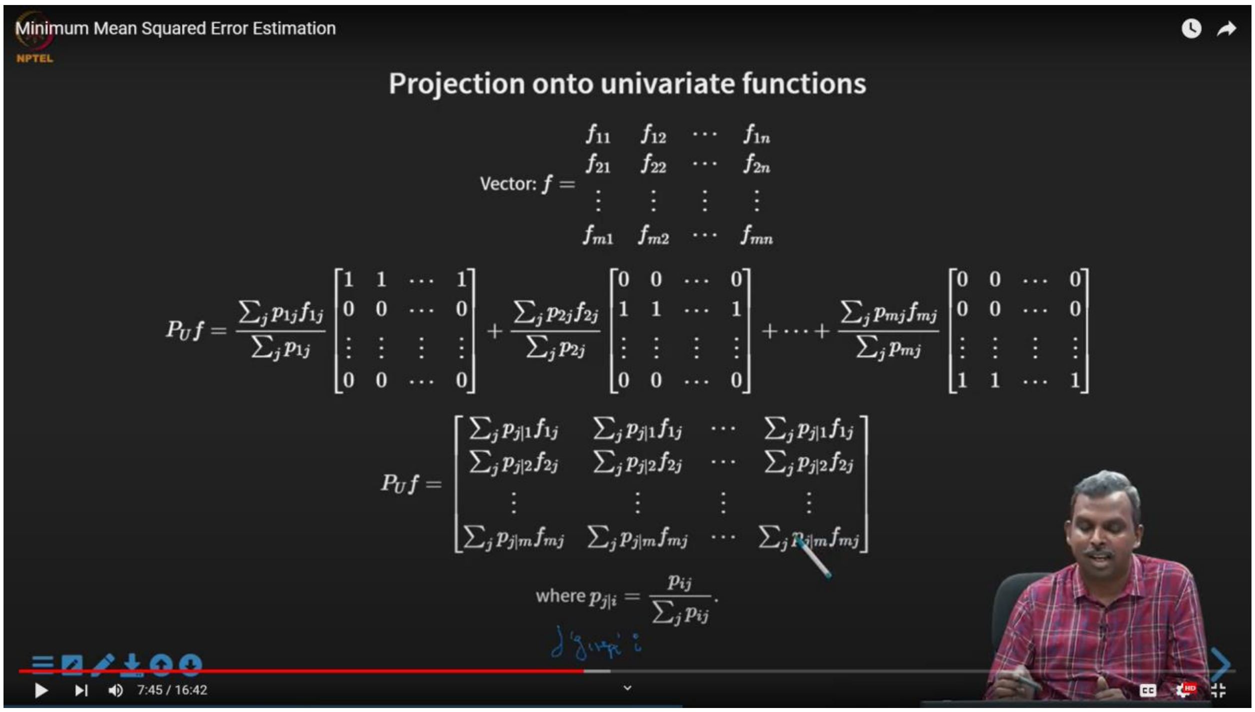
Do we have an orthonormal basis for this? Right? Remember the inner product. Inner product had weighting, okay? So I am looking at a weighted inner product here. There is that p_{ij} which multiplies. So you have to be mindful of that. Even so, you can quickly see that you can come up with a very nice orthonormal basis for this subspace U, okay? So I have given you one here. I mean this is not the only one. Maybe there are others, I don't know. But this is at least a valid orthonormal basis, you can see I have kept $(1, 0, 0, 0)$ $(0, 1, 0, 0)$, I've just kept everything the same and I have normalized, right? So you notice the normalization involves a summation over j of $p_{1j}, p_{2j}, \dots, p_{mj}$, ..., you can check that it works out all correctly. Clearly the inner product between any two of these vectors is zero. So it's an easy exercise to check that this is a valid orthonormal basis, okay? So you have a subspace of univariate functions and that has an orthonormal basis, okay? So it's all easy stuff that we are doing, it's not anything very complicated.

(Refer Slide Time: 05:49)



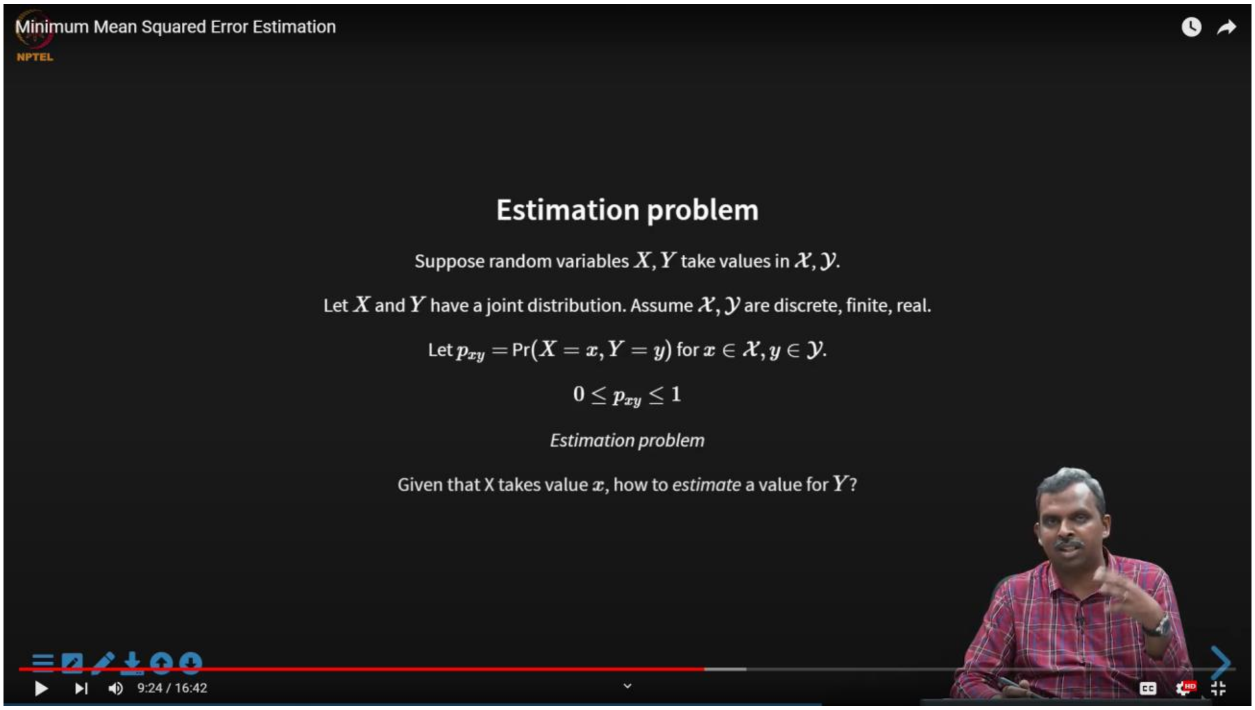
So now we can ask this question: what is this projection on to this subspace, right? You have a subspace, you have an orthonormal basis, of course you have to think of projection next. What is projection? That is also easy enough to do, okay? So you take an arbitrary bivariate function f and then you want to look at the projection, you simply apply the standard formula you know, right? You take the inner product, okay? So that inner product with each of your basis vectors is going to multiply that basis vector itself and then you add, right? So it's very easy to do. And you can do the mechanics of it, I'm not going to provide too much detail. You will see it will come out like this, okay? So you will get a $\sum_j p_{ij}$ in the denominator. In the numerator you will get $\sum_j p_{ij} f_{ij}$ like that it will come, okay? So you will get all of these things and you will get P_U. You can collect it together. And when you collect it together you can write it in this sort of interesting little way. I am going to define this $p_{j|i}$, okay? So this is always read as p j given i, okay? So it's a common terminology arising from probability and all that. So you can define that as, say for instance p 1 given j or p j given 1, I am sorry, $p_{j|1}$ is $p_{1j}/\sum_j p_{1j}$. So this fraction, you know, you take this summation inside this bigger summation and whatever the fraction you get there I am going to call as p j given i. Now if you notice, $p_{j|1}$ also has a lot of interesting properties. So that is why we have defined it like that. Once you define it like that, this $P_U(f)$ I can sort of add up everything together, you will get $\sum_j p_{j|1} f_{1j}$ like that for the first one. And the second one will be $\sum_j p_{j|2} f_{2j}$. The last one will be $\sum_j p_{j|m} f_{mj}$, okay? So this is a valid projection, okay? So you got a univariate function after projecting from a bivariate function, okay? So that is easy enough to see. And this $p_{j|i}$ is this interesting little definition.

(Refer Slide Time: 07:45)



Okay. We have this vector space. We have this weighted inner product, we have the subspace of univariate functions. What can we do with it? Turns out you can do, you can solve estimation problems, okay? So this I'm going to assume a little bit of knowledge of probability. Even otherwise I think you can just look at the previous example. If you don't have too much of a background in probability, you can ignore this estimation part of it. But the previous example was good enough, was clear enough without any additional assumptions. So now let's say you have two random variables X, Y. They take values in discrete finite real sets $\mathcal{X}$ and $\mathcal{Y}$, okay? So this is an assumption you can make. It's reasonable, we'll say X and Y have some joint distribution, okay? So they are jointly distributed in some way. Y has some information about X, X has some information about Y. So you have a joint distribution and that probability is p_{xy}. p_{xy} is the probability that X takes the value x and Y takes the value y. So this is a joint distribution. For all that, we know that the joint distribution probability is between 0 and 1, it is non-negative for sure, okay?

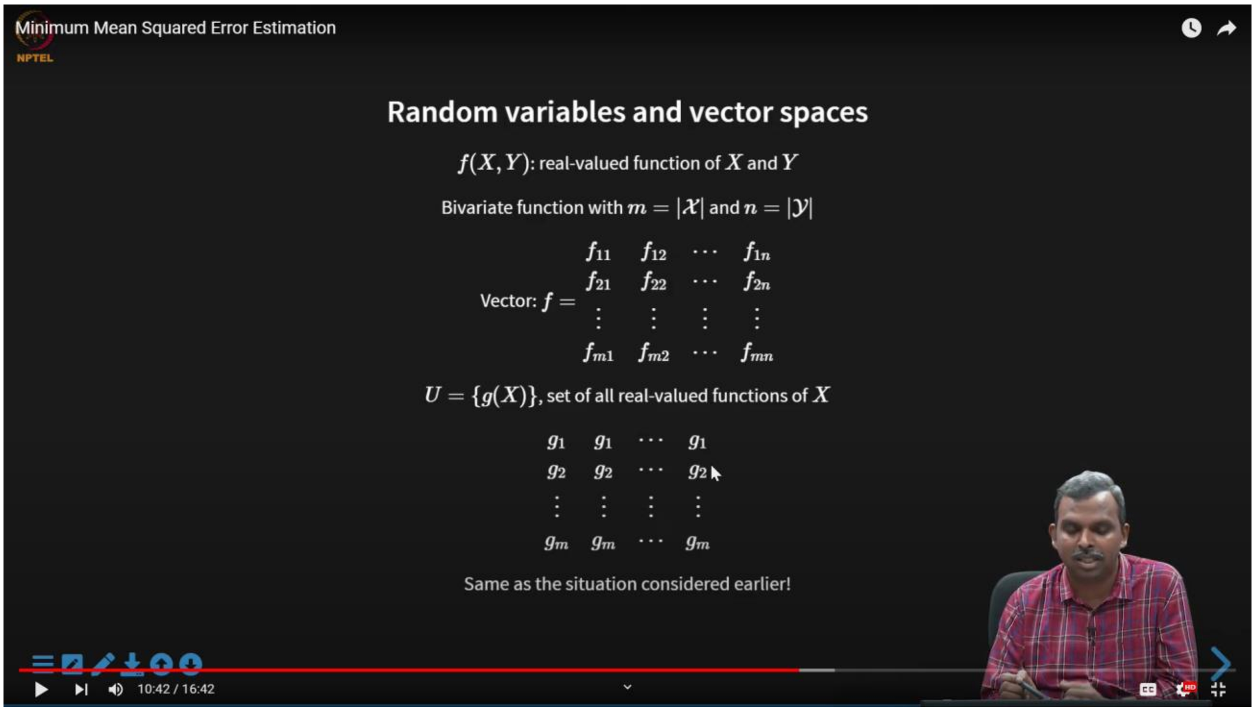
(Refer Slide Time: 09:24)



So the estimation problem asks this question. Supposing I observe X, okay? X takes the value x. How do you estimate a value for Y, okay? So this is at the heart of the estimation problem. This can also be thought of as a learning problem, but usually estimation is in a probabilistic setting, fixed probabilistic setting, not a real life setting where you have to, you know, plug in the probability model and figure out those things which is what you would do in learning. So here in estimation, you are already usually given a probabilistic setting and you have to figure out how to estimate a value for y. So it turns out this problem has a nice connection with what we have been studying so far, okay? So that is what I want to, hope to show you, okay? So now these random variables have connections to the bivariate function vector space that we defined before, okay? So it is very easy to see the connection.

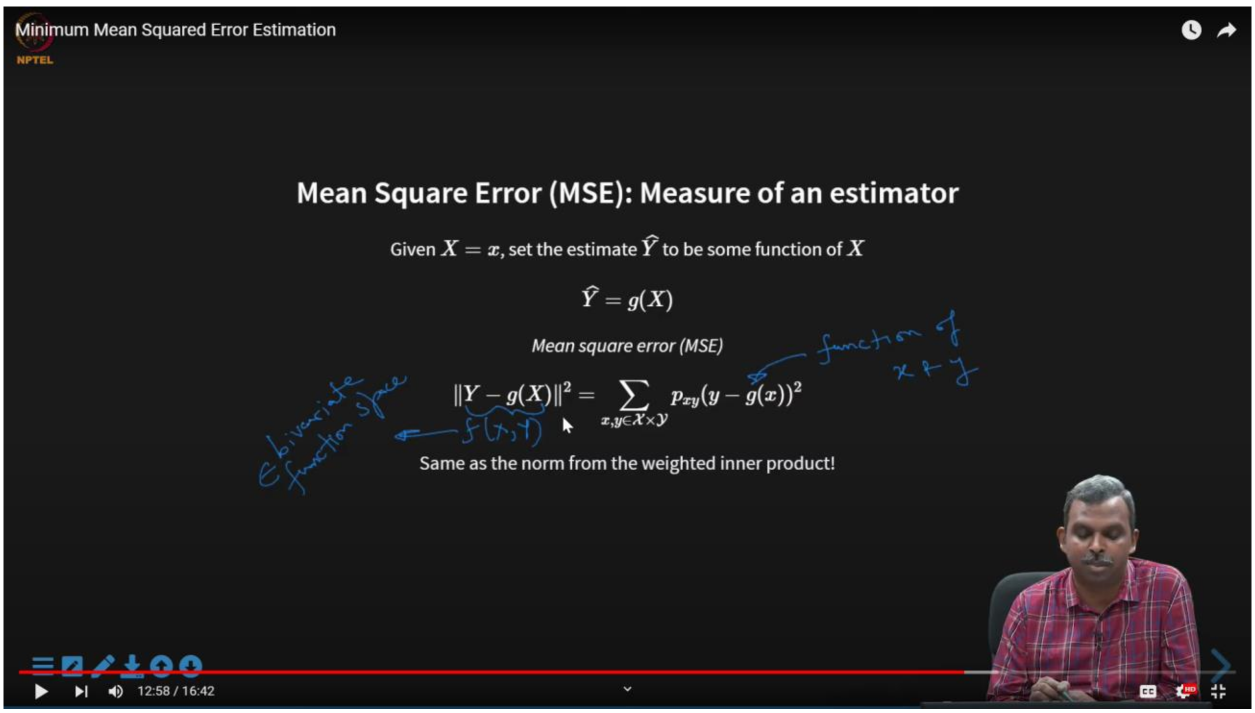
Let us take, let's suppose I take some function of the two real valued random variables $f(X, Y)$. It is actually a bivariate function, isn't it? X takes values in the $\mathcal{X}$. I can think of it as going over the rows. Y takes values in $\mathcal{Y}$. I can think of it as n values going along the columns and each f_{ij} is simply this function evaluated when X becomes x and Y becomes y, right? So that is not too bad to imagine. So we can think of any function of X and Y as being my big bivariate vector, okay, bivariate function vector, that is easy enough to see, okay? And there is also the subspace which is very well defined. Subspace is functions of all, you know, real valued functions of X alone, okay? And this, again this picture of this column repeating comes into the picture. The same thing as before, okay? Now when you want to estimate Y given X, you are observing only X. So your estimate is going to be some form of a function of X, right? So that is the crucial idea which connects everything together, okay? So given that $x \in X$, X takes this value x, we usually set the estimate of Y which is usually called $\hat{Y}$ to be some function of X, okay? So this is the important idea that connects everything together, okay? So you are observing one random variable X. There is another random variable Y which has a joint distribution with this and maybe you understand the joint distribution in some way. You observe X. If you want to estimate Y, the best you can do is come up with some function which, you know, function of X which will give you Y, okay? That is not an unreasonable thing to do. It is very reasonable. So this is what we will do as well.

(Refer Slide Time: 10:44)



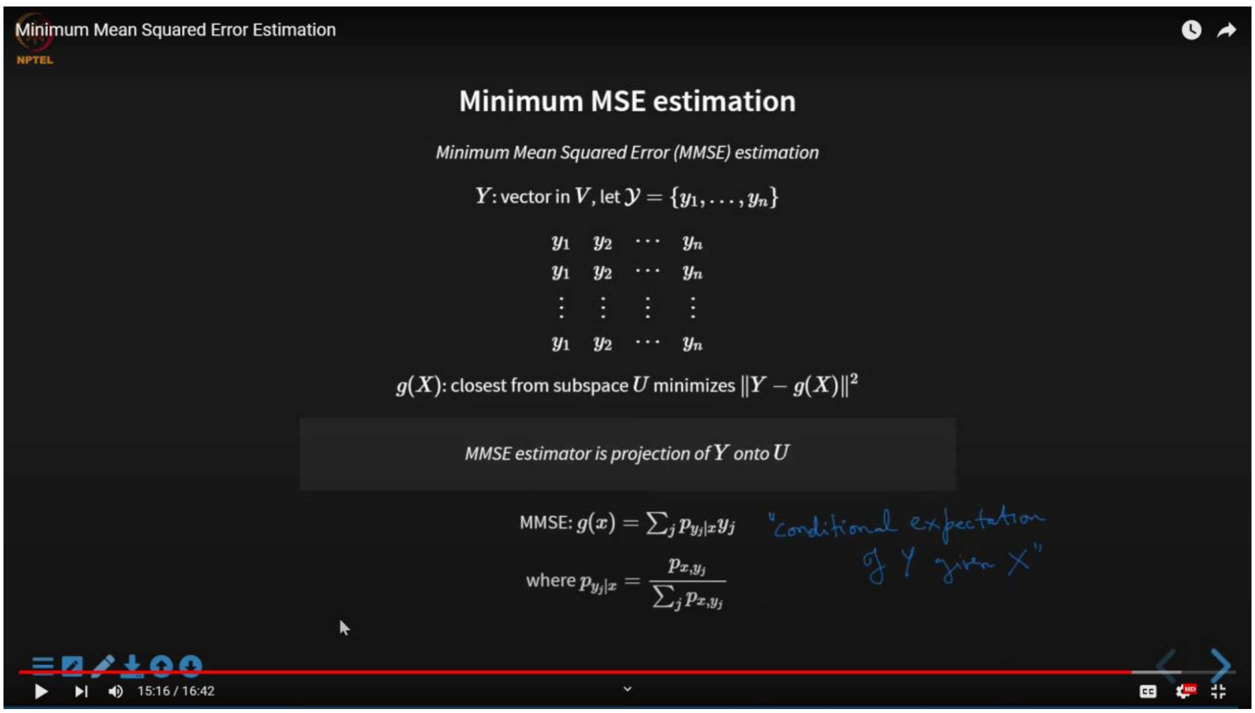
But then how do you know how good an estimator it is, okay? Is it a very good estimator? How do you measure the estimator? You use something called mean square error. This mean square error, it has this definition, okay? Look at this definition, you $\sum_{x,y} p_{xy}$ which is your probability. Non-negative, you know, weighting function so to speak. And $(y - g(x))$, so whatever this $(y - g(x))$ might be, this guy is nothing but, you know, some function of x and y, right? This guy is a function of X and Y, okay? It belongs to bivariate function space that we've been considering so far, right? This simple bivariate function space that we've been considering, okay? So this mean square error which is usually defined as summation $p_{xy}\big(y - g(x)\big)^2$, its, you can see why this is a very good measure of how good an estimator is. You want to weight it by the joint probability and the error you have when that actually happens, isn't it? $\left(y - g(x)\right)^2$. Now that you see is nothing but the norm that we defined in this bivariate function space for this function $(y - g(x))$, okay? So this mean square error which is a very good, you know measure of an estimator is actually the norm from the weighted inner product if you choose the weights to be p_{xy} which are the joint probability functions, okay? So this mean square error has a very nice connection. So if you want a good estimator, it should have low MSE, mean square error or low weighted norm where the weights come from the joint distribution. That's it. So that sort of brings the estimation problem into the Linear Algebra world where you just have a bivariate function space with this weighted inner product and all you're doing is norm-minimization from a subspace, right? $g(x)$ comes from a subspace, isn't it? It's a subspace of all univariate functions of X, right? So again you are doing a norm minimization for estimation, okay? So this norm-minimization will show up in so many different guises and so many different names, it's all the same, okay?

(Refer Slide Time: 13:00)



So what is minimum MSE estimation? It's Minimum Mean Square Error estimation. We know how to do that now. Y is actually a vector in this bivariate function space V. If you let the $\mathcal{Y}$ be $\{y_1, \dots, y_n\}$, it is just simply rows, you know, $[y_1, \dots, y_n; y_1, \dots, y_n]$, it's not even a bivariate function. It's a univariate function but a function of Y, not X, right? That's important to know. $g(X)$ is closest, if you want to do minimum mean square error estimation, it's the closest to Y, okay? The closest vector in U which minimizes $\left||Y - g(X)|\right|^2$, right? So the MMSE estimator, the estimator that is going to minimize your mean square error is simply a projection of Y onto this subspace U, okay? How do we do projections? We know how to do that, right? You have to define this given probability $p_{y_j|x}$ which is p_{xy_j} by summation and $g(x)$ is simply the summation over j this given times y_j, okay? So this is also called, in the parlance of estimation, this is called conditional expectation. Expectation of Y given X, okay? So at least in the simple finite case we are able to quite nicely show that the MMSE estimator is the conditional expectation which is defined in this fashion. In fact the assumptions that I made on, you know, finite domain and all that is not really needed. As long as you can extend, as long as the inner product space is still valid, you can see why this will always be true, isn't it? So it will work out quite decently, okay? So that's minimum MSE estimation.

(Refer Slide Time: 15:17)



I'll close with a couple of, just mentioning couple of applications for estimation. For instance in the lab when you measure a physical quantity, you know, any quantity, it could be, when you measure, you will see that you always take multiple repetitions and then take the mean. What you're doing is actually an estimation, right? Every measurement is not the actual physical quantity, it's the physical quantity plus noise. And when you take the mean of multiple measurements, you're estimating the original physical quantity and eliminating noise in some way. And you're doing an MMSE estimation. So there are fantastic applications of MMSE estimation. Every time you measure something in a lab and other very more interesting, much more interesting applications are things like telecommunications where you transmit a bit and a noisy version is received and the receiver has to estimate the transmitted bit from a noisy version of it that it receives at the receiver, okay? So these are the kinds of problems. Electrical engineering and signal processing are full of such problems of estimation, okay? That's the end of this lecture. Thank you very much.

(Refer Slide Time: 16:29)