INDIAN INSTITUTE OF TECHNOLOGY MADRAS

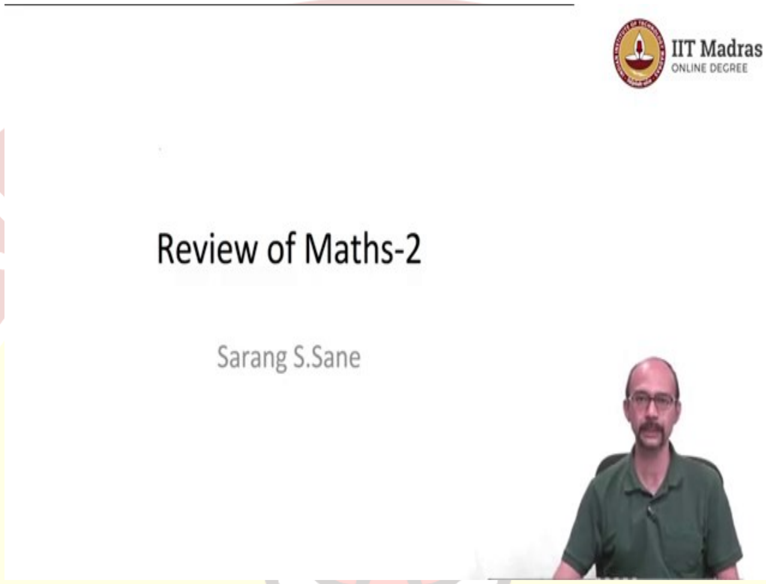
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Sciences - 2
Professor Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras
Review of Maths - 2

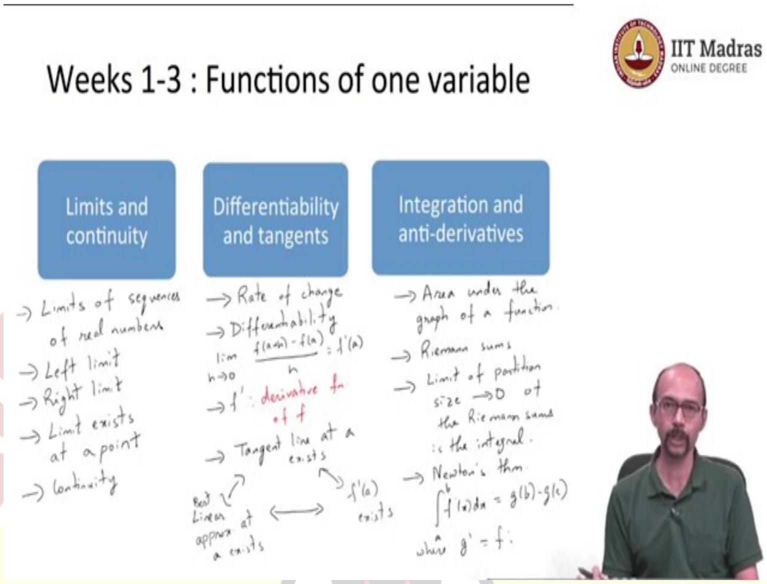
(Refer Slide Time: 0:19)



Hello, and welcome to the Maths 2 component of the online BSc program on Data Science and Programming. This is the final video of this course. So, let us do a review of whatever we have studied in this course, I am going to do a week-by-week review. But I will take bunches of weeks, depending on the topic. So, topics which are sort of clustered together, I will try to address them in one shot.

So, this is a very brief review, it is not meant to exhaust the entire course, in the sense of I would not be able to cover all the ideas or all the details of what we have done in the course. But this is a broad outline of what we have learned. And yeah, it is my hope that if not the details, at least the words and the topics that we have learned are something that you can keep in your mind as you go ahead, especially for the upcoming courses of Machine Learning, and other courses also, where they may not use the same ideas that, the exact same ideas that you have learned in this course, but they will use the maturity that you have gained while studying this course.

(Refer Slide Time: 1:27)



So, let us start with weeks 1 to 3 where we function, where we studied functions of one variable. So, the broad outlines of what we studied were the notions of limit and continuity, differentiability and tangents, and then integration and anti-derivatives, so just as, just to recall within this, what did we do.

So, first, we did limits of real numbers, so limits of real numbers. So, that means limits of sequences of real numbers, maybe I should write limits of sequences of real numbers or limits of sequences in $\mathbb{R}$. So, the idea here was that the sequence comes closer and closer to a point as your sequence increases, as n increases.

So, we use this idea to talk about limits of a function. So, we said, if you have a limit, if you have a sequence, which approaches some real number x, then the function value at the sequence points must approach the function value at that number x, and if that happens for all sequences, which converge to x, then the function is said to be continuous at that point.

So, we made several intermediate steps. So, we had the notions of left limit and right limit, which is when we said, we look at what happens when we come from the left, from the right. Sometimes it happened that both sides there were limits, then we demanded that the limit matches and if they, if it did, left and right matched, then we set the limit as x tends to some point A that exists.

So, limit exists at a point. And then finally, we add continuity, which said that the limit must exist, and that limiting value at the point must match the function value. So, the function evaluated at A is the same as the limit of x tends to A f(x). So, in other words, you can interchange limit and function that was the idea of continuity. So, this was a brief outline of what you did for limits and continuity.

Then we studied the notions of differentiability and tangents. So, we saw that limits. The notion of continuity allows us to talk about the existence of limits, but it does not guarantee that the function behaves well. So, what do we mean by behaves well? That the graph of the function does not have corners. And somehow, we wanted to avoid corners, because wherever we have corners, it becomes difficult to talk about rate of change.

So, the main or underlying theme here was that we want to study the rate of change of a function at a point. So, we saw that this rate of change is captured, provided the function is differentiable at that point. Which meant that as you take $\lim_{h\to 0}\frac{f(a+h)-f(a)}{h}$, this should converge somewhere and whatever that value is. So, it should exist and whatever that value is we call f ‘ a. So, that was called, so f ‘ is a function that is created in this way, so, that is the derivative function.

So, f‘ is the derivative function of f. And we saw that the derivative function is or the idea of differentiability is very, very, very closely related to the idea of the existence of the tangent line. So, the key point here was tangent so the tangent line at a point a exists. So, this was the same as f‘ a exists. And this is the same as the linear approximation, best linear approximation at a exists. This was the idea of tangents and the idea of rate of change, fine.

So, then the third topic sort of that we studied here was integration. And so, the integral computes the area under the graph of a curve, sorry the area under the graph of a function so the way we do this is to take what are called Riemann sums, which means you take rectangles, which sort of cover the part that you want to find the area of and then you sum up the area of those rectangles, and then you make those rectangles, you make your, the length or the height of the rectangle is measured by the function value at some point within the interval.

So, you have to first take a partition of your interval, then for some one of the sub intervals in that partition, there is a point that you choose whose function value you take that is the height of your rectangle, and then the sub interval length is the width of your rectangle, and then you multiply height by width.

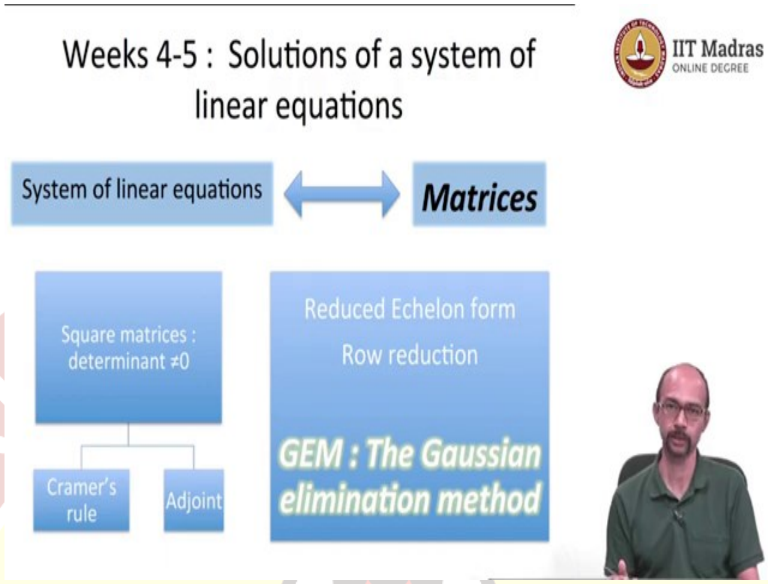
So, that gives you an approximation for the part around the, above that particular sub interval, and then you sum these up. And that is what the Riemann sum is. And then as you make your partition smaller and smaller, so limit over partition size tends to 0 of the Riemann sums is the integral, this was how we defined the integral.

And then we said, well, you can extend this, sometimes you can do 0 to infinity as well, you can do - infinity to infinity for depending on how the function is etc. And then we also saw that this is a very difficult way to compute actually. And that is where Newton's theorem comes in very, very, very handy.

So, Newton's theorem says that if your function is continuous, then first of all the integral exists. And you can find it by finding the anti-derivative. So, that means you find a function, so that its derivative is the function that you have. And then if you can find such a function, so then suppose that function is g, then $\int_a^b f(x)dx = g(b)-g(a)$, so you evaluate it at the endpoints of the interval that you have, and then take the difference of the larger one from the smaller one, meaning the larger endpoint from the smaller endpoint.

So, integral f ‘ x, or $\int_a^b f(x)dx = g(b)-g(a)$ where $g' = f$ this is in a nutshell what Newton's theorem said and this is what allows us to compute our integrals. And you already saw applications of the integrals in this statistics course, you will see more applications of all these things as you go on ahead, fine. So, this was a brief summary of what we did in weeks 1 to 3, where we studied functions of one variable.

(Refer Slide Time: 10:20)



Next, we started looking at a system of linear equations. And then we started asking how do we solve such a thing? So, the main thrust here was that given a system of linear equations, you can find a corresponding matrix. So, you change your question about solving a system of linear equations, instead do a question about matrices. And this allowed us to use the theory of matrices in order to solve these equations. So, this was the main, the big idea in these two weeks was you can rewrite your system of linear equations in the form a $x = b$. And then you can use the matrix a and column vector b, to solve for this.

So, in order to solve of course, we had to define the various things about matrices. So, in particular, we had to define determinants. And once we define determinants for a square matrix, we saw that there were two methods to solve an equation of the form $a x = b$ for a square matrix if the determinant is nonzero.

So, the determinant being nonzero was the same as the matrix being invertible. This was a key point. And the first method was what was called Kramer's rule, which was where you took your column vector B, and you placed it in appropriate columns, you replace the appropriate columns of have a and then you found the determinant of that, and then you divided that by the determinant of a. And that was the i-th, so if you replace the i-th column, then this process gave you the value of x_i, which was a solution. And in this case, there is a unique solution. So, this completely solves the problem of finding the solutions.

The other way was to find a inverse, the inverse matrix. So, to do that, we had to compute what the inverse was. And to do that, there was a formula wherein we define what was called the ad joint. So, to do that, we defined the notions of the minors and so on, then you have to place all these, you have to multiply by a suitable + or - 1, and then you have to place them in appropriate places. And then that divide the whole thing by determinant of a, and that gives you your inverse matrix. And so that that is the solution via ad joints.

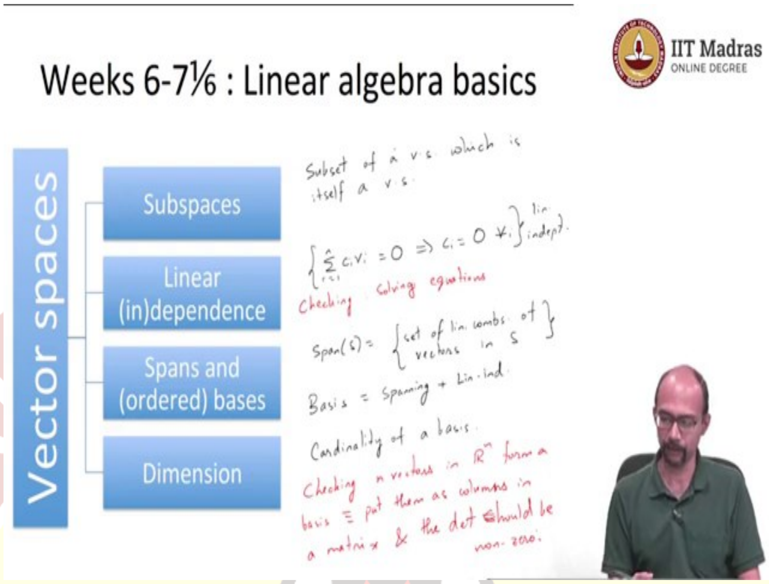
So, these were the two methods we saw for square matrices where the determinant is nonzero. After that, we studied extensively, what was called the reduced epsilon form. And the idea of row reduction for matrices, here the matrices were not square, you could have any kind of matrices.

And the really big point in this study was, at the end, we had something called the Gaussian elimination method, which in short, some people called GEM, because it really is a gem. And the Gaussian elimination method allows you to solve equations by taking a matrix a here, putting your column vector here, and then doing row reduction.

So, this was the method, so where you row reduce a and carry out those same operations on b. And at the end, you get a matrix in reduced epsilon form. And then we know how to write down those solutions. And then those set of solutions are the same for the original system $a\,x = b$, this was the main point.

And the Gaussian elimination method or row reduction was the really outstanding idea that we studied in these weeks. And this idea was later used as well, which we will come to in a minute. So, just to reiterate, we studied matrices over here and we studied row operations. Specifically, how to row reduce and how to use it via Gauss elimination to get solutions of system of linear equations, fine.

(Refer Slide Time: 14:43)



So, after this, we had weeks 6 and one video in week 7, which dealt with the basics of linear algebra. So, here we studied for the first time what is an abstract vector space, for a vector space we studied what are subspaces? So, just to recall, a subspace is a subset of the vector space V of a vector space, which is itself a vector space. What does that mean?

That means it is closed under the operations of scalar multiplication and addition. So, once that happens, it will become a vector space in its own, because all the other properties are already derived from the fact that it is a subset of this vector space. So, this was one of the things we studied, we studied the notion of linear dependence and independence of vectors. So, this was where you took a bunch of vectors. You said, well, let us see if I can solve this equation.

So, if you can solve this equation in a non-trivial way, then it is they are linearly dependent. And if it is the only solution is the trivial solution, meaning all the constants CIR 0, then this is linearly independent. So, this is linearly independent, this is what we mean by linearly independent.

And one of the ideas behind linear independence is that linear independence allows you to write any vector in the span of the vectors that you have in a very easy form or in a sort of the most compact form. So, of course, before that, we had to talk about spans. So, span means the collection of those vectors that you can obtain from the vectors that are given to you by taking all possible linear combinations.

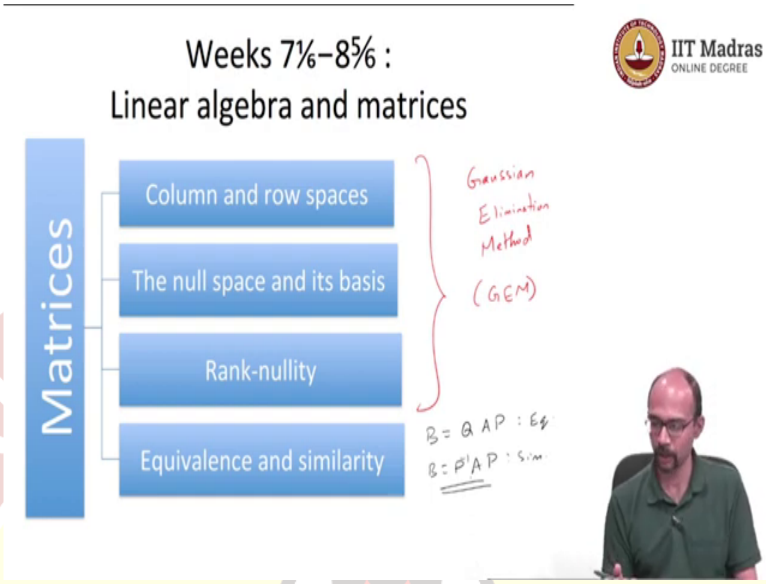
So, span of a set S is set of linear combinations of vectors in S. And then we saw that if you have a set S such that it is both spanning and linearly independent, then that is called a basis. So, basis means spanning + linearly independent, and then we had various ways of characterizing this. Namely, it is a maximal linearly independent set or a minimal spanning set, etc and the number of elements, the number of vectors that you need in order to form a basis that is exactly dimension.

So, dimension is the cardinality of basis. And this is well defined, because any two basis have the same cardinality. And we used all these ideas that we had studied for matrices in order to make computations for this. So, we saw how to if you have a spanning set, how to throw away vectors in order to get a basis, we saw if you have a linearly independent set, how to take in more vectors in order to make it a basis, and in general, how to see whether a set of vectors forms a basis or not. So, all of these were things that we did using matrices.

So, maybe I should say here that so checking for linear independence means so checking involves solving equations, so we are solving this equation, solving equations. Checking something is a basis means, again, we have to solve a bunch of equations or you can, if you, if it is in $\mathbb{R}^n$, then you must have n vectors first of all, if it is more than n, then it cannot be a basis. Because they will not be linearly independent. If it is less than n, then they will not be a basis because they cannot be spanning. So, it must be n.

And then we saw that checking n vectors in $\mathbb{R}^n$ form a basis is the same as them as columns in a matrix and the determinant should be nonzero, should be nonzero. So, if it is nonzero then we know they form a basis, if it is 0 then they do not form a basis. So, our previous theory of matrices, all the operations that we did on matrices helped us with solving these kinds of problems.

(Refer Slide Time: 20:37)



Then we did a little bit more of linear algebra, where we connected linear algebra with matrices in the sense that first of all, we studied matrices by themselves. So, we studied the notion of column spaces and row spaces for a matrix, so that is the span of the columns of the matrix or the span of the rows of the matrix, so they will be if the matrix has size m $\times$ n, so that means it has m rows. So, each row is an element of $\mathbb{R}^n$. And so, the row space will be a sub space of $\mathbb{R}^n$, and the column space will be a subspace of $\mathbb{R}^n$.

And so, then we studied these column and row spaces, and we saw how to get a basis for the column space and the row space again, the underlying idea was Gaussian elimination because you have a spanning set already, so, you can use Gaussian elimination to throw out some vectors, some of these columns or rows, and then you can get a basis for the column space and the row space.

So, similarly, we define something called the null space of a matrix. So, the null space means all those vectors x such that $a\,x = 0$. So, again this means solving equations, and we know how to do this using Gaussian elimination. So, using Gaussian elimination, we could tackle the problem of the null space. And in fact, the Gaussian elimination method also gave us what are the basis vectors for the null space.

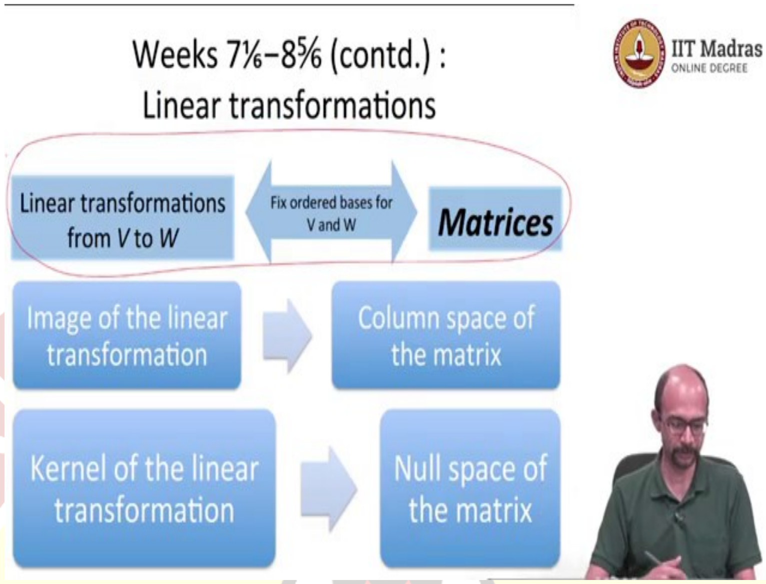
And putting these two together, we could find the rank of the matrix which is the dimension of the column space or the row space, which happened to be the same and the nullity of the matrix which is the dimension of the null space. And it also showed us that something called the rank nullity theorem was true, which meant that rank + nullity = n, where this matrix is of size $m \times n$.

So, all of these are the checking part is a byproduct of Gaussian elimination. So, this is a really powerful method as I commented earlier, which is why it is often called GEM. And finally, we studied the notion of equivalence and similarity of matrices. So, this came towards the end of the 8th week after we had also studied linear transformations, which I will do in the next slide. And after which it made me a little bit more sense.

So, the idea of equivalence is not so difficult to understand. So, $B = Q^{-1}AP$ or Q A P, so this is equivalence. Similarity is a much, much deeper idea. So, the idea of equivalence is just to say that when do two matrices have the same rank, once two matrices have the same rank, they are going to be equivalent. So, it is a precise condition. And where do we need this?

Well, if you change your basis on both sides, then you change your matrix that is coming to linear transformations. And that is why this is defined in the first place. Similarity, though, is much, much more involved idea, because here what we are doing is we are changing our basis simultaneously on both sides. And that is why we had similarity and the idea of similarity is going to come in again, it will come when you study machine learning and foundations and you will study things like Eigen values and characteristic polynomials and so on. So, although we had only one video on this, this idea is going to come again. So, I would encourage you to pay attention to this, do it once more before your ML course if you want.

(Refer Slide Time: 24:58)



So, the same set of weeks. We also studied linear transformations. So, linear transformations essentially functions from vector spaces to vector spaces. Well, of course, we know, the idea of functions already. If you have two sets, a function takes something in here and says I will associate to this element of the set, I will associate some element of that set. This is what a function does.

But when you have extra structure on those sets, then you will often like to preserve that structure. So, here for vector spaces, what is the main structure that you have? It is coming from addition and scalar multiplication. So, you would like to preserve that structure and linear transformations are exactly what do that.

So, linear transformations are defined by the fact that if you have a linear combination, and you apply the linear transformation, then you get a linear combination on the other side, where you can take your transformation inside the linear combination. So, linear transformation means when you apply your function to a linear combination, you can take your function inside the linear combination. This is exactly what is a linear transformation.

Now, again, the punch line in this entire set of videos was, once you fix a ordered basis for V and W, you fix an ordered and remember here it has to be an ordered basis, not just a basis, yeah, you fix a set of vectors $V_1, V_2, \ldots V_n$, and $W_1, W_2, \ldots W_m$, then you get a matrix. How do I get a matrix, if I express f of one of the basis vectors of W in terms of the basis vectors for V, so I have V to W sorry, so if I take a basis vector for V, and apply the transformation and express that in terms of the basis vectors of W, the ordered basis of W that we have chosen.

So, those coefficients are what come into your matrix, that gives you a matrix and essentially, this entire thing said, well, I have this linear transformation, but what I am going to do is I am going to choose ordered bases on both sides convert it to a matrix, then whatever I want to find about that linear transformation, I am going to use my matrix, which we developed a lot of theory for in order to find it. So, what are some of the things that we could find, the image of the linear transformation, this goes to the column space of the matrix, this corresponds to the column space of the matrix.

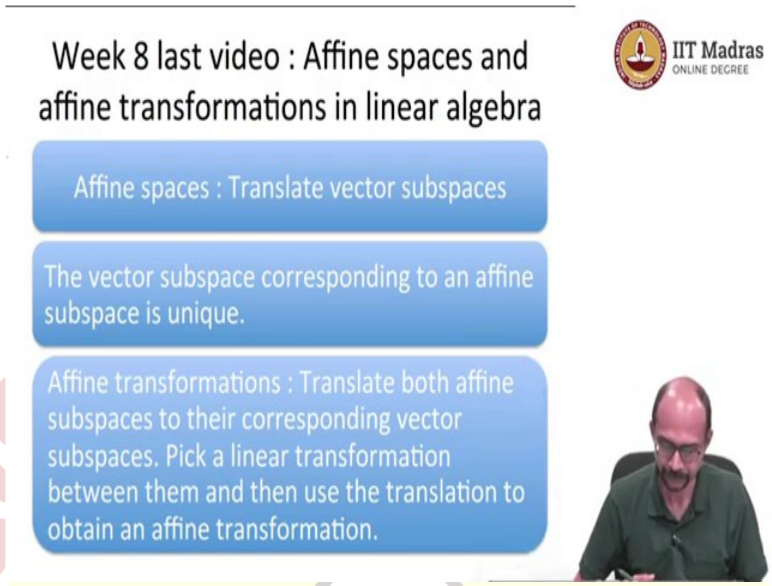
So, in particular, if I want the dimension of the image of the linear transformation, then so sometimes that is called the rank of the linear transformation, then that is exactly the dimension of the column space or the row space of the matrix, and that can be found by finding the rank of the matrix, which we have seen before can be done using Gaussian elimination.

Similarly, we define something called the kernel of the linear transformation that is exactly what corresponds to the null space of the matrix. So, again, you find, you can find the nullity or a basis for the null space using Gaussian elimination and then use that to get the basis or the dimension of the kernel of the linear transformation. So, that was again called nullity of the linear transformation.

And then because we had this very nice formula for matrices that rank + nullity is n, you get the same thing for linear transformations, this was broadly what we did. So, the key point here is this thing of how we did this and I think this is a fairly if you are seeing it for the first time this is a fairly involved process, it is not difficult but you have to keep it in mind that what are you doing, so how do you get your matrix, this is what you have to keep in mind.

So, express the basis vector of V in terms of the apply your linear transformation and express that in terms of the basis vectors for W and those coefficients are what you should take and that put them in the appropriate places in your matrix. This is exactly how you get this matrix. Yeah. So, a large part of linear algebra as you have seen so far is to convert your problems from abstract vector spaces into a tangible thing that we understand like matrices. This is exactly what linear algebra does for you.

(Refer Slide Time: 29:41)



And then finally, we had one last video in week 8, which is called, which was on affine spaces and affine transformations in linear algebra. Now, I want to warn you, some of you might have studied mathematics before and you may have come across the term affine space. It is used in different context. So, this is the context in linear algebra, there is a different context in higher if you have done more mathematics, and that is not what we are talking about in this set of videos, fine.

So, what are affine spaces, they are very easy, you take now vector subspaces, you have a vector space, then you take vector subspaces, and you translate them. See, remember that vector subspace means it must pass through 0 when you take it as a geometrical object, like a plane or a hyperplane, or something like that, it must pass through the 0 vector.

So, if you translate it, it will not pass through the 0 vector. But that still has some nice structure. After all, it is still a plane, or it is still a hyperplane, although it is not passing through 0. So, you would want, you want to study such things. And indeed, later on, we saw why we want to study such things, because we studied the notion of tangent planes, or tangent hyperplane. So, those are affine spaces.

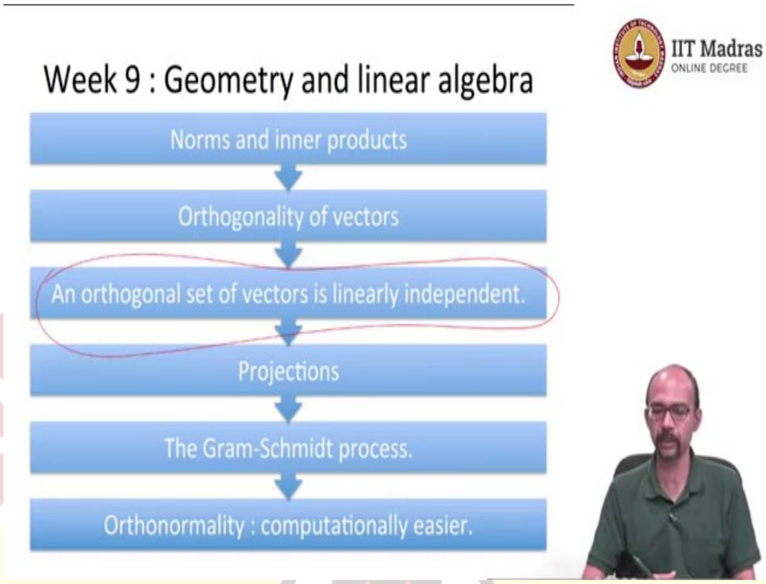
So, that is one of the reasons to study affine spaces. So, they are just translates of subspaces, nothing really at all deep happening in this, only a lot of notation, so you have to get used to it. So, as I said, one of the good reasons for studying this course is to develop your mathematical maturity, that is really what, many of these things are for that reason, fine.

So, the vector space corresponding to an affine subspace is unique. So, this is a, it is very intuitive, because you have translated it, if you translate it back to 0 parallelly, then you get something unique. So, that is a unique vector of space that you get. And then how, what are affine transformations?

Again, this was a lot of notation when we defined this, but what they are really doing is very simple, you translate your, you have two affine subspaces, translate this to pass through the origin, that translation means already you have a function. So, you map some vector your to the vector on this thing. Similarly, you have a affine subspace here, you transfer it to the origin, meaning passing through 0. So, now you have two vector subspaces.

Here, you can talk about linear transformation. So, you have a linear transformation here. So, what you are doing is you are saying affine transformation means what, you are first asking for this map, then you are asking for this linear transformation, and then you are asking for this translation, natural, once you put these three together, that is exactly what is an affine transformation. So, writing it down in terms of algebra is a little tedious, but the idea is not at all difficult, it is exactly what is happening, translate, taken in your transformation, translate back that is all.

(Refer Slide Time: 32:54)



Then week nine was very studied a little bit of geometry allied with linear algebra. So, linear algebra, as we saw is mainly about, so the word linear has something to do with it, so it talks about linear combinations and things like that. So, now we also want to do geometry, because after all, much of this is happening in $\mathbb{R}^n$, $\mathbb{R}^n$ has many good property. So, if you have two lines, you can talk about the angle between them and so on.

So, how do you capture that in terms of linear algebra? And to do that, we talked about norms and inner products. So, norms were the general version of lengths and inner products was the general version of dot products or scalar products. And just to remind you, we saw norms and inner products coming in later, the inner products came for example, when we talked about directional derivative, so the directional derivative was the inner product of the gradient vector with the unit vector in whatever direction you are, you want to find the directional derivative.

And the norm came about when we talked about tangents and the best linear approximation because, after all, when we start talking about best linear approximation, in many dimensions, we have to talk about how close it is. And once we say how close that means, we need to know some notion of distance or length. So, that is where norms came in.

So, of course, we did not talk about gender norms when we went ahead, we talked only about the dot products and the usual distance metric, but again, you can think about maybe in some world, there may be a different way of computing distance and there, you may want to do all of this but with a different notion of distance.

And I need such things do occur, for example, in physics and so on. Anyway, so that is why we studied norms and inner products, so the main emphasis even if you may not have understood, general norms or general inner products is on the dot product and the length that at least you must know very well.

So, once we talked about the norms and inner products, we talked about orthogonality of vectors. So, essentially, remember that the dot product picks up the angle between vectors, that is one of the things that the dot product does. And orthogonality is saying, when are two vectors perpendicular to each other, at least for the usual inner product, that is what it means that they are perpendicular to each other.

So, sometimes instead of perpendicular, we say orthogonal. That is the standard language when we do linear algebra. So, if you have an orthogonal set of vectors, then it is linearly independent. This was really the punch line of whatever we did. And the point here is that this allows you, this idea of inner products allows you to easily find nice basis, this is one of the main reasons why we study orthogonality. Or why we would like that things are orthogonal, fine.

And the way we did this was to use projections. So, projections is like shadows. This is a very easy idea. If you have something like this, you want to project it down, depending on where your light source is, it will cast a shadow. And that is exactly what the projection did. So, the projection we wrote down concretely as a linear transformation, what it was, so it projects a vector from its original vector to some subspace, this is what a projection does.

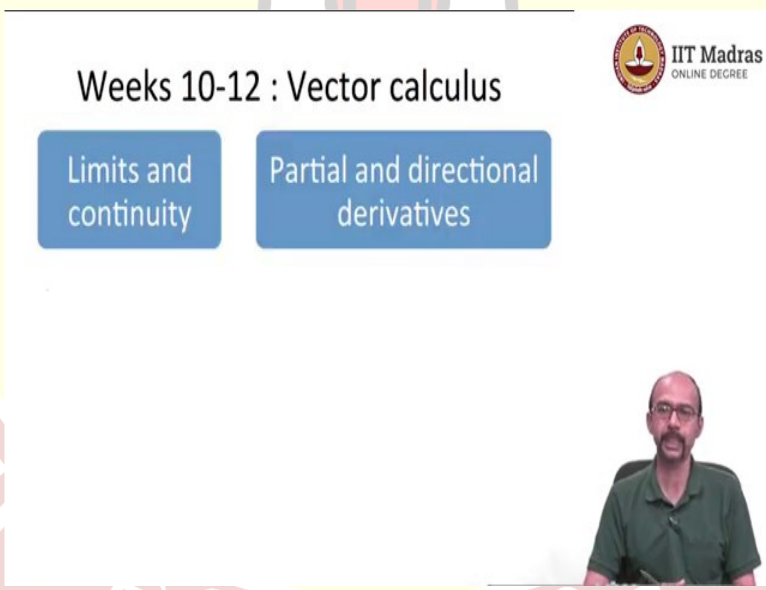
And if you have an orthogonal set of vectors, then you can use these projections in a very nice way. It is a very beautiful, clear cut formula that you can use. In order to make it into an, if you have a general set of vectors, in order to get a basis, so we have a spanning set, you can create a basis out of that, by using the Gram Schmidt process.

So, the Gram Schmidt process sequentially created vectors wherein it said well, what I will keep doing is I will take projections and subtract those out. So, whatever extra things remaining, that is not already there in my previous subspace. So, I will take that in. And that will actually be orthogonal to whatever I already have. And that way it created a orthogonal, in fact, an orthonormal basis.

And the reason we prefer orthogonality is because it makes computations easier, so you do not have to keep dividing by the length of the vectors in all. So, it is computationally easier process, I should have, yeah, since we have come to computationally easy process I should also mentioned from the previous slide on matrices, that Gram Schmidt sorry, that the Gaussian elimination method was computationally the fastest method amongst the various methods to solve equations, fine. So, this is a overview of what we did in week 9.

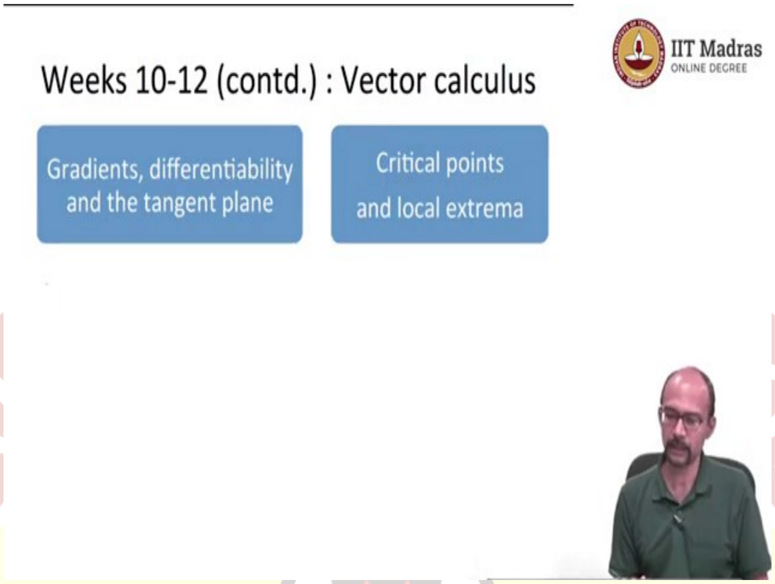
Yeah, maybe I should also say, in the middle, we had a break of one week, so I am just doing this according to the weeks in which we had videos coming in. So, you have to delete that week when we had a break, otherwise, some weeks will get shifted by one.

(Refer Slide Time: 38:46)



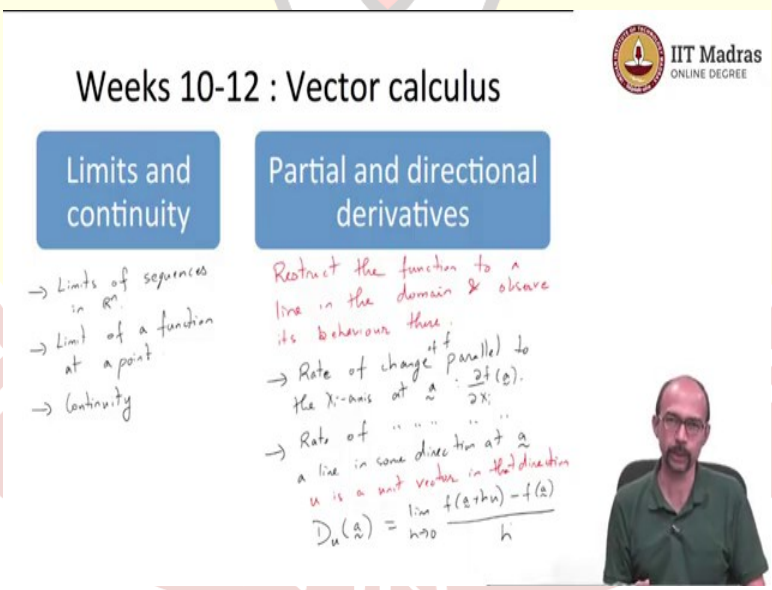
And then finally, in weeks 10 to 12 we studied vector calculus, this is exactly where the one variable calculus that we studied in the first three weeks, at least the first two topics, and the linear algebra that we studied in the subsequent weeks came together and we could study functions of several variables. So, the main themes here were limits and continuity again, from similar to one variable calculus, then partial and directional derivatives.

(Refer Slide Time: 39:15)



And then gradients, differentiability and the tangent plane and then critical points and local extrema.

(Refer Slide Time: 39:23)



So, let us, let me quickly recall, what we did about limits and continuity. So, the notion of a limit in several variables is slightly deeper than the notion of a limit in one variable. Because in one variable essentially, there are only two directions in which you can come in, there is nothing more that can happen.

Whereas in many variables, there are many different possible ways you can come, not just along straight lines, but even along curves. And that makes it a little difficult to understand the notion of limits. So first we had limits of sequences in $\mathbb{R}^n$, then using this idea, we defined the limits of functions. Or let me say a limit of a function at a point, limit of a function at a point.

So, the idea here was that, you if there is a point you are interested in, you have a scalar valued multivariable function, you allow all possible sequences which come to that point, find how the function values at along that sequence behave, and you hope that it converges to some point, meaning to some number.

And then if you obtain the same number, independent of what sequence you choose, which has limit as that point, then we will say the limit of the function at that point exists. And then if that limit happens to be the same as the value of the function at that point, then we say that the function is continuous at that point. So, using this we could talk about continuity of functions.

So, again the idea is the same that limit x tilde tends to a tilde f(x tilde) = f of you can take that f outside, f limit of x tilde tends to a tilde x tilde which is a tilde. So, that is f(a tilde). So, you can interchange the limit and the function that is exactly what is continuity. And yeah, so, we studied the notion of continuity.

So, this was more or less what we did for limits and continuity, and then we studied the notion of partial and directional derivatives. So, this we did in a fair amount of detail. So, the partial derivatives was so what, maybe before describing what they are, I should say, what was the main punch line of partial and directional derivatives.

So, the punch line is, I have a function of several variables. So, one way of studying that function is let me see what how the function changes, when I change just one variable, keep all other things constant. And let me see how it changes when I change one variable. So, the underlying theme here is restrict the function to a line in the domain and observe its behavior there. This was the main thrust of much of this.

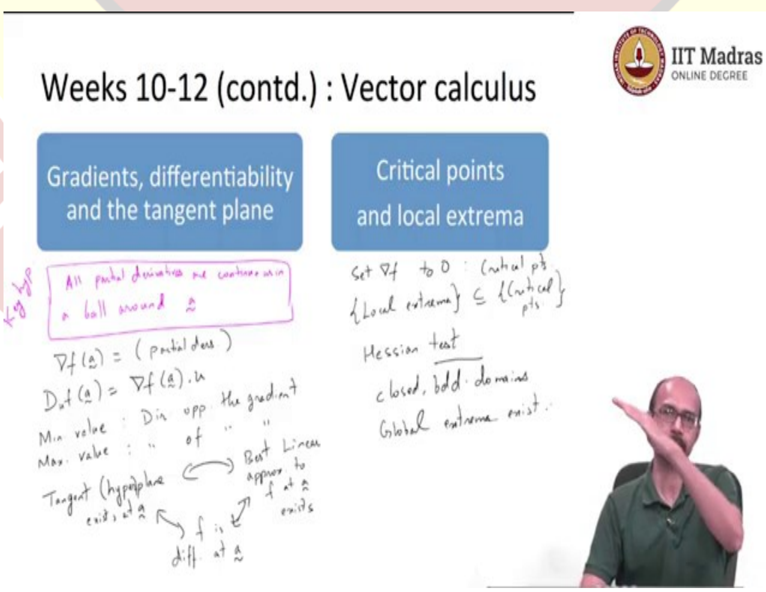
And so, first we restricted to all the axes and then we found the rate of change along that axis. So, rate of change along the parallel to the X_i axis maybe, so rate of change of f parallel to the Xi axis at the point a tilde. So, this was exactly the partial derivative of f at a tilde, this is exactly what it is, I mean, we wrote down the formula for this, meaning expression for this.

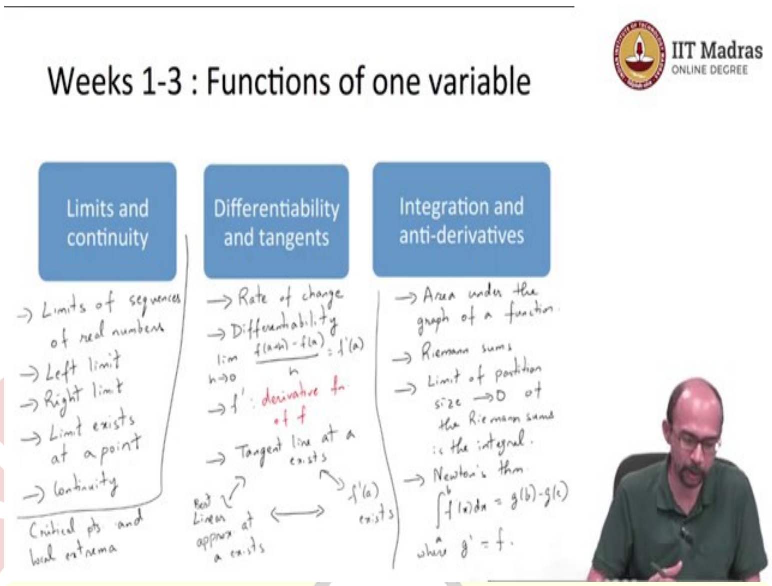
And instead of the axes, if you take a general line, which need not be the axis, so you have to take some direction, and you take the line in that direction. So, rate of change of f parallel to a line in some direction at a tilde, this is exactly what was the directional derivative. So, in order to compute it you have to choose a unit vector.

So, if u is a unit vector in that direction, that direction in the direction of the line. So, then we had this thing called Du of a tilde. And we defined it as a limit h tends to 0, f(a) tilde + h × u - f(a) tilde divided by h. And, of course, there is no guarantee that this limit exists, but if it does, then the directional derivative does exist. And if you take some particular values of u, meaning if you take the standard, which is vectors, if I take Ei, then I get exactly my partial derivative.

So, essentially, what we are saying is, in order to study a function of several variables, I will try to study it along one particular variable. So, if I have a plane, and I want to understand what is happening above the plane, what I will try to do is I will cut that plane, meaning I will take particular lines, see how it behaves on those lines, and then the hope is that I can put together all that information and understand the global behavior of that function.

(Refer Slide Time: 45:50)





Now, the directional derivative started becoming harder to compute as we went ahead. And so, we hope that there is some easy way of doing this and indeed there is provided your partial derivatives are continuous in the neighborhood. So, one of the key hypotheses in this entire business is this thing about the partial derivatives being continuous.

So, $\frac{\partial f}{\partial x}$ or maybe I should say all partials, all partial derivatives. So, key hypotheses I will write it here, all partial derivatives are continuous in a neighborhood or in a ball around the point a, this was really what this hypothesis was what drove all the theorems and computationally efficient methods of finding the directional derivative, etc.

So, if this happens, then we, first we define the gradient vector and this was easy, so the or the gradient function. So, this is you put together all your partial derivatives into a vector. So, put together all the partials in a vector, so first one, second one, third one, nth one. And then under this hypothesis, with u as a unit vector, we found that you can compute the directional derivative very easily using the gradient vector. And this was very important, first of all note here that we have, as I remarked earlier that we have used dot products. So, dot products that we studied earlier in linear algebra, this was one of the reasons why we did that.

And the second thing that we want to observe is that so the dot product we can write in terms of angle and because we, well angle and norm, angle and length, and because we can do that, we know that the minimum value and the maximum value, so this happens in the direction opposite the gradient. And this happened in the direction of the gradient.

And this is very, very important as I remarked, even in the video there, this is what you are going to use in order to do gradient descent which you will learn more, in a more streamlined way when you do machine learning basics, fine. And maybe I should also say that the that no change happens when you are orthogonal to the gradient vector. So, the notion of norms and lengths was used over here from, coming from linear algebra, fine. And of course, we use the vectors in general because the gradient is a vector, fine.

And then, we studied the notion of the tangent plane, we wrote down its equation using the gradient vector. And most importantly, we saw that, well in my slide I should have written tangent hyperplane. So, we saw that whatever we had seen first, earlier in the one variable situation namely the existence of the tangent plane, this happens exactly when the best linear approximation to f at a exists.

So, for both of these cases, we wrote down explicit equations, the equation to the tangent hyperplane was z = f(a tilde) + gradient of f at a tilde dot with x tilde - a tilde. So again, you can express that in terms of a dot product. But we had not really talked about differentiability here. And then later, we saw that or I remarked, actually, that this is the same as the function, f is differentiable at a, f is differentiable at a.

So, differentiability exactly means that the tangent hyperplane exists, which is the same as saying that the best linear approximation exists, fine. So, this is a general summary of gradients differentiability and the tangent hyperplane. And, yeah, I should have, again, I remind you that, for all of this, this was the key hypothesis. So, this hypothesis is what will often be used in when you want to do this.

Now, when you do courses outside of mathematics, often the theory involved, this hypothesis may not be checked. But you should know from looking at the function that this will work on that. For example, if it is a polynomial function, immediately you should say, this hypothesis works. So, there is no problem with using any of these things. Or if it involves a sin function, cosine function, you have good functions that you understand, and where you do not have denominators, or you maybe have denominators, but they are not becoming 0, etc. So, such cases, we can use all of this with impunity, because now this hypothesis is going to hold.

So, this you should remember as you go ahead, because in the further courses, remember that they are not going to do mathematics. So, they will not, they may not emphasize this hypothesis, so that you should mentally do if you are, if you have questions about how do I use this formula, etc, fine.

And finally, we also studied critical points and local extrema. So, critical points means this is again, something that we did for the one variable situation where we studied. Yeah, so I think maybe I did not mention that when we went there. So, here I should have said it here as well. So, when we did differentiability and tangents, we also did.

So, we also did a critical points and local extrema. And how did we find this? We had a critical point means you set your derivative to 0 and those points are called critical points, well, they also include those points where the function is not differentiable, but in our examples, usually it was differentiable. So, those are the kinds of functions you are likely to see when as you go ahead.

And then out of those, some are going to be local extrema, meaning local minimum or local maximum, and then we had the second derivative test in order to check for local maximum and we also saw that it could be inconclusive, fine. And there was also the optimization part where we had a closed interval, and there we know that the global minimum or global maximum exists.

And we saw that in order to find it, we have to find all the critical points include the neighboring sorry, the points, the endpoints of the interval, or if there is a function which is broken, then you include all those points at which there are breaks in the function. And meaning broken in the sense piecewise defined, so you choose all those points where the end of the pieces, and then add each of those values, points you check the values, and that will give you your maximum or minimum.

So, the same thing happened when we did vector calculus. So, critical point here, instead of the derivative, we now have the gradient vector, so set gradient to 0. So, that gives you your critical points. And out of these, some of them are going to be local, some or all are going to be local extrema.

So, the set of local extrema is a subset of the critical points. Here also, I should add the caveat that if the gradient exists, then you set it to 0 and find the critical point, there will be some vectors for which or some points for which the gradient does not exist, meaning some partial derivative is not defined. So, those are also critical points. So, in general, you have to take those also. Of course, in our examples, we did not really see examples of that type.

So, your local extrema is a subset of the critical points. And so, we had a hessian test in order to find this. So, this was what was the analog of the second derivative test in the one variable calculus case. And so, from the hessian test we could find the nature of the critical ones, of course, the hessian test is not a absolute test, meaning it could be inconclusive. So that is, that is part of how it works. So, if it is inconclusive, we have to find other ways of finding out whether they are local extrema.

Yeah, and then finally, we did all of this for closed bounded domains. So, again, as in the one variable case, the global extrema exist. And so, you have to find the critical point, also find the points on the boundary, find critical points there, then maybe you have boundary of boundary, and you have to keep doing that process, collect together all those points, find the function value at all these points. And whichever point it is maximum, that is your global maximum, whichever point is minimum that is your global minimum, this is how you find them.

Yeah, so you have to again use the gradient function. So, this is the overall summary of what we have done in these 12 weeks. This is a fairly fast course, it was, it went at quite high speed. So, I am aware of the fact that some of you may have had trouble keeping on. But really, there is you really do need to know, many of these things, or at least be familiar with many of these things, in order to appreciate the courses that are coming next.

So, this material is something that you do have to study. And well, in a way, it is my personal opinion is that if you are climbing up a mountain, the steepest part is best when it is at the beginning, then it becomes more gradually incline, and towards the top, it might actually become like a plateau.

So, if you have gone trekking, this is often something that happens. And so often in a lecture series or a degree program, the steepest claims are in the beginning, and then slowly, actually starts easing out. So, the moral of what I am trying to say is that, if you found this course difficult, do not worry, keep going, you will understand it. And it will actually I feel like if you understand this, well, right now, it will actually become easier as you go ahead, you will be very well prepared for what is coming next.

It is been a challenging time for all of us, especially in this last year, to do a lot of things. And this is one of the more challenging courses that I have given because it was very different from any course that I have taught before. In terms of both what has happened in the past year, as well as the fact that it is an online course, it is the first time, it is happening, at least in IITM or in the Indian setup. And also, it is a lot of material, which we usually teach in separate courses.

So, sort of bring it together was a little bit of a challenge. Of course, this course would not have been possible without all the team that has made these contents possible. I am just the face sort of I think many of you have also interacted with the team, especially in the live sessions. And then there is many others, those who edit and so on. So, I want to thank all of them. And thank all of you for your attention. Thank you.