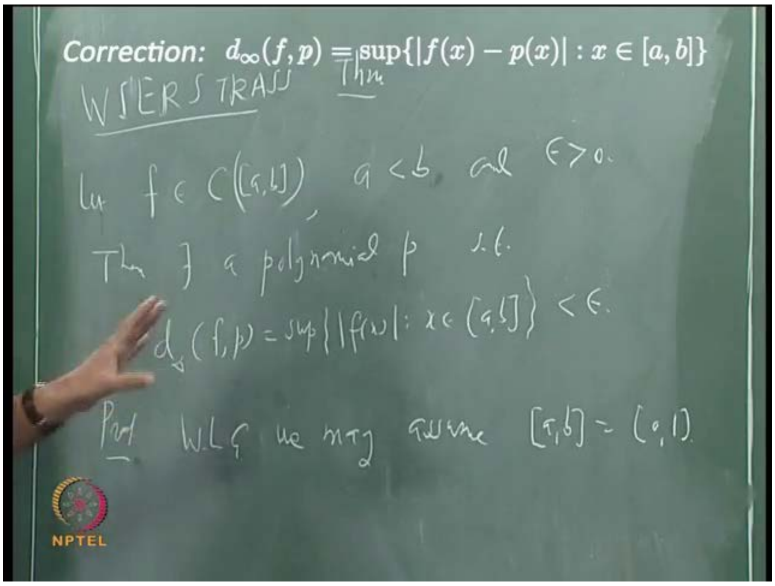
Real Analysis
Prof. S.H. Kulkarni
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 51
Approximation of a Continuous Function by Polynomials: Weierstrass Theorem

(Refer Slide Time: 00:22)

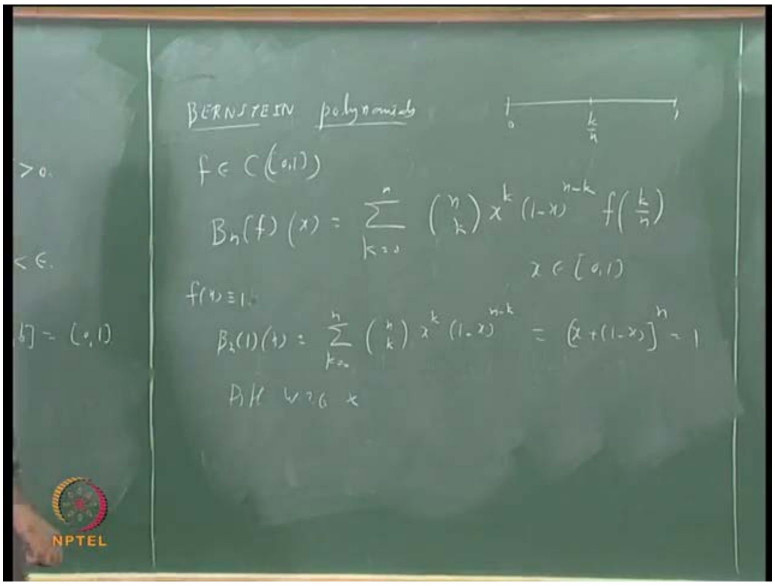


So we were discussing Weierstrass theorem, so let us recall the statement once again. So, Weierstrass theorem. Suppose, f is a continuous function on some interval a b. so let f be in C a b and we shall assume a less than b. And suppose some epsilon bigger than 0 is given then there exists a polynomial p such that the distance between f and p is less than epsilon. Let us recall what is this once again that is nothing but supremum of mod f x, where x belongs to a b, this is less than epsilon. In short you can assume any continuous function uniformly by a polynomial that is what the theorem says.

And another way of saying the same thing is that given any continuous function, you can find the sequence of polynomials, which converges to f uniformly. And this is a very well known and celebrated theorem and quite also, so there is used also in several places. 1 more thing is that there are large numbers of proofs of this theorem available in the literature. In fact you will find different proofs in different books. There is a 1 proof in Prudin, there is another in Siemens, and then there is another interesting proof in book by Lemay functional analysis.

So, what I am planning to discuss today is a proof as given in Siemens that is relatively simpler of course. The proof given in Lemay functional analysis is also good and quite interesting, so you should read it sometime. Now, the first step of the proof which we have done the other day, it is this that is first of all without loss of generality we may assume that the interval is 0 to 1. This is something we have discussed already, so let us not bother about it once again. Advantage of the proof which I am going to discuss from the Siemens book is that, it is a sort of constructive proof. It means it tells you what exactly the polynomial that approximates the given function is and those polynomials are also very well known polynomials, those are known as Bernstein polynomials.

(Refer Slide Time: 03:06)



Bernstein is a name of a Russian mathematician, Bernstein polynomials. We can define those polynomials for any f in C 0 1 and those is defined as follows. Suppose for each n you have 1 polynomial, so suppose I called that polynomial as B n f that will be a polynomial. So, we should know how it is defined. So, B n f at x this is defined as sigma k going from 0 to n is something similar to as something depending on the binomial coefficients.

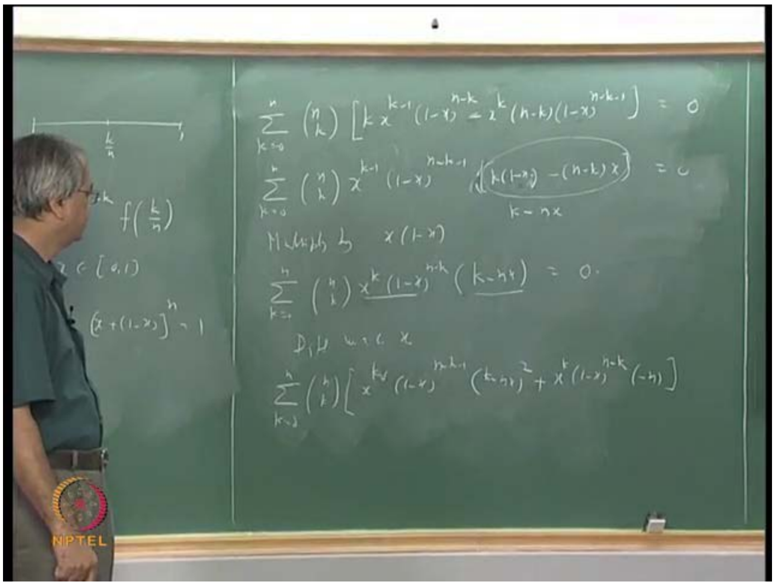
So, sigma k going from 0 to n, n c k that is n choose k, x to the power k multiplied by 1 minus x to the power n minus k and multiplied by f at k by n. This is for any x in 0 to 1. So, you can try to understand what is happening here is that suppose this is the interval 0 to 1, for any fixed n you subdivide the interval into n subintervals each of length 1 by n.

So, those n points will be 0 1 by n 2 by n etcetera arbitrarily k by n. So, evaluate suppose this is k by n, you evaluate f at k by n and multiply it by this x to the power k etcetera you will get some value. And you will do it for each point and take the sum.

Now, it is obvious that this is a polynomial because x to the power k into 1 minus x to the power n minus k is a polynomial and you are taking combinations of such term. So, this is a polynomial, it is a polynomial of degree n degree n. Now before proceeding further, we will need some properties of this problem not exact case of this problem some special case of these polynomials. So, suppose I take f as a constant function 1, so what will it be ,it will be b n 1 of x, that is you take f x as a constant function 1.

Then this will be a sigma k going from 0 to 1 n, n c k x to the power k 1 minus x to the power n minus k and this f at k by n will be 1 is a constant function 1. So, this is nothing but the binomial expansion of this x plus 1 minus x whole raised to n. So, this is the thing, but 1 and similarly, for your one interest if you can try to convert what is what suppose take f x equal to x, f x equal to x square, what is the corresponding Bernstein polynomial?

(Refer Slide Time: 07:03)



You can try to calculate on your own for things like x, x square, x cube etcetera, you need to have some simplification. It will not be as simple as this, but not very difficult also. What we plan to do further is that we want to get, this is one thing which we shall be using in the proof that is this is equal to 1. But there is 1 more identity, which we need and that is obtained by differentiating this with respect to x.

So, what we will get is sigma k going from 0 to n this n c k is anyway number, so I will just put it outside and then you differentiate. So, it will be k x to the power k minus 1 into 1 minus x to the power n minus k then plus x to the power k into it will be n minus k into 1 minus x to the power n minus k minus 1. And because of this minus x you will have minus sign here and this is equal to 0. Now let us simplify because in both of these terms what you have here?

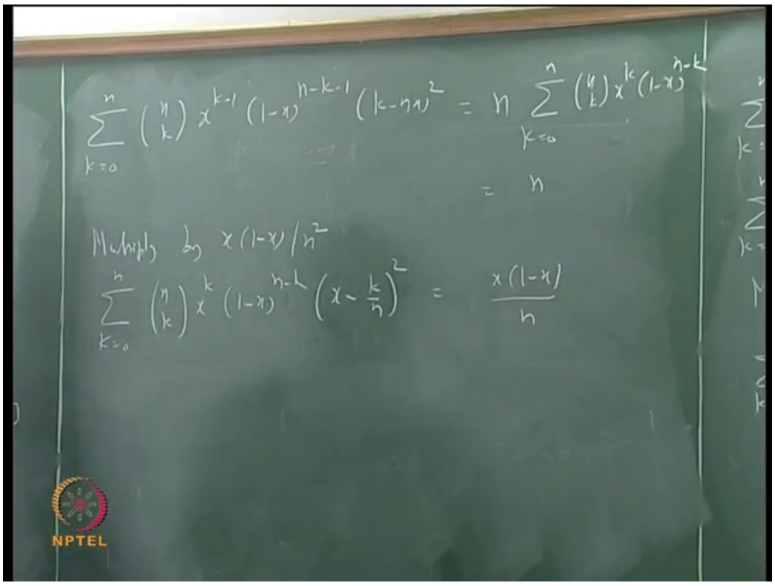
You have is x to the power k minus 1 into 1 minus x to the power n minus k minus 1 that will be common in both of these terms. So, this is same as saying that sigma k going from 0 to n, n c k I will write it here as x to the power k minus 1 into 1 minus x to the power n minus k minus 1 and what remains. From here you will get k into 1 minus x because 1 minus x to the power n minus k minus 1 is a common factor k into 1 minus x and from here minus n minus k into x that is equal to 0. So, simplify this so this k minus k x minus n x plus k x.

So, k x will cancel so this term will simplify to k minus n x is it k minus n x. Now, what I wanted to do further because we wanted to have obtained one identity by differentiating this once more, but before doing that I will simplify one more. So, that it looks more like this original polynomial. To do that I will multiply this by x into 1 minus x, so multiply by x into 1 minus x, so that power of this 2 will increase. So, by that we will get sigma k going from 0 to n, n c k x to the power k into 1 minus x to the power n minus k and then multiply it by k minus n x and that is equal to 0.

We have given this now differentiated this once again with respect to x. And since in the earlier step we have differentiate this step x to the power k into 1 minus x to the power n minus k and its derivative we have readymade. So, we can treat this as one term and this as another term in the parrot formula. So, while doing that we will do sigma k going from 0 to n is n c k remains as it is. So, first I have written derivative of this so what is that? It is x to the power k minus 1 into 1 minus x to the power n minus k minus 1 multiplied by k minus n x and this has to be multiplied by k minus n x.

So, this is k minus n x square k minus n x square then plus this term as it is x to the power k 1 minus x to the power n minus k and then derivative of this that is minus n. So, what I will do now is that this term is minus and here I will make plus. So, minus and this term I will take to the other side, so what we will get is that is sigma.

(Refer Slide Time: 12:42)

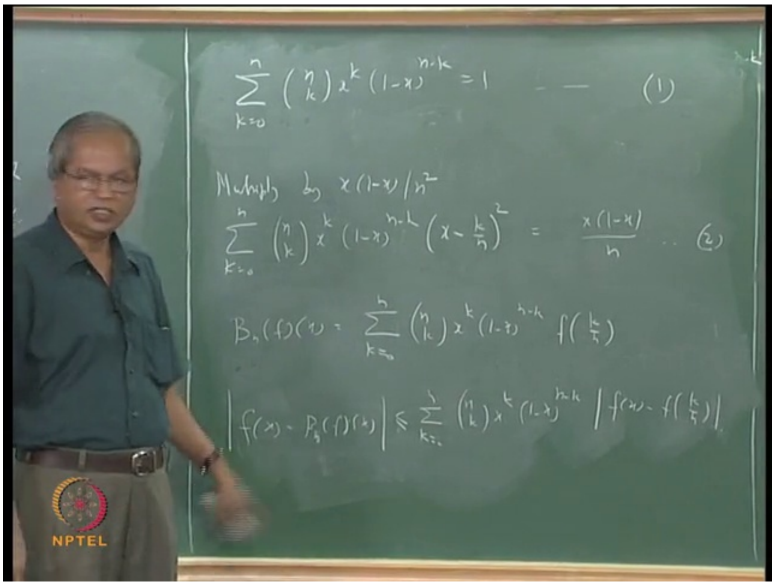


I think I will continue here sigma k going from 0 to n, n c k x to the power k minus 1 into, but minus x to the power z minus k minus 1 into k minus n x square into k minus n x square that is equal to n into sigma k going from 0 to 1 n c k x to the power k into 1 minus x to the power n minus k. Do you agree? But we know that this is 1 fine that is what we have. So, this is same as n then I will do the same thing once again whatever I did the previous step.

Multiply the whole thing by x into 1 minus x, so again multiply by x into 1 minus x. So, what we will get is this we wanted anyway sigma k going from 0 to n, n c k x to the power k 1 minus x to the power n minus k because we have multiplied by 1 minus x also multiply it by k minus n x square that will be same as n x into 1 minus x and I will do this simplification here itself. I will divide the whole thing by n square so here itself I have said multiply by x into 1 minus x by n square. So, what we will have right hand side will be x into 1 minus instead of this it will be x into 1 minus x by n.

And here it will be k by k by n minus x correct, so that is we are dividing this by n square. So, k minus n x by n whole square, so this will simply I will rewrite this as term as x minus k by n whole square that is equal to x into 1 minus x by n. This is an identity which we shall require.

(Refer Slide Time: 15:35)



So, this one and the first one which we started with I think this in between things I will remove it. What was the first one sigma k going from 0 to 1 n c k x to the power k 1 minus x to the power n minus k that is equal to 1. In fact this is we started with this differentiating this and all that, so this is 1 and this is 2. Now, we will look at let me remind it what was definition of B n f of f of x, it was sigma k going from 0 to n, n c k x to the power k into 1 minus x to the power n minus k into f of k by n f of k by n. So, suppose I want to consider this f x, our idea is to show that for large n mod f x minus b n x is less than epsilon. This is the required polynomial for large values of n that is what we want to show.

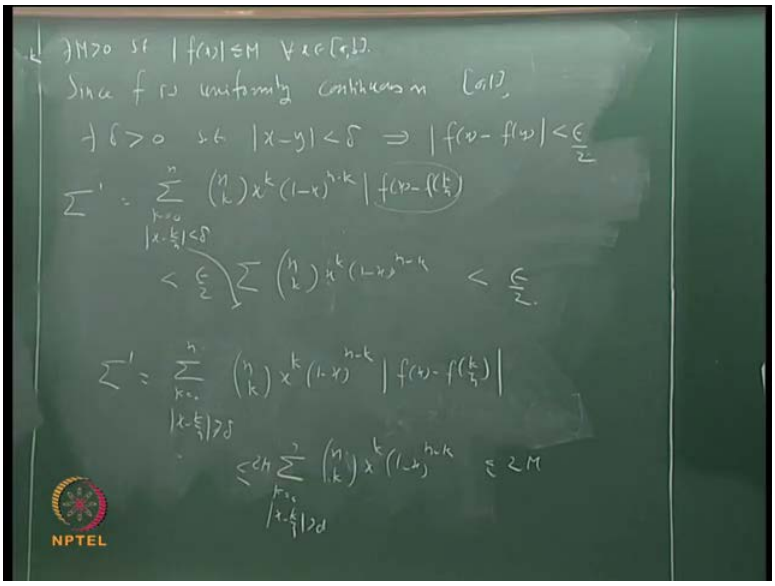
Now to do that we will have to estimate the difference between f x minus this polynomial, so to do that we will again do as follows. We will write f x minus B n f of x B n f of x is here that is what we have seen already. For f x I will use this identity I will multiply both sides of this identity by f x. So, what will happen it will be f x into this is equal to f x and I will take that f x inside this, what is the advantage? All these terms are common and what we will get is f x minus f of k by n.

So, I will write it like this, this is same as sigma k going from 0 to n, n c k x to the power k 1 minus x to the power n minus k and multiplied by f x minus f of k by n. Do you agree with this? So, here we have already used this identity and this identity. We want the difference between the absolute values here I will take the absolute values here. So, this equality will be replaced by less than or equal to this n c k is anywhere non-negative. Similarly, x is between 0 and 1 remember, everything is happening between the intervals.

So, x to the power k 1 minus x to the power n minus k all those are nonnegative terms. So, that will not be affected by taking absolute value. This will be this will become mod f x minus f k by n. Now, till now whatever we have done could have been done for any function. We have not used anywhere the fact that f is continuous till now. These are just algebraic manipulations. Up to this point I could have done for any function. Now, we shall use the fact that f is continuous, to show that this difference can be made less than epsilon, we have to show that each this sum can be made less than epsilon.

So, first we shall use the continuity and we also need to use this identity. Let us see how to proceed. Again since f is continuous in 0 to 1, you know that every continuous function on a compact set is uniformly continuous, so we can find some delta. So, we can say that since f is uniformly continuous on 0 to 1.

(Refer Slide Time: 20:01)



We can say that there exists delta bigger than 0 such that, wherever mod x minus y is less than delta, this implies mod f x minus f y is less than epsilon. Now, you will look at this sum here. Here f x minus f y f x minus f k by n. Now remember x is a fixed point, x is a point which in the interval. What are these k by n s that will depend on what is k, this is

0, 1, etcetera, it will be k by n here, x maybe somewhere here. So, let us say this is k by n for various values and let us say x is some such point here.

Now given any delta, there may be some case for which mod x minus k by n is less than delta and there maybe some case for which this is false, but at least for those case for which mod x minus k is less than delta, we can say that mod f x minus f k by n is less than let me make it epsilon by 2. The remaining sum also we shall make epsilon by 2. So, first what I can say is that I will split this sum into 2, so split this sum into two. So, the first term is for the sake of convenience let me simply write it here, suppose I call this as sigma prime plus sigma double prime.

So, what is sigma prime? Sigma prime is I take only those indices k for which mod x minus k by n is less than delta. Let us write sigma prime is this, sigma prime is sigma k going from 0 to n and mod x minus k by n less than delta, that is take only those delta for which mod x minus k by n is less than delta, whatever is this n c k x to the power k 1 minus x to the power n minus k and multiply by mod f x minus f of k by n.

Now, we know that this part is less than epsilon by 2, so for all such case that will come outside the summation sign, so this is less than epsilon by 2 into this sigma, whatever all this sigma etcetera. It is sigma k going from not k going from it is sigma over some case, but the terms involved are the same as this. Remember once I take this epsilon by 2 outside what remains inside is, let me just write once again here n c k into 1 minus x to the power k into 1 minus x to the power n minus k.

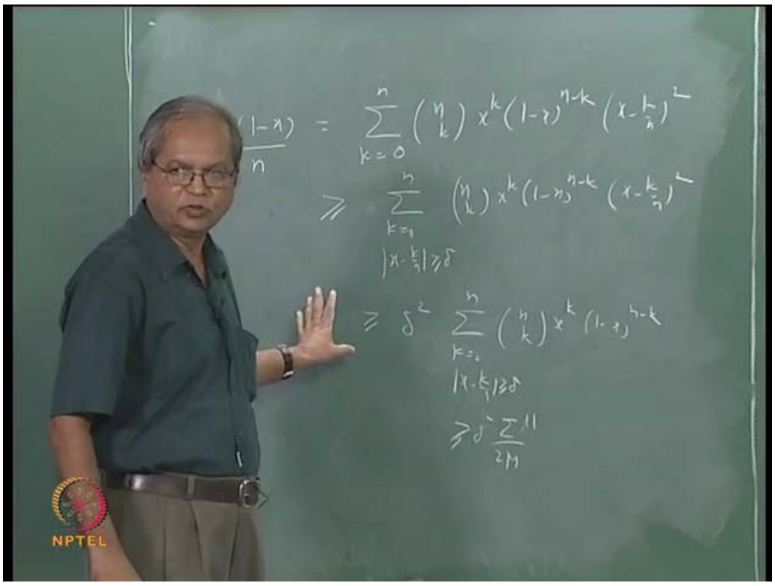
Only thing is it is not sum over all case, but sum over some case. Is it clear that this must be less than or equal to this because this is sum over all case going from 0 to n and we have just taken some terms from them. So, that should be less than or equal to 1, so this must be less than epsilon by 2 at least as far as this point is. Now, our idea is to show that this remaining part sigma double prime that is also less than epsilon by 2, but here it may not be for all n, but for large values of n, that is what we shall show. What is sigma double prime?

Sigma double prime that is nothing but sigma let me write here sigma double prime that is equal to sigma k going from 0 to n and with what case what x minus k by n is bigger than or equal to delta. And what are the terms here n c k x to the power k 1 minus x to the power n minus k into mod f x minus f at k by n. Now, as such we cannot say that mod f x minus f k by n is less than epsilon by 2 for this case, but we can say something. Since f is continuous function on the compact set it is bounded, so we can always find some number m such that mod f x is less than or equal to m for all.

We can say that 1 a b also we can say that there exists m bigger than 0 such that, mod f x less than or equal to m for all x in a. So, what we can say is that this is less than or equal to sigma k going from 0 to n mod x minus k by n bigger than delta due to all this n c k less than or equal to 2 m into this, 2 m into sigma k going from 0 to n, n c k x to the power k into 1 minus x to the power n minus k. And again we can say that since x to the power k 1 minus x again whatever remains here that is less than or equal to this.

So, this is less than or equal to 2 m, but of course that is something we could have got even otherwise what we also use here is the following. We have not yet used this identity. The way in which we planned to use is this because for those case for which mod x minus k by n is bigger than or equal to delta, this term x minus k by n square that will be bigger than or equal to delta. So, since sigma primes involves those term x to the power k into f x minus f of k by n etcetera, we shall use this fact here.

(Refer Slide Time: 28:42)



This is from here what follows is the following that is x into 1 minus x is this sigma k going from 0 to n, n c k x to the power k 1 minus x to the power n minus k into x minus k by n square. That is something that this divided by n x into 1 minus x divided by n.

What we can say is that this is bigger than or equal to sigma k going from 0 to n, but instead of taking some over all k.

Suppose, I take sum over only those k for which mod x minus k by n is bigger than or equal to delta n c k x to the power k 1 minus x to the power n minus k into x minus k by n whole square. All that we are doing is we are taking only sum of this terms here and all the terms are positive. So, if we drop some of this terms whatever remains should be less than or equal to the original sum, but what we know is that for this case mod x minus k by n is bigger than or equal to delta.

So, this term x minus k by n square whole square that will be bigger than or equal to delta square. So, we can say that this is bigger than or equal to delta square into sigma k going from 0 to n, n c k x to the power k 1 minus x to the power n minus k not k going for 0 k going from 0 to n mod x minus k by n bigger than or equal to delta. And what now you compare what expression we have got here with this. Here you have the same 2 m into n c k x to the power k 1 minus x to the power k mod x minus k by n.

So, what we have got here is that here we have got that, this expression is bigger than or equal to delta square into this. And this expression which we have here is same as this last not forget about this. This is same as this expression that we have here. What is the requirement there, for the time being I will write as this is bigger than or equal to sigma double prime by 2 m.

Forget about this part, this is something that we have not used. So, what we have is this expression sigma k going from that is sum over this over this case is bigger than or equal to sigma double prime by 2 m. Is that clear whatever we have done till this?

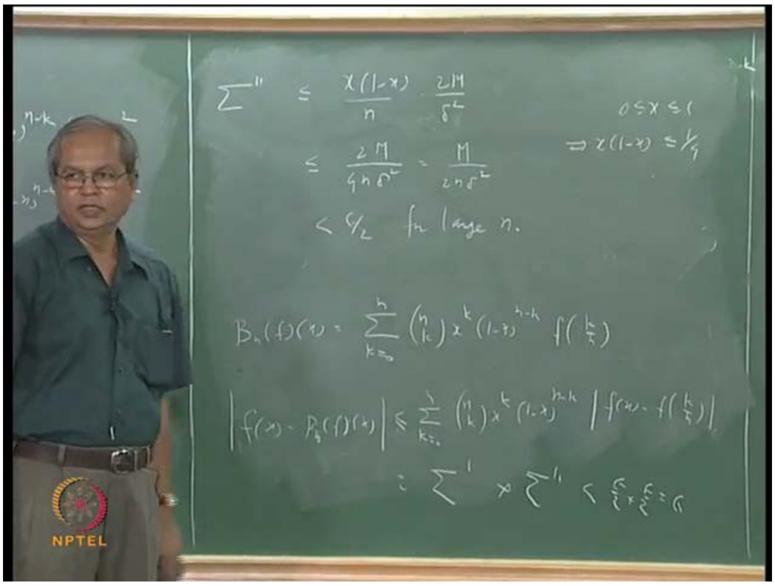
Student: ((Refer time: 32:22))

What is the doubt?

Student: ((Refer time: 32:32))

This delta squares, that is delta square sigma double prime by 2. Now we will just take this, whatever we have got we want to write sigma double prime less than or equal to something, that is our objective.

(Refer Slide Time: 32:57)



So, what I have got sigma double prime less than or equal to x into 1 minus x by n into 2 m delta square. Do you agree with this? Of course, you can check all this calculations once again I mean those are fairly simple. So, these are just adjusting some terms in the in the sum. Now, again remember the whole thing is happening in the interval 0 to 1 0 less than or equal to x less than or equal to 1. Do you agree with in this interval x into 1 minus x will be less than or equal to 1 by 4 in the interval 0 to 1 that is for 0 less than or equal to x less than or equal to 1.

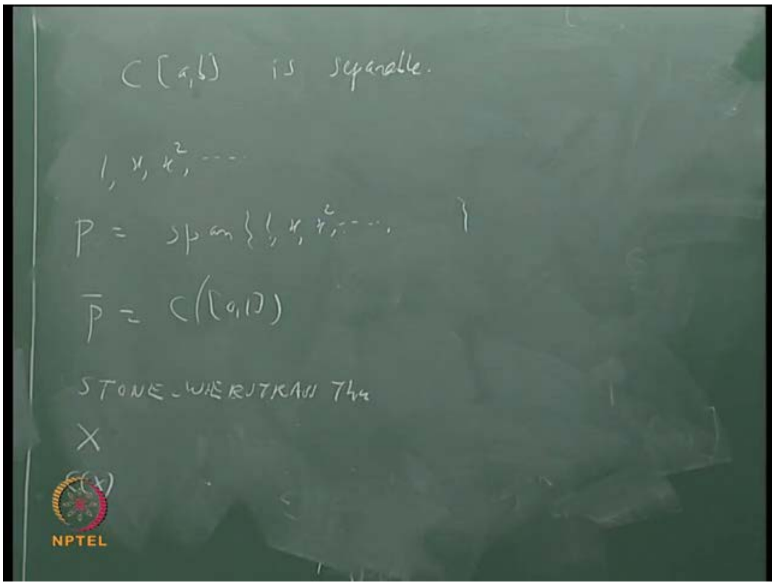
This will imply that x into 1 minus x is less than or equal to 1 by 4 for all x. It is a elementary maximum problem, so this I can remove now. So, using this we will get, this is less than or equal to 2 m by 4 n delta, that is this is less than 1 by 4. Already you have 2 m and delta square or that is same as m by 2 n deltas square. Now, is it clear that by choosing n large enough this can be because this m and delta square are fixed m 2 delta square are fixing.

So, by choosing n large enough this can be made less than epsilon by 2, so we can say that this is can be made less than epsilon by 2 for large n. So, again coming back to this, this we wanted to look at this difference. This difference is less than or equal to this two parts sum, this is sigma prime and sigma double prime. Sigma prime is less than epsilon by 2 that is what we have shown. And remember this is for every x whatever estimates we have done, we may not start with some particular x, but ultimately we have removed the dependence on x.

We have removed the dependence on x because for example, if you look at here up to this x was fixed, but after you come to this part; this is less than epsilon by 2 for every x. Similarly, up to this there was a dependence on x, but after this we say that whatever be the x, x into 1 minus x is less than or equal to 1 by 4. So, that what I wanted to say that this n does not depend on x because there is no x here.

So, this inequality is that is this will be less than epsilon by 2 plus epsilon by 2 or which is equal to epsilon, that is mod f x minus B n f of x is less than epsilon for large and for that delta, that is independent of x. In other words I have removed the statement, the difference between that is what is the supremum of this for x in 0 to 1 that supremum is less than epsilon. Now, I have already mentioned that one of the ways of using weierstrass theorem is that, you can show that this C a b is separable.

(Refer Slide Time: 36:42)



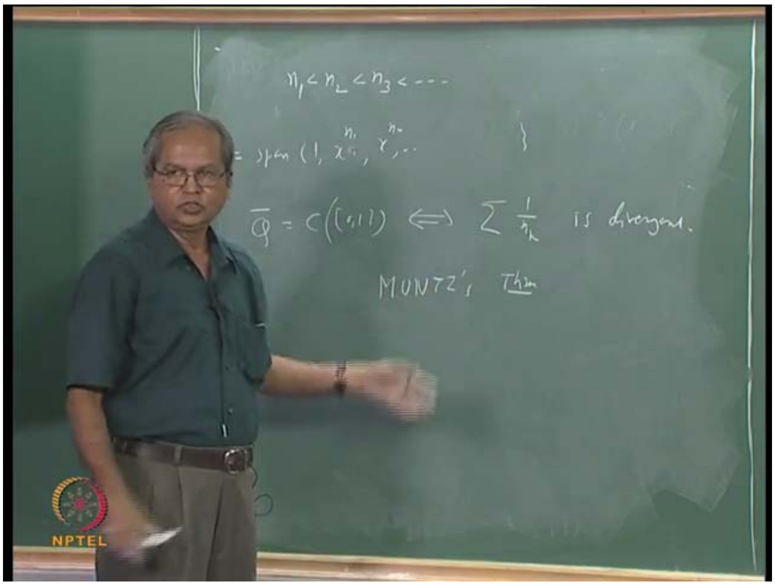
This is something that we have discussed already. And let us see what this weierstrass theorem says, if we look to that once again if I look at this space suppose I look at these polynomials. Let us say 1, x, x square etcetera, etcetera if I look at this polynomial, and see these all are polynomials. Suppose, I look at the span of this 1, x, x square etcetera suppose at this is nothing but a set of all polynomials. Previous span of this function 1, x, x square etcetera any element in the linear span will be a polynomial and every polynomial is in the linear span.

Student: ((Refer time: 37:37))

So, that is p is dense in C end or that is same as saying that p closer is C 0 1. As I said there are several generalisations of Weierstrass theorems, one generalisation is to replace this set a, b by any compact metric space or the compact what is called compact house doss space. And those are that theory was also very famous theorem, it is known as stone Weierstrass theorem. Stone Weierstrass theorem where it says that suppose you take x as any compact metric space then what are the subsets which are dense in C x.

That is theorem which gives characterisations of those sets, but we shall not go into the discussion of that because that will take us too much away from these, that we are discussing here. But there is one more interesting thing that one can ask is that whether all this functions are essential, whether we can drop some of these functions and still the span of the remaining things will be dense and which can be dropped. And in that direction there is one very interesting thing and that is the following, suppose instead of this set p, suppose we take some other set.

(Refer Slide Time: 39:08)



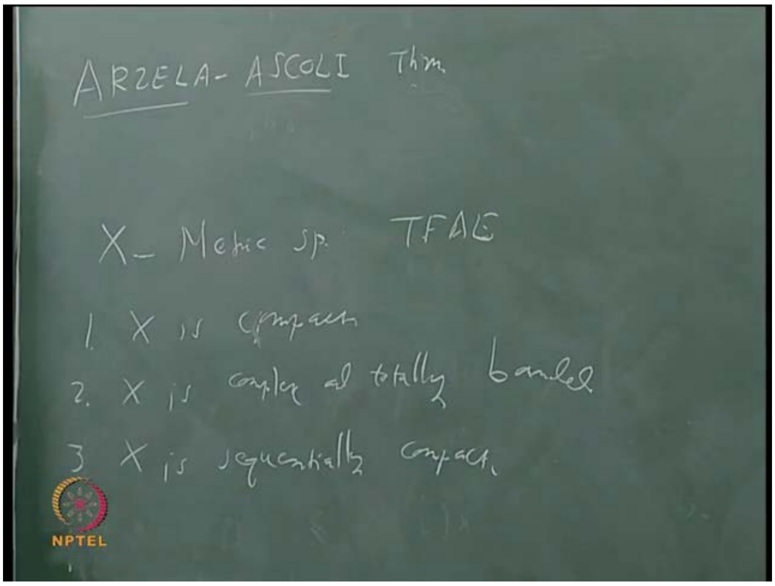
Suppose, we take this set 1 x to the power n 1 x to the power n 2 etcetera where, n 1 n 2 etcetera they are some natural numbers n 1 less than n 2 less than n 3 etcetera. Suppose you pick up some such numbers and suppose you take the span of this and call it this as q. And suppose we ask a question when is this dense in C 0 1, let us say. When is this or for what values of n 1, n 2, n 3 this is dense in C 0 1 the answer is very interesting and totally unexpected.

Totally unexpected in the sense from whatever we know at this level it. The answer is the following it is cube that is q bar q is there that is q bar is equal to C 0 1 if and only if, you take this series sigma 1 by n k, this is a divergent series. This is also a very famous theorem known as Muntz theorem. Can you see that Weierstrass theorem satisfies this condition because if you take all of the if n 1 is 1 n 2 is 2 etcetera, this will be sigma 1 by n, that is a divergent series.

So, the question is which n 1 n 2 n 3 you can select; you can select only those if they form if sigma 1 by n k forms a divergent series. That is why it apparently it is totally unrelated things and so that is why somewhat unexpected in the beginning. Now, coming to the discussion back with the uniform conversions, what we have already seen is that if we take any sequence of functions and if it is uniformly convergent, there it is point wise convergent.

And we have seen that converse is false and we have seen some special cases where, the converse maybe true, for example, Denis theorem. Now there is one more theorem of a similar type, which we would like to see. And here we ask the question suppose, we do not bother so much about the uniform convergence of the given sequence, does there exist a uniformly convergent sub sequence. Suppose, even if the original sequence does not converge at least does there exists some sub sequence which converges uniformly.

(Refer Slide Time: 42:20)



So, there is a theorem which answers this question and that is also very-very famous theorem and that is called Arzela-Ascoli theorem. Arzela and Ascoli both were Italian mathematicians. Now, to again discuss this theorem let us discuss a few preliminary things. You would have already discussed the various properties of metric spaces particular compact metric spaces. And there are various equivalent characterisations of compact metric spaces.

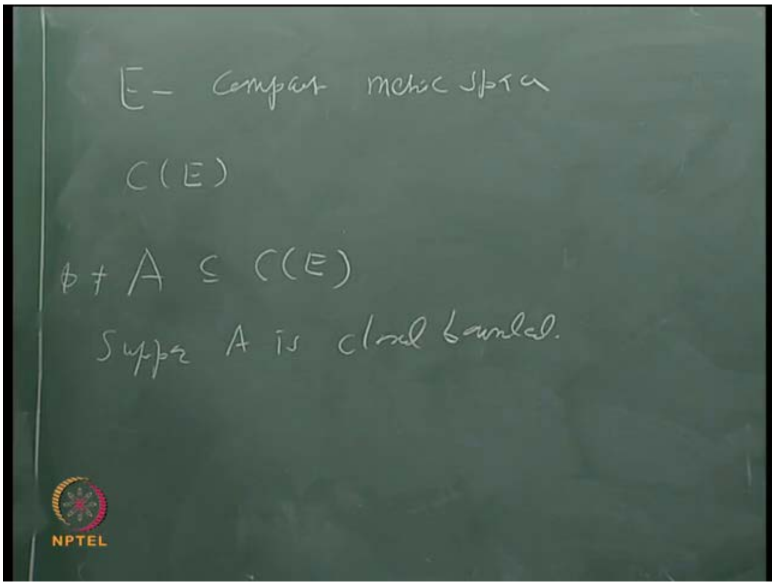
And just now I mentioned about the convergence of a sub sequence, you know that there is a characterisation of compactness also in terms of using sub sequence. And what is that, we say that the space is compact metric space is compact if and only if, it is sequentially compact, that is every sequence has a convergent sub sequence and it is also equivalent to saying that it is complete and totally bounded. We will find that this is something we should remember always that is suppose we take x is a metric space, you may not have studied this, but it is useful to recall this.

Then the following statements are equivalent first is that x is compact 2 x is complete and totally bounded x is complete and totally bounded and 3 x is sequentially compact. I simply write x is sequentially compact x is sequentially compact and these are the characterisations, which are useful to decide which sets are compact in the well known well known metric spaces. So, suppose we take the let us say the usual spaces R n etcetera then you know that there is another famous theorem that known as harboral theorem, which says that a set is compact if and only if it is closed and bounded.

It is closed and bounded, but you also know that this is not true in general. It is true only in r 2, r n etcetera it is true only for example, in other metric spaces, you can find the spaces sets which are closed and bounded, but not compact. In fact it can be shown that this happens only in finite dimensional spaces that is if you take non-linear space that is vector space on which a norm is defined and which gives a metric then every closed and bounded set is compact if and only if the space is finite dimensional.

This is a very well known theorem, but that you will learn in functional analysis course that requires some techniques in functional analysis. But anyway it raises one obvious question that suppose, you take an arbitrary infinite dimensional space for example, the space which we are considering here let us say that E is a compact metric space. And consider C of E then suppose I want to know what the compact subsets of C of E are?

(Refer Slide Time: 45:35)

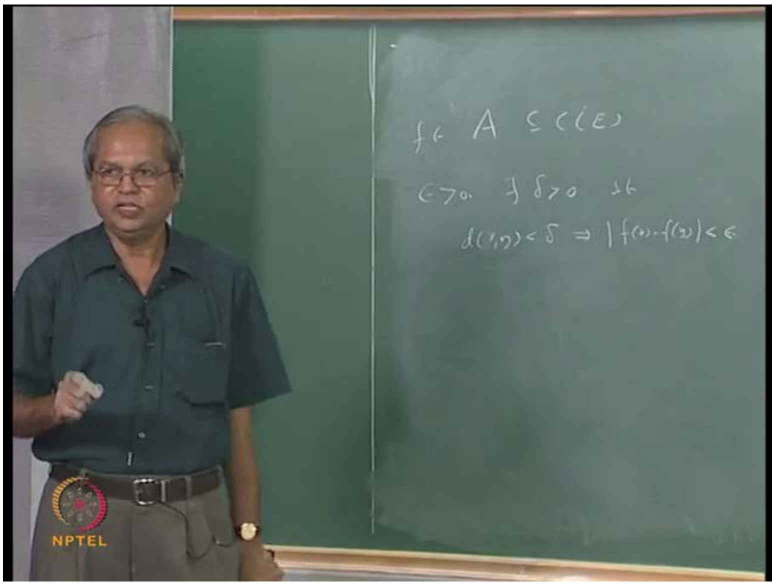


This is an infinite dimensional space since, it is an infinite dimensional space you can obviously find the sets, which are closed and bounded, but not compact. So, the question that we are asking is the following. Remember compact set is always closed and bounded, only thing is converse cannot be true. And for the converse to be true it is just boundary is not enough what we require is totally bounded. And that is where the difference in finite dimensional spaces boundedness and total boundedness are the same, but these are different in infinite dimensional spaces.

So, what are the sets which are in addition to being closed and bounded? What gives the compactness that is the question that we will ask and that is the question, which this theorem answers in this particular space C. Suppose, you take the set of all continuous functions on a metric space and what is the answer? Let us say what is the question let us take this set or non empty set A, A is contained in C. Of course, let us take non empty of set and suppose A is closed and bounded then the question is what additional conditions are required to make a compact and that is the question that this Arzela-Ascoli theorem answers.

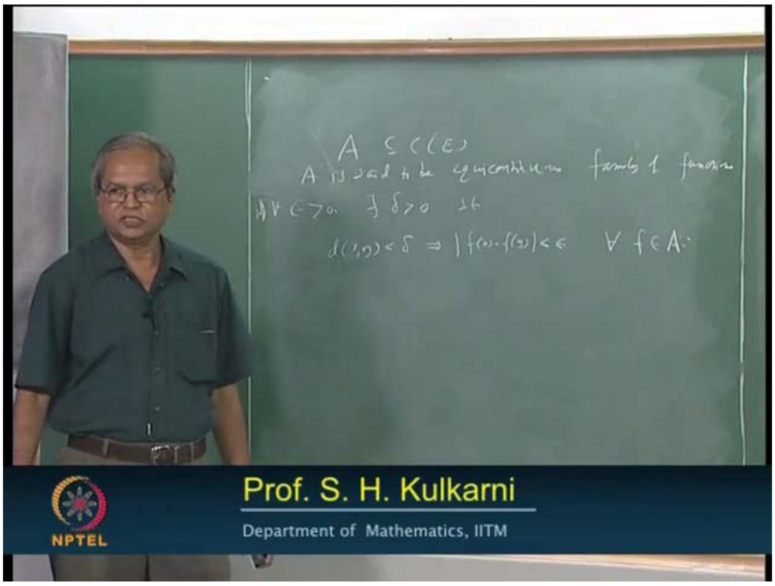
And there is a property that makes this compact and that is the property, which is called equicontinuity. So, the theorem will ultimately say that A is compact if and only if A is equicontinuous. Now, to understand that we will have to do what is meant by saying that A is equicontinuous? So, we shall just make this definition today and the details we shall do in the tomorrow class.

(Refer Slide Time: 48:12)



Remember this A is a subset of C of E, so if you take any f in A then that is a continuous function in C of E. And since it is a compact metric space, it is also uniformly continuous. So, suppose you are given epsilon bigger than 0 then there exists delta bigger than 0 such that, distance between x and y less than delta, this should imply mod f x minus f y less than epsilon. That is whenever the distance between x and y is less than delta that is the meaning of saying that f is continuous and of course, in this case continuous is same. So, given epsilon you can find the delta, we already know that this delta will not depend on what is x because it is uniformly continuous, but it may depend on what is f?

(Refer Slide Time: 49:32)



If A is a family of functions so given an epsilon for different f there may be different delta, but suppose you can find one delta which works for all f then that family of functions is called equicontinuous family of functions. So, we will say that A is said to be equicontinuous family of functions, if for every epsilon bigger than 0 there exists delta bigger than 0 such that, distance between x y less than delta implies mod f x minus f y less then epsilon for every f in A.

So, when we talk of equicontinuity, we are talking about the family of functions and in particular it can be a sequence of functions also, but it can be any family of functions. So, equicontinuity is a property that a family of functions may or may not have. We will we will discuss this things more in tomorrow.