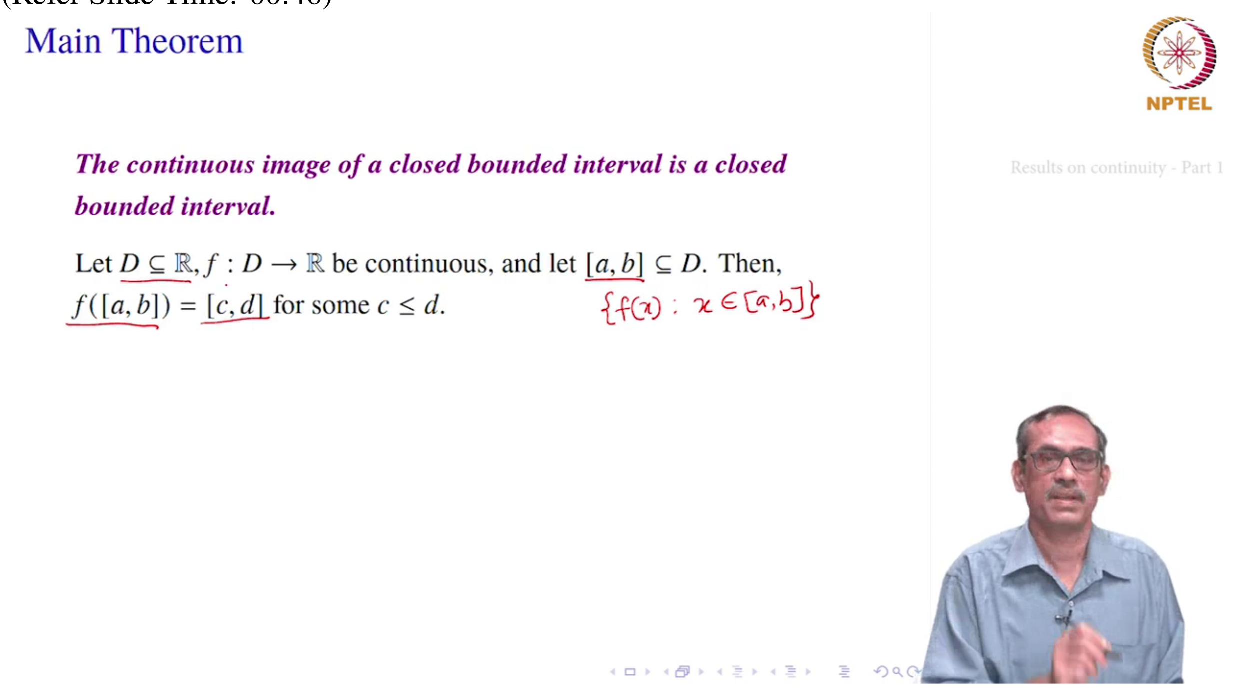
Basic Calculus - 1
Professor. Arindama Singh
Department of Mathematics
Indian Institute of Technology Madras
Lecture 12 - Part 1
Results on Continuity - Part 1

This is lecture 12 of Basic Calculus 1. Recall that we had discussed continuity, that is, continuous functions and some algebra of continuous functions. Today we will be discussing one result, a very specific result, and then some of its corollaries which are themselves very important results on continuous functions, and solve some problems basing on those ideas.

(Refer Slide Time: 00:46)



The main theorem here is that the continuous image of a closed-bounded interval is a closed-bounded interval. That is our main result. We will not prove this result. In fact, it requires the completeness principle. Some rigorous analytical tools are needed to show that this really holds. You remember it this way: the continuous image of a closed-bounded interval is a closed-bounded interval.

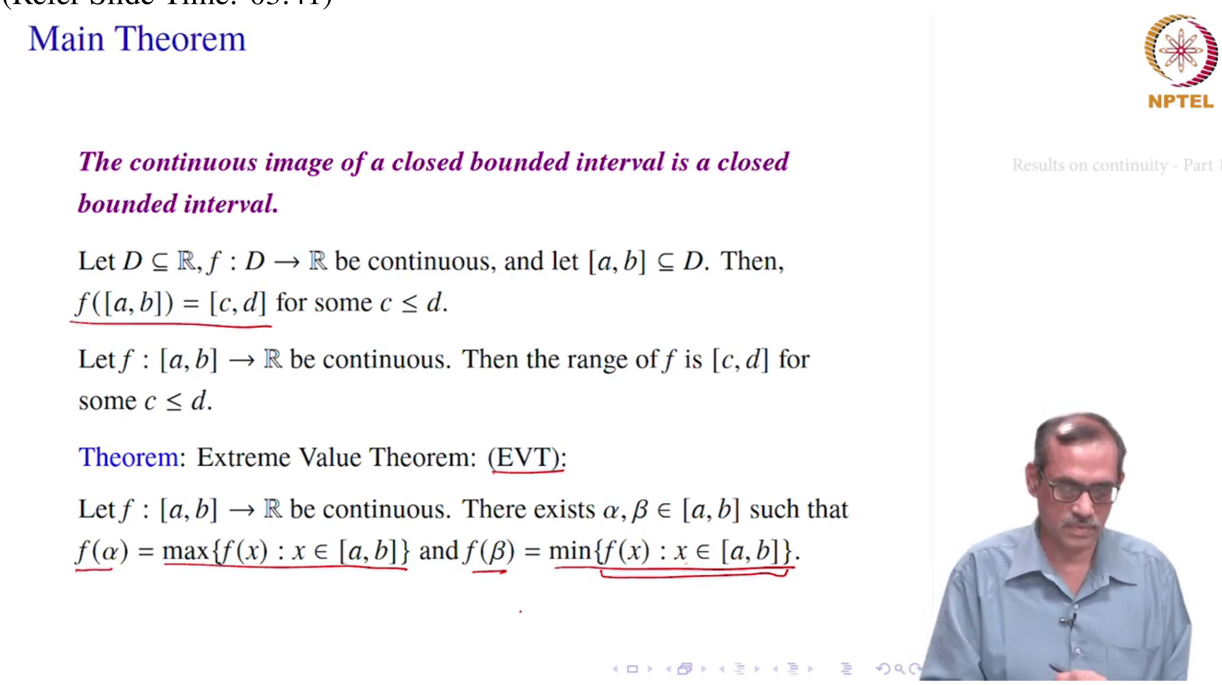
Actually it means this. Suppose D is a subset of $\mathbb{R}$; it is some set of real numbers; and you have a function defined on this D taking real values. Let that function f be continuous. Let us choose any closed-bounded interval, which is $[a, b]$ with $a \leq b$. We will really take $a < b$; the other one $a = b$ will give a singleton. To avoid it, let $a < b$; that is assumed here implicitly. Let us take one such closed-bounded interval which is a subset of D. That means D is such a set that it contains at least one closed-bounded interval. Of course, we are only interested in defining continuous functions on intervals or unions of intervals. So, these assumption is not a big thing for us.

Suppose D contains such a closed bounded interval. Let us take any closed interval $[a, b]$ which is a subset of D. Then its image which is $f([a, b])$ is the set of all values $f(x)$ such that x varies over the closed interval $[a, b]$. Recall that this is denoted by $f([a, b])$. This $f([a, b])$ will be equal to a closed-bounded interval, which is $[c, d]$ for some $c \leq d$. If f is a constant function, of course, you may get 'equal to' here. So, we will take 'equal to' sign also as a possibility. This is what we mean by the continuous image of a closed bounded interval, that is, $f([a, b])$ is a closed bounded interval which is $[c, d]$, where f is assumed to be continuous. This is the main theorem.

This may not happen if f is not continuous, or if you do not start with a closed bounded interval. Even if f is continuous, if you do not start with a closed bounded interval, then you may not reach a similar interval there.

We will not prove this, but we will use it. Usually, in calculus, we do not define functions over any arbitrary subset of $\mathbb{R}$ to $\mathbb{R}$; we only take domains of functions as intervals.

(Refer Slide Time: 03:41)

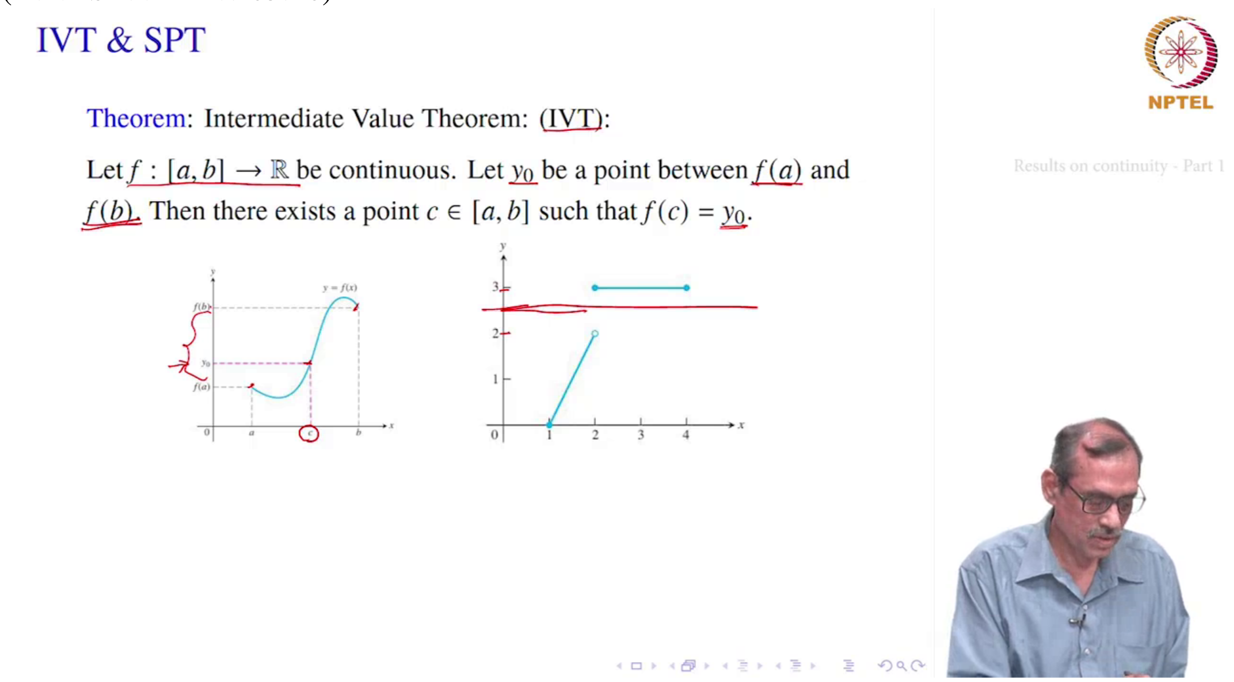


Again, there is a reformulation of this theorem, where we start directly with "f is defined on a closed interval $[a, b]$. That is how it is usually mentioned. Let $f : [a, b] \to \mathbb{R}$ be continuous. Then its range, which is $f([a, b])$ is equal to $[c, d]$ for some $c \leq d$. This is a very convenient form to use.

Once we understand this, we can go to the next theorem, which follows from this. It is called the extreme value theorem. Let us try to understand what it says. It says that if $f : [a, b] \to \mathbb{R}$ is continuous, then there is a number which is minimum of all $f(x)$ for $x \in [a, b]$. Now, that follows because $f([a, b]) = [c, d]$; and here, c is the minimum of all values of $f(x)$. That is, the minimum of this exists, and that is really c. Similarly, d will be the maximum of all values $f(x)$. So, maximum of the range also exists. That means there is one point on the left side, in $[a, b]$ such that f of that point is equal to c; similarly, for d. So, there exist $\alpha, \beta \in [a, b]$ such that $f(\alpha)$ is the minimum of all $f(x)$ and $f(\beta)$ is the maximum of all $f(x)$, where $x \in [a, b]$.

So we say that a continuous function achieves its maximum and minimum values. Of course, this way, it is informal, but that is how we remember most of the things by looking at them informally. And that is why this is called the extreme value theorem, that a continuous function achieves its extreme values, minimum and maximum values. Of course, that follows from the other theorem.

(Refer Slide Time: 05:46)



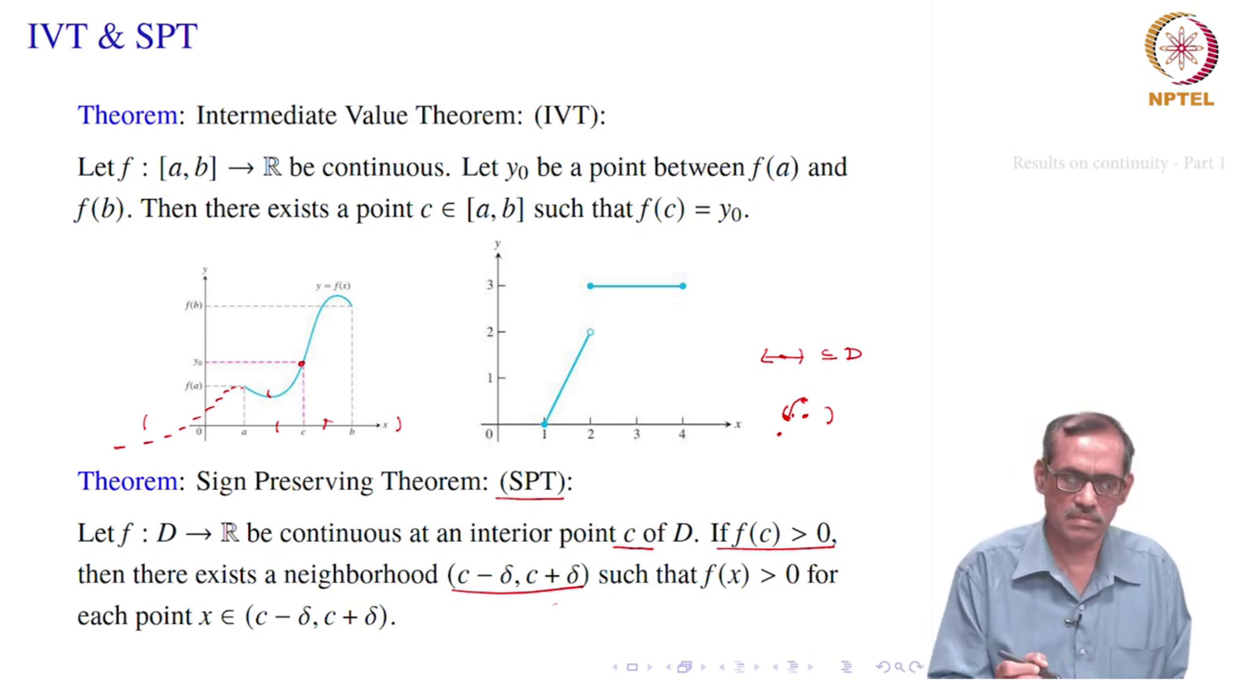
There is one more important thing which is called intermediate value theorem. What it says is this. Suppose f is defined over a closed bounded interval taking real values and it is continuous. Now, $f(a)$ and $f(b)$ are certain values in $\mathbb{R}$. You take any point in between $f(a)$ and $f(b)$, say, it is y_0. Look at the picture. Then it says that there exists a point $c \in [a, b]$ such that $f(c) = y_0$.

Extreme value theorem says that the minimum and maximum of f in the range are achieved; that is, there are some points α and β where $f(\alpha)$ and $f(\beta)$ give us the minimum and the maximum. But here goes a bit further; it says that between $f(a)$ and $f(b)$ if you choose any number, then that number is also achieved. That is, there is a point $c \in [a, b]$ such that $f(c)$ is equal to that number. You can see geometrically why it is so. Now, $f(a)$ is somewhere here, $f(b)$ is somewhere here; y_0 is in between them; that is, if you take a horizontal line $y = y_0$, then it will cross the curve somewhere. That gives us the point.

Of course, it does not say that there is a unique one. Suppose, it goes on, comes down and again and you draw one line, it may hit at many of the points. So, there can be many points but there exists at least one. That is what it claims. On the other hand, if it is not continuous, for example, like this, then this may not happen; see the second picture. If I choose a point between 2 and 3, something, say 2.5, then draw that horizontal line $y = 2.5$, it will never hit the curve anywhere. There is no point in the x-axis such that $f(x) = 2.5$, but 2.5 is between 2 and 3.

So, that is what the intermediate value theorem says. The extreme value theorem says that the two extremes are achieved, whereas the intermediate value theorem says that even in between points are also achieved. Again, this follows from the mother theorem, the main theorem that $f([a,b]) = [c,d]$. Clearly, if you choose some two numbers between the minimum and the maximum, then those two points also are achieved. That is what extreme value theorem says. There are nice applications of these results, which we will see soon.

(Refer Slide Time: 08:56)



We will see one more result, which comes from this. It is called the sign preserving theorem. This is also very important theorem. It gives an idea of what continuity in a small neighborhood means. Suppose f is a continuous function. We will start with an interior point c of the domain D of f. For example, if f is defined on the closed interval $[a,b]$, then c will not be the end points a and b, it is something else, in between a and b but not equal to a or b.

Once you say it is an interior point, recall that it means there is a neighborhood of that point, which is contained inside D. That is why it is called an interior point. Always there is a neighborhood around it which is a subset of D. So, that means f is defined on that neighborhood at least.

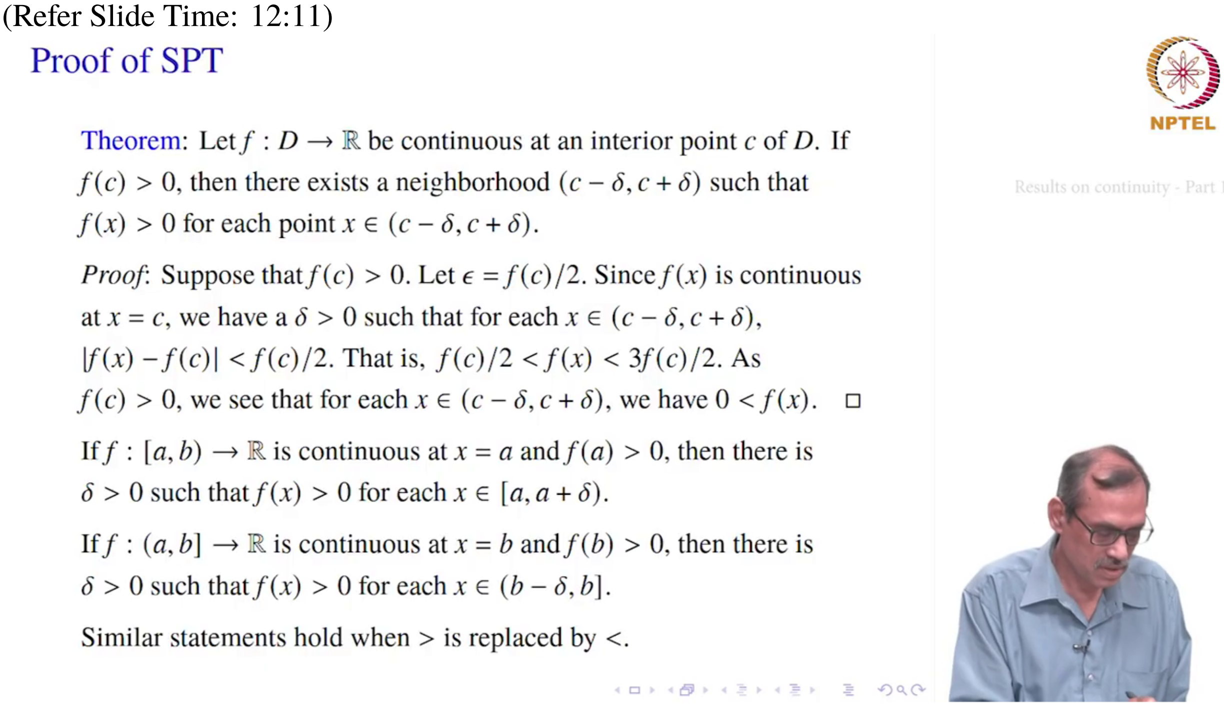
Suppose $f(c) > 0$; something like here, c is greater than 0. Look at this, and forget the earlier ligend. Suppose f is a continuous function and c is there so that $f(c) > 0$. Then, it says that there is a neighborhood of c, $(c-\delta, c+\delta)$ such that on the whole neighborhood, the function remains positive.

See, I can choose δ here something, between this, in fact I can choose bigger. But suppose the function comes down to like this, somewhere, then I cannot choose my δ to be so big, because f of this will be negative. It says that there exists a neighborhood where f remains positive. How does it follow? Why does it follow intuitively? Let us see.

Suppose it is remaining positive, and it does not happen. That means whatever neighborhood you take around that point, you will find some negative point or 0. In fact, if you get a negative point, then because of intermediate value theorem, 0 is also achieved. So, 0 is there somewhere. That means, you have the function, you have c, you take any neighborhood, then, you will find there is a point where it is 0. In fact, if you take any point, it will be 0 or negative. This is in contradiction to what the theorem states. So, on the contrary, assume that. Because this is the result to happen. If it does not happen, then you take c, any neighborhood of this, around this you take any point, there it is either less than 0 or equal to 0.

Now, here it is positive and here it is negative or it is equal to 0. Then by the Intermediate value theorem, all these values should have been achieved, between 0 and that. So, there is a point where it is 0. Once all these values, say this, there is no 0 elsewhere here; we can always choose one point, after that 0 is never taken, otherwise everywhere it is 0. And directly from that point, it will fall to 0, it will not be continuous. That is our intuitive feeling. And of course, we can prove this. It is easy to prove without using the intermediate value theorem; it is more fundamental.

(Refer Slide Time: 12:11)



Let us give a proof. Let $f : D \to \mathbb{R}$ be continuous at an interior point c of D. If $f(c) > 0$, then there exists a neighborhood such that $f(x) > 0$ in that neighborhood. So, what we do: suppose $f(c) > 0$. It looks something like this; c is somewhere here, and $f(c) > 0$. Then what we do is, we start with something which is the difference between 0 and this, we will take that height, say, that is our ϵ. We will get a corresponding neighborhood so that $f(x) - f(c)$ will lie between that epsilon. So, let us choose something less than that, say, less than half of that. We will see what is the advantage of choosing that. Suppose $f(c) > 0$. Now, let us take $\epsilon = f(c)/2$, which is less than the difference of 0 and $f(c)$. Now, for this ϵ, since f is continuous, there is a $\delta > 0$ such that whenever $x \in (c-\delta, c+\delta)$, this neighborhood, we have the difference between $f(x)$ and $f(c)$ less than ϵ, which is $f(c)/2$.

If you rewrite this as an inequality, this says that $|f(x) - f(c)| < f(c)/2$. This gives you $f(c)/2 < f(x) < 3f(c)/2$. So, there is a neighborhood which is $c - \delta, c + \delta)$ such that for each point x in that neighborhood, $f(x)$ lies between $f(c)/2$ to $3f(c)/2$. As $f(c)/2 > 0$, we have $f(x) > 0$. That is, for each x in the neighborhood $(c - \delta, c + \delta)$, our $f(x)$ is bigger than 0, and that is the proof.

This talks about the interior point c. It is an interior point means around c, there is a neighborhood, which is contained inside D. Suppose c is not an interior point. We are concerned about intervals. Say, D is an interval and c is one of the end points of D. It looks something like $f : [a, b] \to \mathbb{R}$, it is also continuous at $x = a$ and $f(a) > 0$. Then what can we say about the neighborhood? We cannot say there is a neighborhood around a, which is inside $[a, b]$. But we have only a right neighborhood which is contained inside the domain $[a, b]$. In that case, the result comes to this. There is a $\delta > 0$ such that $f(x)$ remains positive in the neighborhood $[a, a + \delta)$. That, of course, is covered in this same proof because by definition when we say continuous at a left endpoint, we will have a right neighborhood only where these things happen. The same proof with a little modification holds for the right endpoint. Here, instead of this, we will be taking c to $c + \delta$ and continue the rest of the proof. That is, when it is the right endpoint, $f[a, b] \to \mathbb{R}$ is continuous at $x = b$ with $f(b) > 0$, then we have a left neighborhood $(b - \delta, b]$ where $f(x)$ remains positive. I think this is understood.

But then this is what is positive. If you have negative, say, $f(c) < 0$, then of course, you will get a neighborhood where $f(x)$ remains negative. The proof, of course, is easy because you consider $-f$ instead of f throughout. The same proof will work.

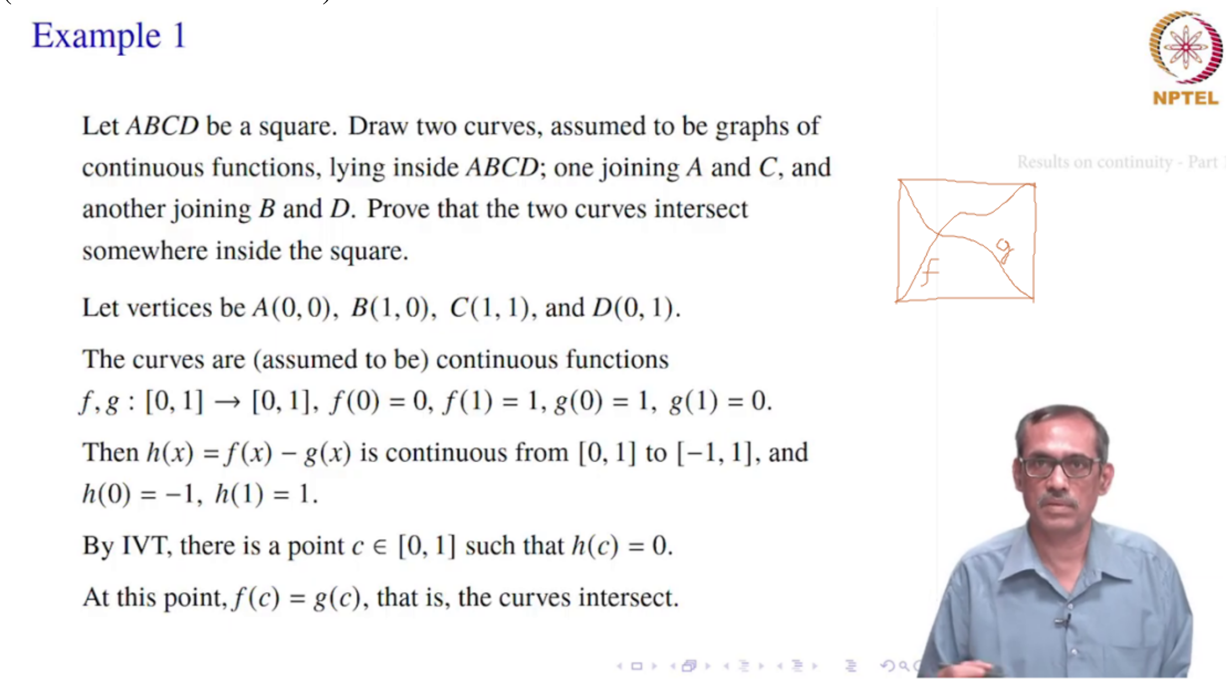
That is why we give a comment that a similar statement holds when you replace this greater than with less than. That is, if f is negative at a point and it is continuous, then it will remain negative in a neighborhood of the point. In fact you can think of it is bigger than some number k, right, $f(c) > k$. Then, there is a neighborhood of c where $f(x)$ remains bigger than k. Fine? Same holds for less than also.

These are our main results. We have the extreme value theorem, intermediate value theorem, and the sign preserving property of continuous functions in a neighborhood. Let us give some example where we can apply these results.

This is our first example. We have a square $ABCD$. It is something like this. Let us name them the vertices following some order. $ABCD$ is a square. We got two curves, which are assumed to be graphs of continuous functions. That means, you do not do like this, you should not do like this; they cannot come to this side. Is that okay? That will not be happening there. Suppose we start with a curve which is assumed to be a continuous function. In that case, what will happen?

Let us do it this way. We have a square, say $ABCD$; we are trying to draw two curves which are assumed to be graphs of continuous functions, lying inside $ABCD$. They will not go away. One curve joins A and C, and the other joins B and D. The curve joining A and C is something like this and the other which joins B and D is something like this; both lie inside the square. Obviously, what is asked is to show is that two curves intersect somewhere inside this square. It is intuitively very obvious, but we require proof. So, what do we do?

(Refer Slide Time: 19:11)



Let us write A to be the origin. It is done the other way. We take this to be A and this to be B, this to be C, and this to be D. So, this point, let us write $A(0,0)$. Let us write B as $(1,0)1$, C as $(1,1)$; and D as $(0,1)$, taking this as the y-axis and this as the x-axis.

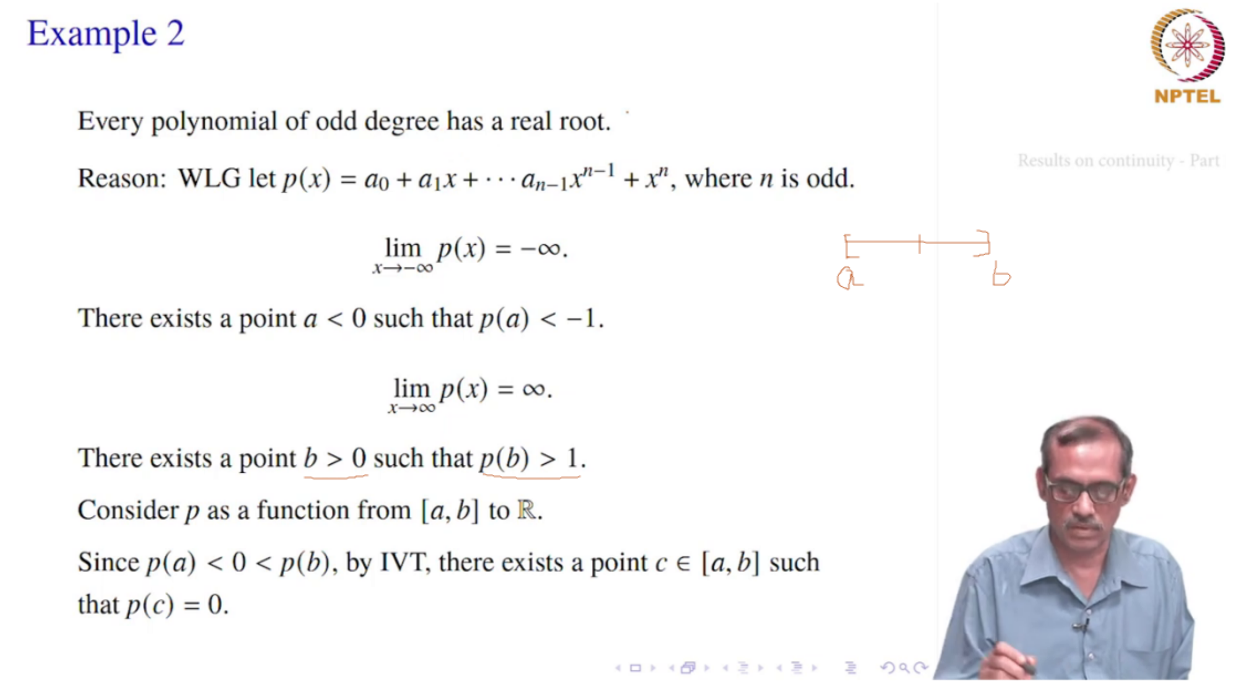
Now, f is a function and g is also a function. Both are defined from $[0,1]$ to $[0,1]$. It is defined from $[0,1]$ to $[0,1]$. One function is like this, and the other function is like this. So, one is defined from this to this. Both of their height remain between 0 and 1. So, f and g are functions from the closed interval $[0,1]$ to the closed interval $[0,1]$; that is obvious. Now, what is $f(0)$? Her, $f(0) = 0$. This is our f and this is our g. So, $f(0) = 0$, $f(1) = 1$, $g(0) = 1$, and $g(1) = 0$. This is what we get. Both are continuous functions. I think this much information is enough. We will not need the picture after this.

What happens is, we define another function $h(x) = f(x) - g(x)$. We know that this is continuous, because of the algebra of continuous functions. This is continuous from where? From $[0,1]$, and to what? What is its range? Its range will be -1 to 1 since both $f(x)$ and $g(x)$ have maximum value as 1 and minimum value as 0. So, the range of $h(x)$ is $[-1,1]$. That is, $h : [0,1] \to [-1,1]$ is a continuous function with $h(0) = f(0) - g(0) = -1$ and $h(1) = f(1) - g(1) = 1$.

Now, use intermediate value theorem. Between -1 to 1, lies 0. So, there is one number here, say, x such that $h(x) = 0$. So, there is a point c such that $h9c) = 0$. And what is the meaning of $h(c) = 0$? It is really $f(c) - g(c) = 0$. That i, $f(c) = g(c)$. And that is the meaning of the curves intersect. So, this is an application of intermediate value theorem.

Let us take another example. So, here it is claimed that every polynomial of odd degree has a real root. We want to show that it is correct. And you take any polynomial of odd degree, it will have a real root. All polynomials may not have real roots. This claims it only about odd degree polynomials, like $x^2 + 1 = 0$ does not have any real root. However, if you take a cubic, then there exists at least one real root.

(Refer Slide Time: 21:17)



So, how do we proceed? We start with a polynomial, say $p(x) = a_0 + a_1x + \cdots + a_{n-1}x^{n-1} + x^n$, where n is odd. Here, we have taken the coefficient of x^n as 1. There is no loss in generality, because anyway, we are interested in finding a root. So, if this has a root, then the general one also will have a root because it is of degree n, $a_n \neq 0$; and then dividing by a_n we get a polynomial where the coefficient of x^n is 1. So, let us assume that $p(x)$ is in this form, where the coefficient of x^n is 1, and n is odd.

Now, what happens if I take the limit of this as x goes to ∞ and as x goes to $-\infty$? In the polynomial $p(x)$, I take x bigger than these coefficients, that is, let $x > |a_0| + |a_1| + \cdots + |a_{n-1}| + 1$. Then, this term x^n will dominate all the other terms. So, when x goes to $-\infty$, the value of this polynomial can be made smaller than any negative number, since n is odd. We thus say that the limit of $p(x)$ as x goes to $-\infty$ is equal to $-\infty$. So, this becomes minus infinity. Had it been even, it would have become infinity. Now, similarly, the limit of $p(x)$ as x goes to ∞ is equal to ∞. So, this dominant term tells you what is the result as $x \longrightarrow \pm\infty$.

Let us see what is the meaning of this. The limit of $p(x)$ as x goes to $-\infty$ is equal to $-\infty$. It means there exists a negative number a such that $p(a) < -1$. In fact, it means that $p(x)$ can be made smaller than any negative number you give me by choosing x. That is you give me something, say, -100. Then I can get one x such that $p(x) < -100$. Here, in particular, we say there exists a number $a < 0$ such that $p(a) < -1$. Similarly, when we have the limit of $p(x)$ as $x \longrightarrow \infty$ is equal to ∞, it means there exists a number $b > 0$ such that $p(b) > 1$.

Now, $p(x)$ is a continuous function. We can restrict it to the domain $[a, b]$ instead of taking the whole of real numbers. Now, there is this function $p(x)$, where $p(a)$ is some number less than -1, and $p(b)$ is some number bigger than 1. Therefore, by intermediate value theorem, since 0 lies between -1 to 1, there exists one point c such that $p(c) = 0$. That is what we wanted, that $p(c)$ must be equal to 0 for some c. So, any odd degree polynomial with real coefficients has a real root; and that is the proof.