IIT Madras

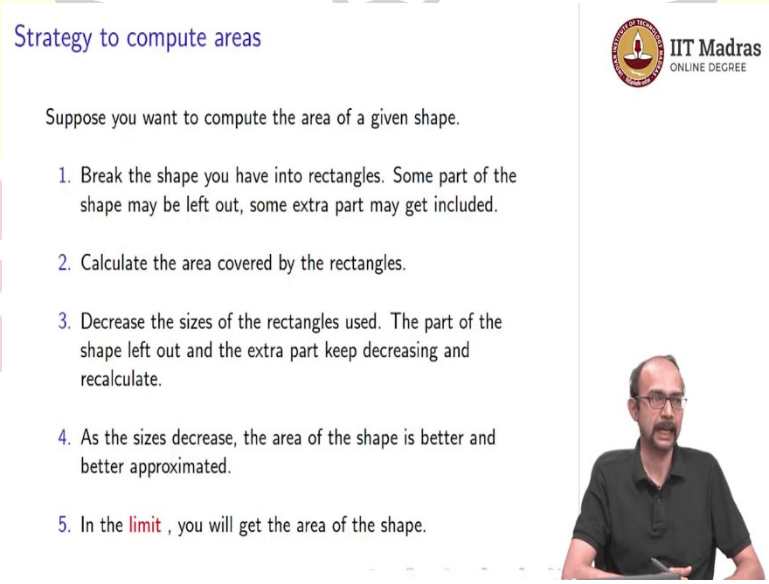
ONLINE DEGREE

Mathematics for Data Science 2
Professor. Sarang Sane
Department of Mathematics
Indian Institute of Technology, Madras
Lecture No. 13
Riemann sums and the integral

Hello, and welcome to the maths 2 component of the online BSc program on data science and programming. This video is about Riemann sums and the integral. So, let us recall what we did in the previous video. In the previous video, we computed areas and the underlying theme over there was, if you have an arbitrary shape, then you try to approximate it as well as you can by rectangles, you keep dividing and dividing and dividing, making smaller and smaller rectangles.

And hopefully, you will cover the shape entirely. And as you take smaller and smaller rectangles in the limit, the sum of the areas underlying those rectangles will give you the shape of, the area of the shape under consideration. So, there were different ways of course, sometimes we instead of rectangles, we use polygons or trapeziums or triangles, so other shapes, which were all derived from the rectangle. So, the rectangle was our basic building block.

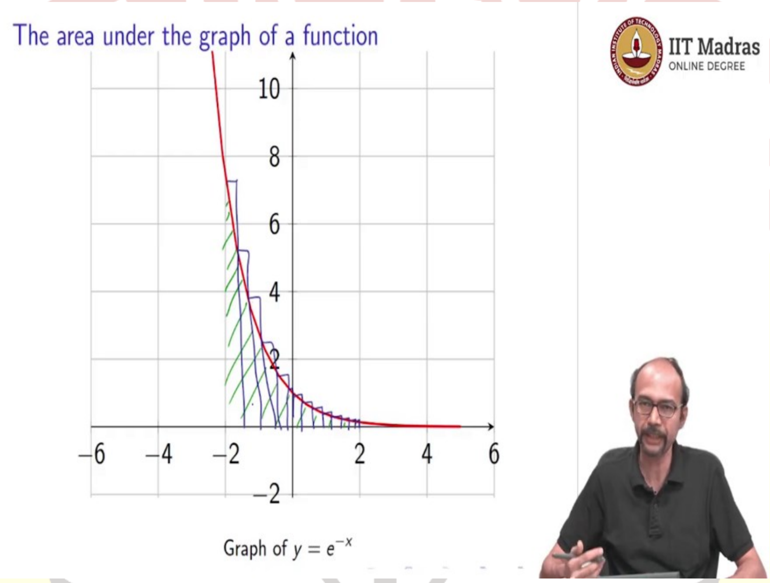
(Refer Slide Time: 1:16)



So, let us, let us make that precise. So, here is a strategy to compute areas. Suppose you want to compute the area of a given shape, break the shape you have into rectangles, some part of the shape may be left out, some extra part may get included, so we saw this in the previous video. Calculate the area covered by the rectangles. Decrease the sizes of the rectangles used, the part of the shape that gets left out and the extra part that is added in, these keep decreasing, because you are taking smaller and smaller rectangles, so you can approximate your shape better.

And then each time you do this, you recalculate the area. And as the sizes decrease, the area of the shape is better and better approximated. So, in the limit, you will get the area of the shape. So, we are going to use this strategy in sort of coherent fashion.

(Refer Slide Time: 2:14)



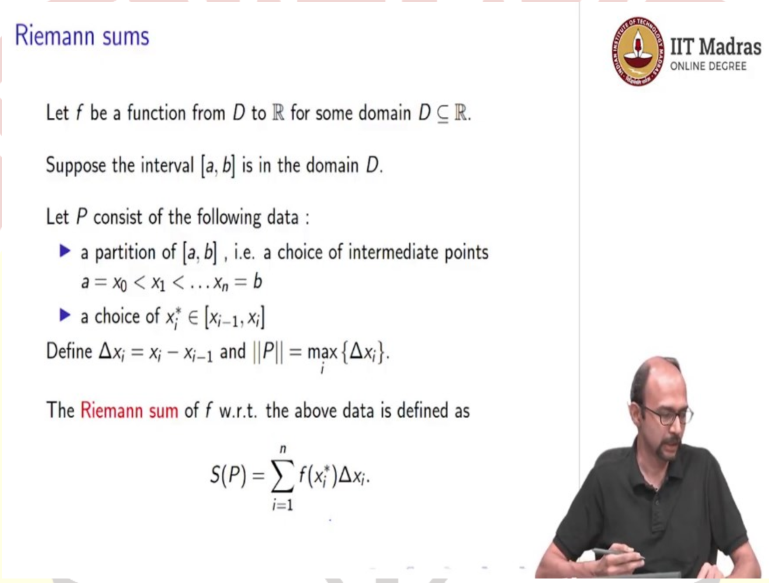
And we are going to talk about the area under the graph of a function. So, here is an example of the function $y = e^{-x}$. So, let us say I want to approximate the area under the graph. So, suppose let us say I want to get this this area over here. Let us say I want to get this area over here between $[-2, 2]$. So, in other courses, you may have come across this notion where the computing this area was important, particularly in statistics, which you either have already seen or will see soon. And so we want to know how to go about doing this.

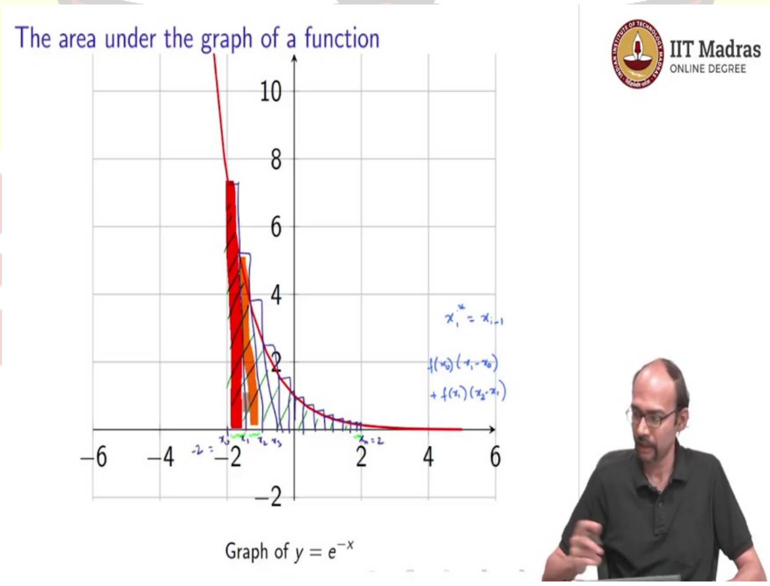
So, here is the area I want to compute. So, one strategy for doing this is to approximate this by rectangles. What kind of rectangles do we take? So, here is one, one idea about what kind of rectangles to take. So, we could choose rectangles like this. So, here is -2. So, I choose a rectangle like this, I choose a rectangle like this. And then this, then here, here, here, here, you can see what I am doing.

So, you take this, this set of rectangles, you compute its area, and then what you do, you shrink these intervals, and takes take smaller and smaller rectangles. And you can see that if you take small enough rectangles, this area that you want, which is in green, you can approximate it as well as you want.

And in the limit, meaning if you keep doing this process and allow your rectangles to get smaller and smaller and smaller, then you are likely to actually compute the area over here. So, you could use this strategy for any graph. So, here in particular, we are doing it for $y = e^{-x}$ between $[-2, 2]$

(Refer Slide Time: 4:40)





So, this process was sort of formally introduced in the 1800’s or 1900’s and there is a name for these kinds of sums, they are called Riemann sums. So, I am going to describe whatever I did in a previous slide in a more mathematical way. So, let f be a function from the domain D to the real line. Suppose the interval $[a,b]$ is in the domain D. So, that means is a function f is defined on the interval $[a,b]$.

So, let P consist of the following data. So, P is here, P is supposed to be for partition, along with some other data. So, partition of $[a,b]$, so that means a choice of intermediate points is $x_0 < x_1 < \cdots x_n$, which is b. So, in our previous picture, these things that I have drawn here, this would be your x_0, this first thing would be x_1, x_2, x_3, all the way up to x_n, which was 2 and $x_3$is -2.

So, this is a partition of your interval $[a,b]$, you could choose your partition so that it is uniform, meaning the distance between x_0 and $x_1$is the same as x_1 and x_2; x_2 and x_3 and so on or you may choose an arbitrary partition. So, along with this partition, you also want to choose for each small interval that is contained in the partition. So, x_{i-1} to x_i, you choose some number x_i^*, star from that interval. This is the data P.

So, let us repeat what we are doing, we are taking a function f, which we know takes values on the interval $[a,b]$, it is defined on the interval $[a,b]$ and then we are dividing $[a,b]$ into 'n' parts. And then in each of these parts, we are choosing a point, this is what we have done so far.

Define $\Delta x_i = x_i - x_{i-1}$, that means the length of the i'th piece. That is what Δx_i is. And this thing ‖P‖ = $\overset{max}{i}\{\Delta x_i\}$, which we will call it, norm of P. So, ‖P‖, let us call it norm of P. And this is just a name for now, do not bother too much about what it means. So, let us call it, so then let us define that as the maximum of these Δx_i's.

So, in particular, if you choose all your lengths to be the same, then this will be just that number. So, in your previous picture, the Δx_1 is this distance here, this distance here, and Δx_2 is this distance here, etcetera. And Δx_n is this distance here. So, at least the way I have drawn it, it is not the same. So, your ‖P‖ will be the maximum amongst these distances. So, that is what Δx_i and ‖P‖ are.

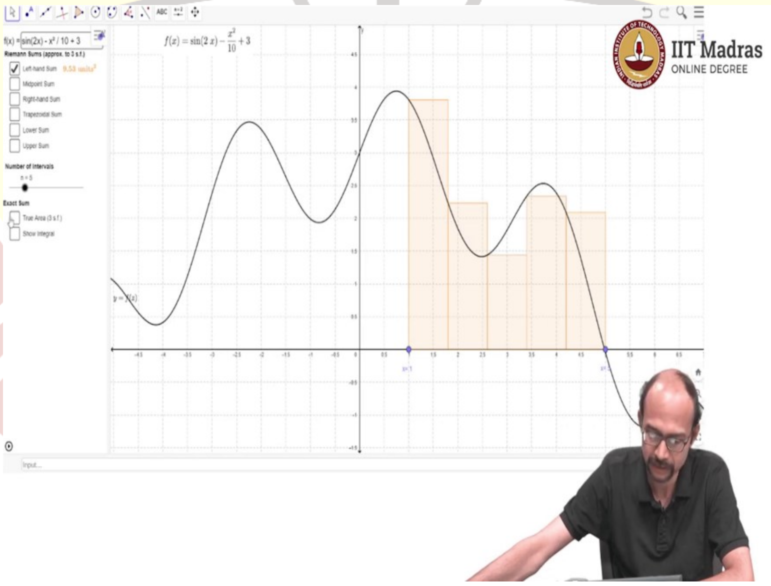
The Riemann sum of f with respect to the above data is defined as, so you have your data P which consists of a partition and a choice of point corresponding to each small interval in that partition, then we define the Riemann sum corresponding to this P. So, S(P) = $\sum_{i=1}^{n} f(x_i^*)\ \Delta x_i$, what is going on here? So, let us go back to our picture over there.

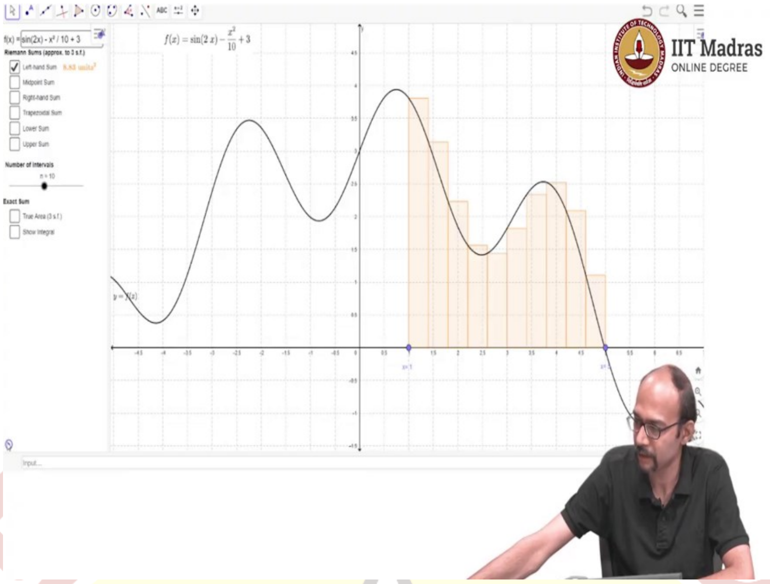
So, suppose I choose my x_i^* to be the first point of that interval. Suppose $x_i^* = x_{i-1}$. x_i^* has to be between x_{i-1} and x_i, so I choose x_i^* to be x_{i-1}, then what do I get? I get, f(x_0) $(x_1 - x_0)$. What is that? That is exactly the area of the first rectangle over here. That is the area over here, this area. That is what is f(x_0) $(x_1 - x_0)$. What is the second one? So, the second one is f(x_1) $(x_2 - x_1)$ that is exactly the second area. This is the second area.

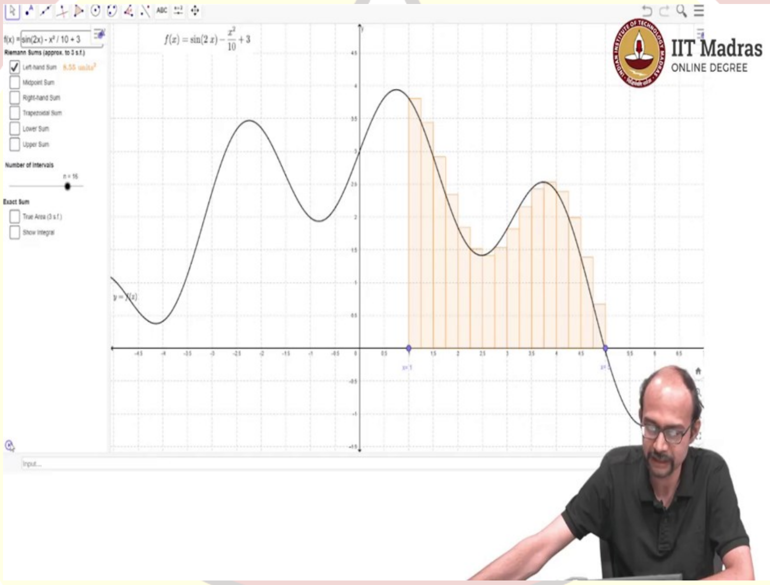
So, what we are doing is, if we had chosen x_i^* to be x_{i-1}, and we had that, that P over there, then the Riemann sum is exactly the area that we have the sum of the rectangles, so what this is doing is it is saying you take you take your interval $[x_{i-1}, x_i]$ over that you take the rectangle of height or breadth $f(x_i^*)$. And then when you multiply the length and the breadth, so the length is Δx_i, the breadth is $f(x_i^*)$, you get the area of that rectangle, and then you sum up all these areas, so that is an approximation for your, the area under the graph.

That idea of what is going on. So, that is a Riemann sum and of course, what we would like is that we keep refining our partitions. So, this P is going to now start changing. So, we will take smaller and smaller P's, and that will give us something interesting. So, let us go to GeoGebra and do some computations.

(Refer Slide Time: 10:34)







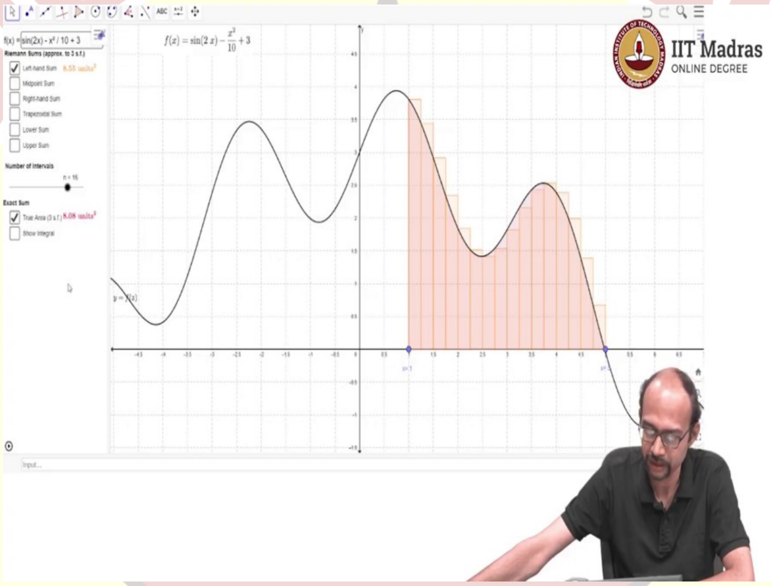
So, I got this from the GeoGebra website. So, here is a function, this is the function $\sin(2x) - \frac{x^2}{10} + 3$, somewhat complicated looking function. And let us plot all these various sums that we have. So, these are the Riemann sums corresponding to different choices of x_i^*. So, here is the left-hand sum. So, we want to approximate, we want to compute this in [1,5]. So, between 1 and 5, let us say we have the number of intervals is 5, and we are taking equidistant intervals.

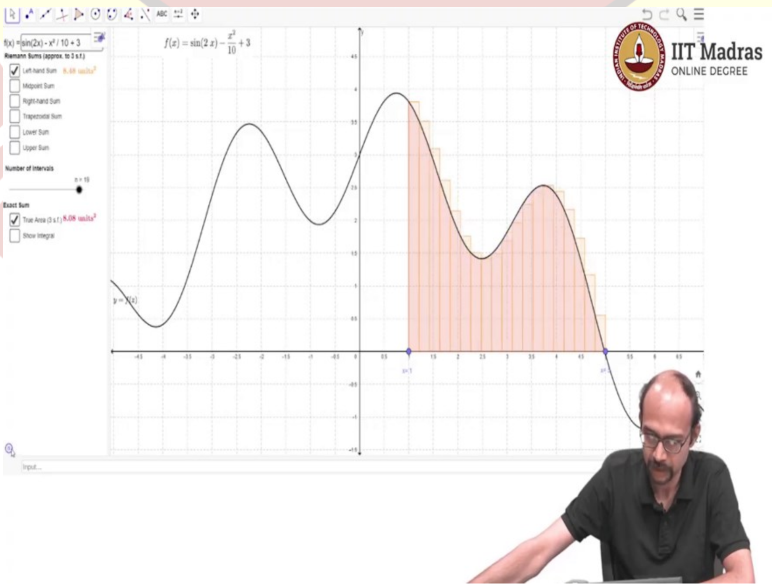
So, that means Δx_i is $\frac{5-1}{5}$, so $\frac{4}{5}$. So, here your first, that is 0.8. So, your first interval is $[1, 1.8]$. The second interval is $[1.8 , 2.6]$, then $[2.6 , 3.4]$, and so on. And now, what we have here is our left-hand side, what that means is, x_i^* is the left endpoint of your interval. So, $x_i^* = x_{i-1}$, which is what we had in our previous example, for e^{-x}.

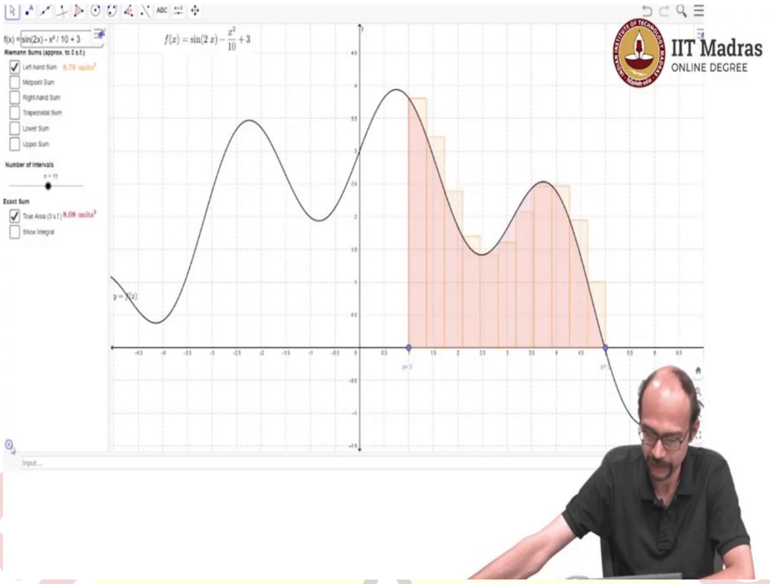
So, in that case, this is a picture we get. And you can see here, the value of the left hand side it is 9.53 units. So this is the Riemann sum corresponding to $x_i^* = x_{i-1}$, and $x_i = 1 + (i \times 0.8)$. And there are 5 intervals. And now we can play this animation, if we play this animation, then the number of intervals increases. And you can see what happens.

Your intervals shrink, the number of rectangles increase, and slowly, you are starting to approximate the area under the graph better than the previous one. So, this is now 8.83 units, you can see it is changing, the area is changing. So, the area is occurring in this panel on the left, where you have the left hand sum.

(Refer Slide Time: 12:56)

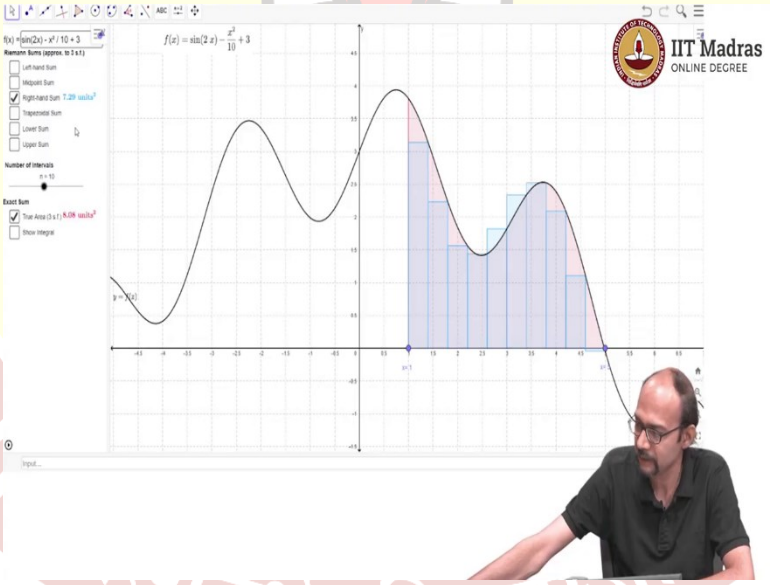






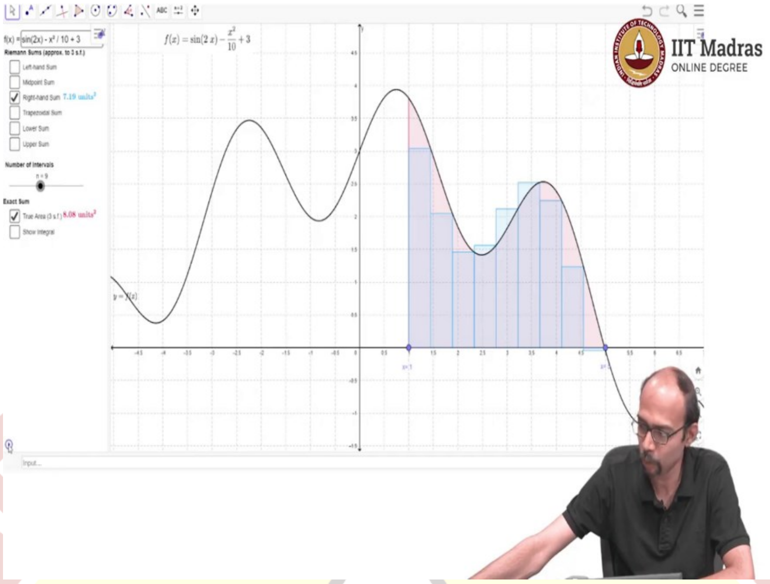
And what is the true area? it is 8.08 square units. And as you go on, it is going to get better and better. Of course, it may not get better immediately, but in the long run, you would expect that it gets better. So, this is with a choice of $x_i^* = x_{i-1}$.

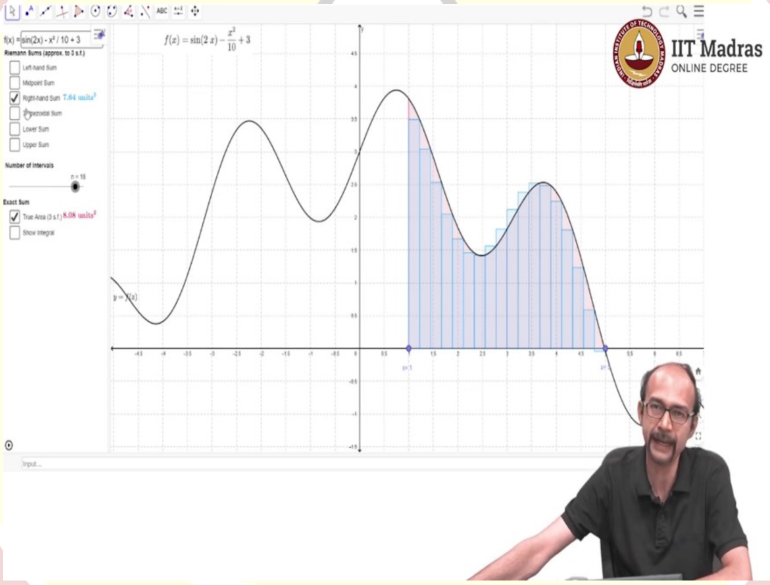
(Refer Slide Time: 13:18)



Instead, you could choose $x_i^* = x_{i+1}$. In that case, you get a different set of intervals. Rather sorry, you get a different set of rectangles. And this is how those rectangles look like. And now, if we play this, let us see what happens. So, let us start with just 1 rectangle. So, you can see if you have one rectangle, this area is actually negative, so it is −0.176 square units, very far from the actual value.

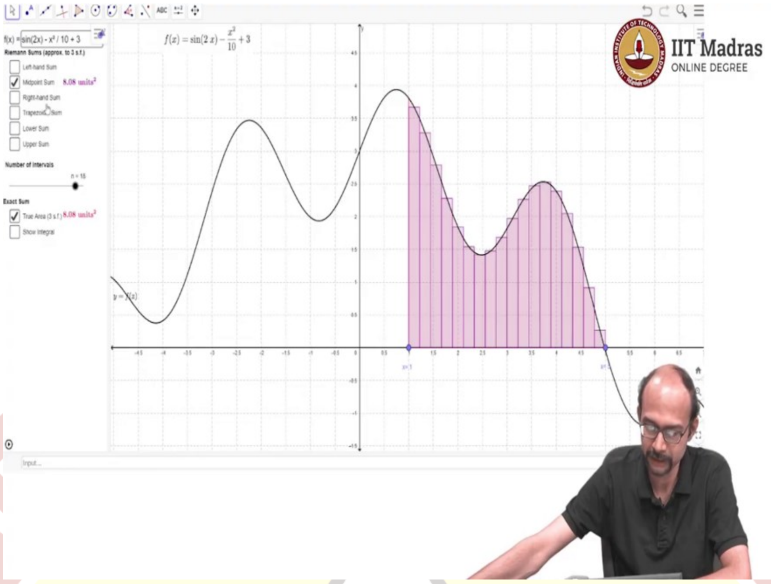
(Refer Slide Time: 13:54)





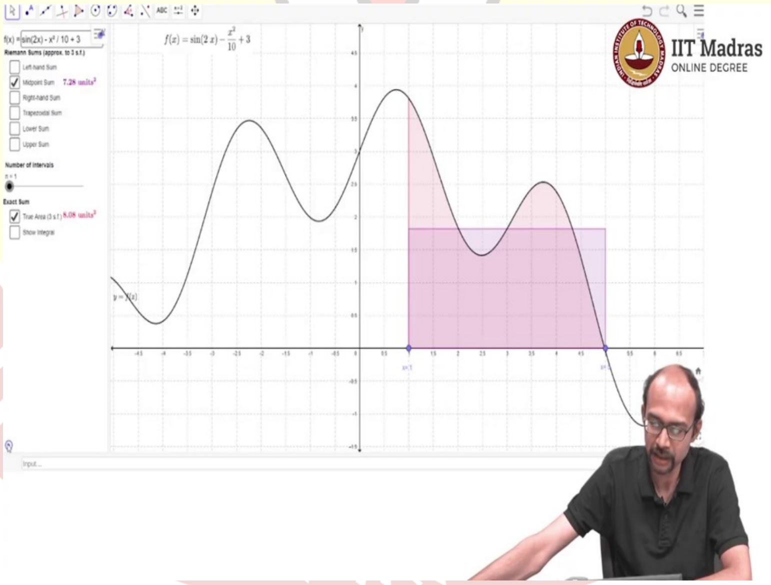
But now as you do better meaning you start using more and more rectangles, which is to say finer and finer intervals, you can see it starts becoming better and better so it is 7.42, 7.47, 7.6 and slowly it is going to come closer and closer. So, if I played this till say n =500, you would have seen it is very, very, very close. So, here x_i^* was chosen to be x_i. So, remember x_i^* was any point between x_{i-1} and x_i. It was a particular choice, here we had made, we made the choice $x_i^* = x_i$, so the right-hand sum.

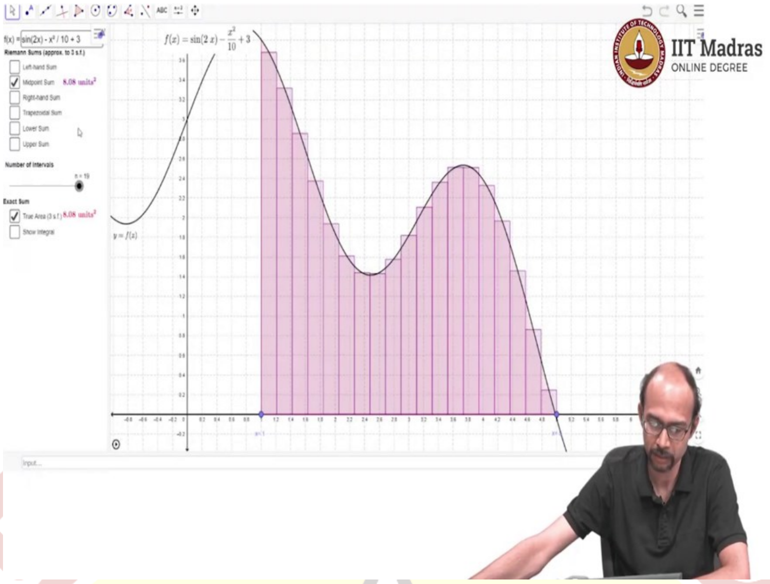
(Refer Slide Time: 14:38)



Instead, you could have the midpoint sum let us say. And if you take the midpoint sum, what do we mean by midpoint some, we mean, you take $x_i^* = \frac{x_{i-1}+x_i}{2}$. So, if you do that, let us play what happens.

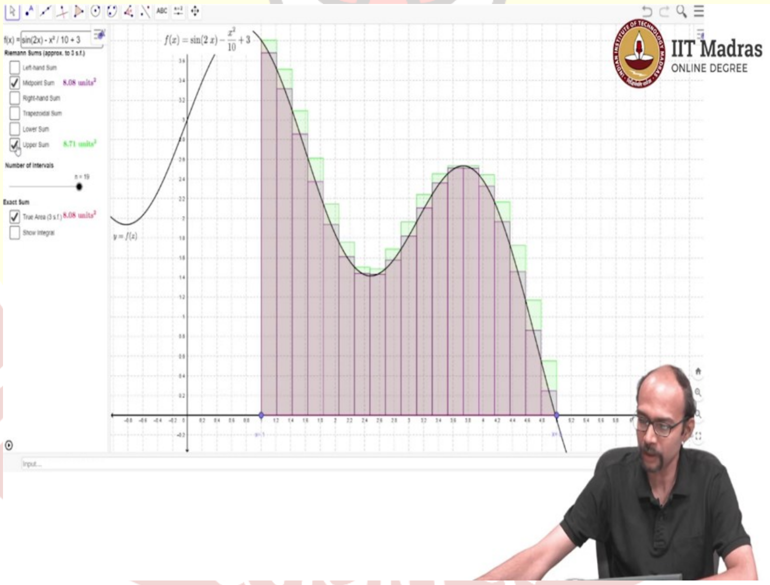
(Refer Slide Time: 14:55)

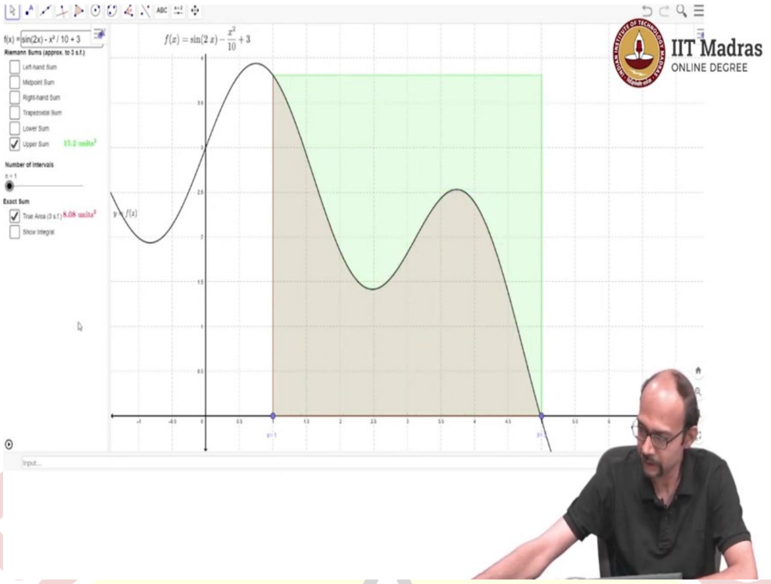


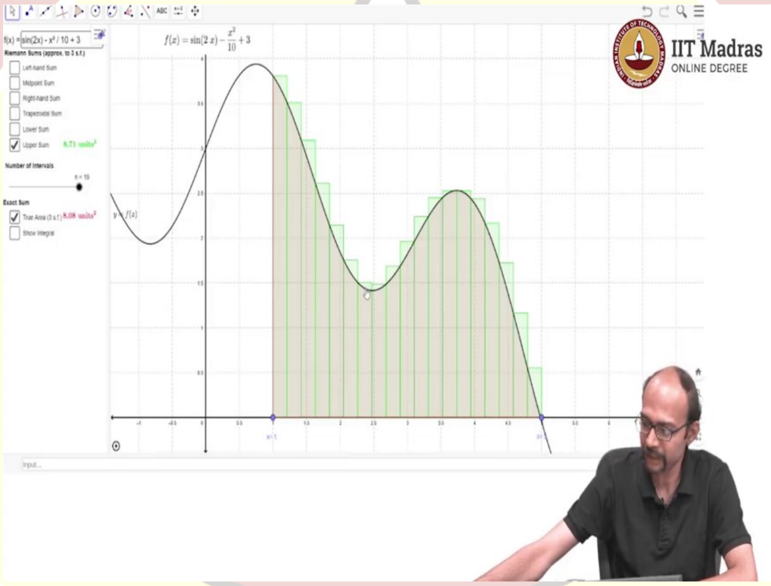


So, right at the start, it is it is fairly close to 7.28 square units. And you can see slowly, it is better and better approximating. So, it is now 8.09 square units, 8.08 square units, which is precisely very, and you can see, it is a really good approximation if you look at the picture. So, that is the midpoint sum.

(Refer Slide Time: 15:20)





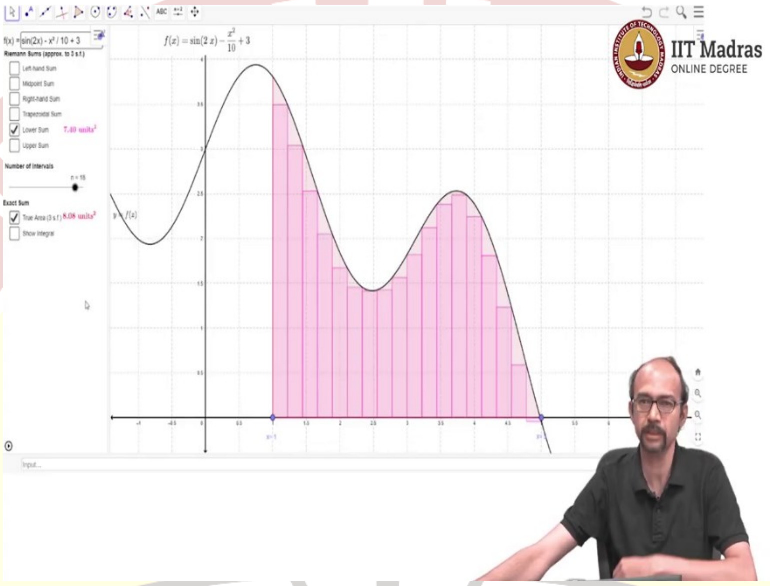


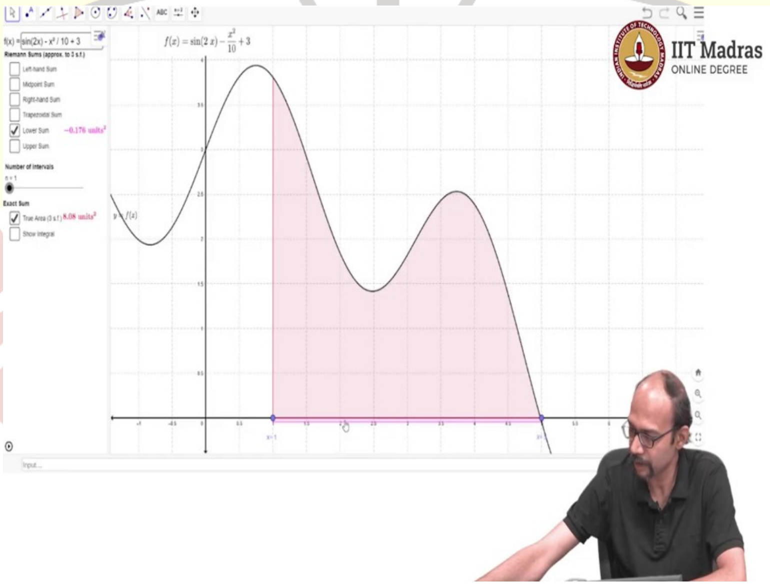
Now, I have defined the Riemann sum with the data, which is you choose an interval and a point in that interval. Instead, sometimes, you could choose a slightly different set of data you could choose, you choose your partition, but then instead of choosing x_i^*, you could choose for example, say the largest value between of f(x) between x_{i-1} and x_i. So, if you choose that, that is called the upper sum. So, that is always going to overestimate your area. So, here is the upper sum. So, as you can see, the chosen number to multiply by is the maximum of f(x) within that interval.

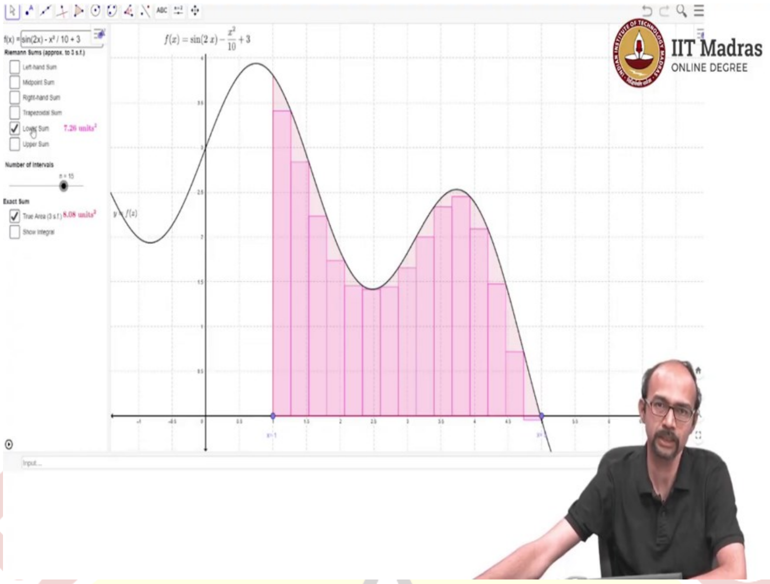
And if you play this in the initial stage, it is really large. But slowly, it is better and better, going to get approximated better and better. And it is an upper bound remember, because it is always going to be larger, but in the limit, it will come closer and closer. And that is, that is what we saw. And this idea is very similar to the idea that we saw for the way Archimedes got to the area of the circle, so he got over estimates based on polygons.

And then he got under estimates based on polygons and then he said it should be between these two, but in the limit, both of them are πr^2. Hence, it must be πr^2. So, you can overestimate by the upper sum, then try to take the limit.

(Refer Slide Time: 16:58)



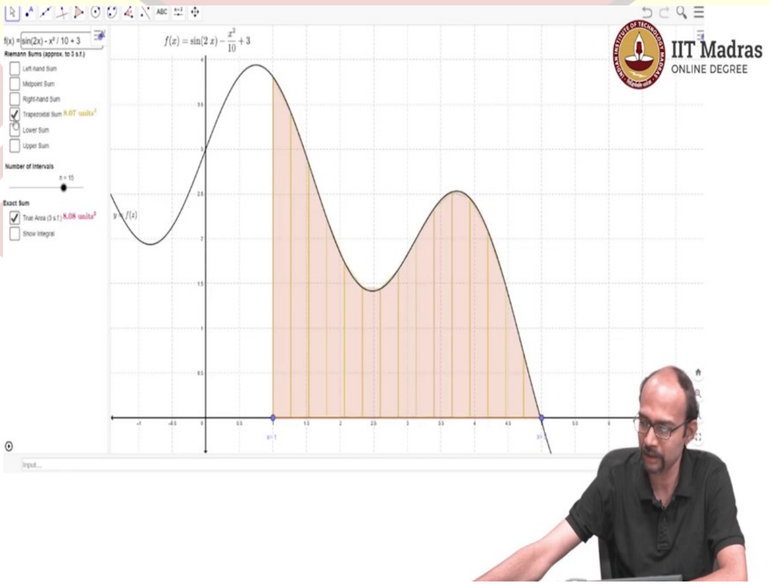


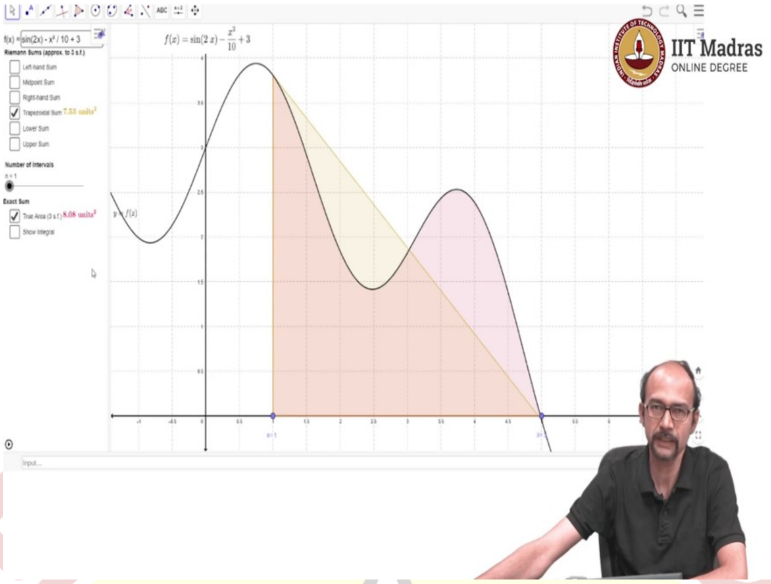


Or you can underestimate by the lower sum. So here, instead of taking maximum, you take the minimum. And again, let us play this and see what happens. So, when you take the minimum, the starting place is you are in really bad shape, because the minimum is actually a negative number. But if you keep going, then slowly it is going to get better and better and you can see the numbers here are 6.98, 7.05.

The picture tells you already what is happening. So, this is a under estimate for the area. And the true area is going to be between the upper sum and the lower sum. So, in the limit if upper sum and lower sum go to the same value, then that is going to be the area.

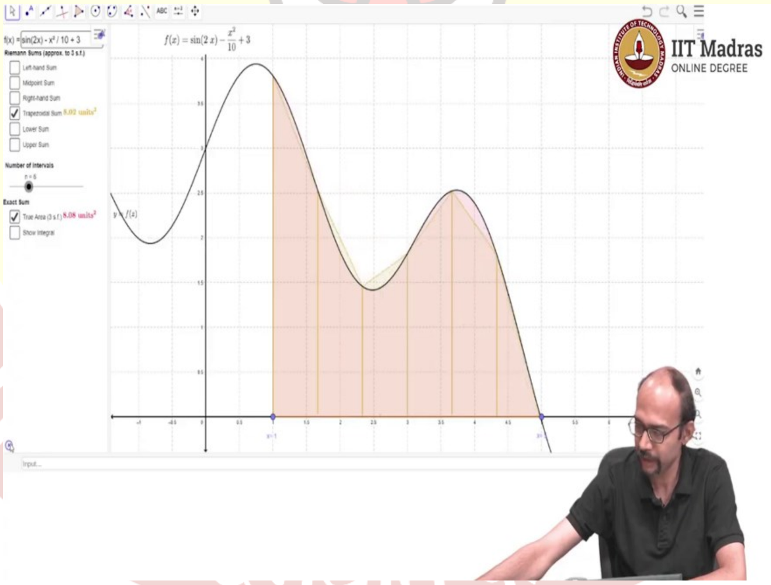
(Refer Slide Time: 17:38)

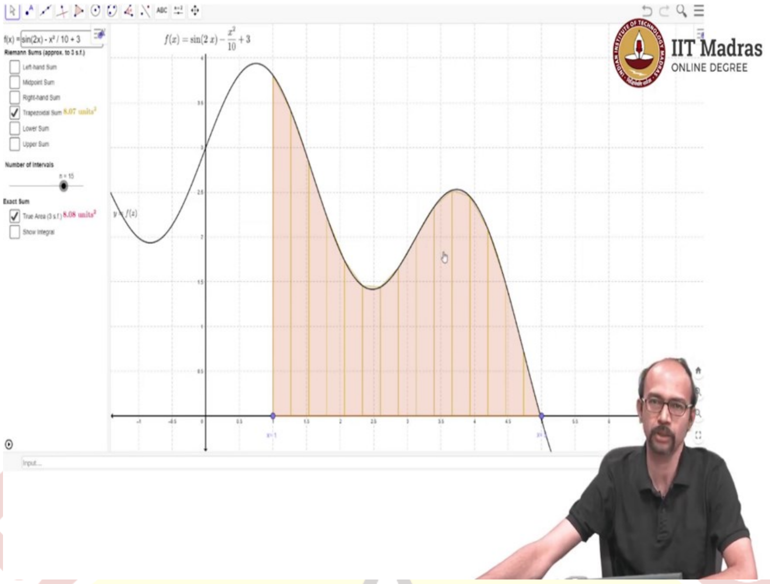




And then one last example of another thing you could do, you could use something called the trapezoidal rule. So, you may have sum, you may have seen this at some point. In school, we have something called a trapezoidal rule to estimate areas. And the idea is exactly what we have here. So, instead of taking rectangles, you take trapeziums and here is how you do that.

(Refer Slide Time: 18:06)





So here is now various trapezium, so what you do is you have x_{i-1} and x_i. And now you join $f(x_{i-1})$ to $f(x_i)$. So it is a right angled trapezium at the base. And then at the top, it is like this. And there is an important takeaway for why we want to do this. You see, we could, instead of doing this, we could take a line slightly above, and we could use something called the tangent. So, and if he did that, well, that would also be an approximation and then that may give us something interesting.

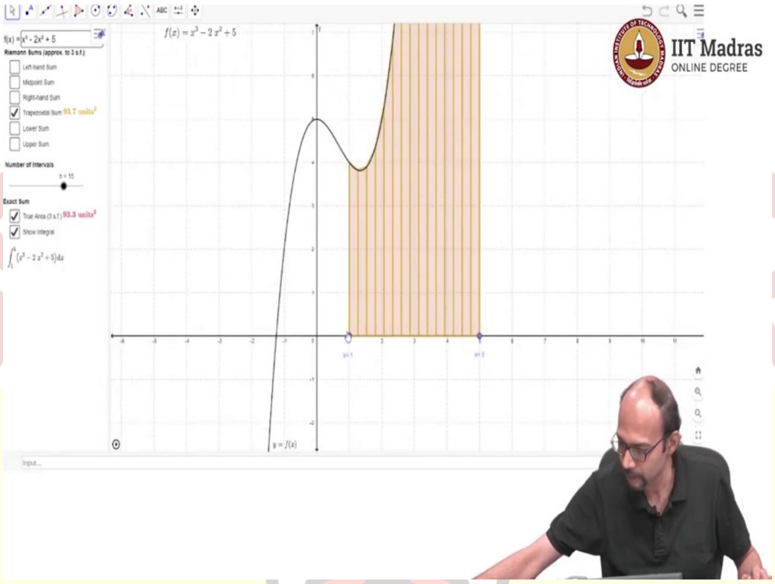
So, keep that idea in mind, it is going to come up in the next video. So, coming back to the trapezoidal rule, you are estimating your areas using trapeziums. So, this is actually a much better approximation as compared to the others. And if we play this further, you can see it very close 8.06, 8.07, 8.07 and it is going to remain over there. So, fairly close to 8.08. And you can see in the picture also that it is really well approximated.

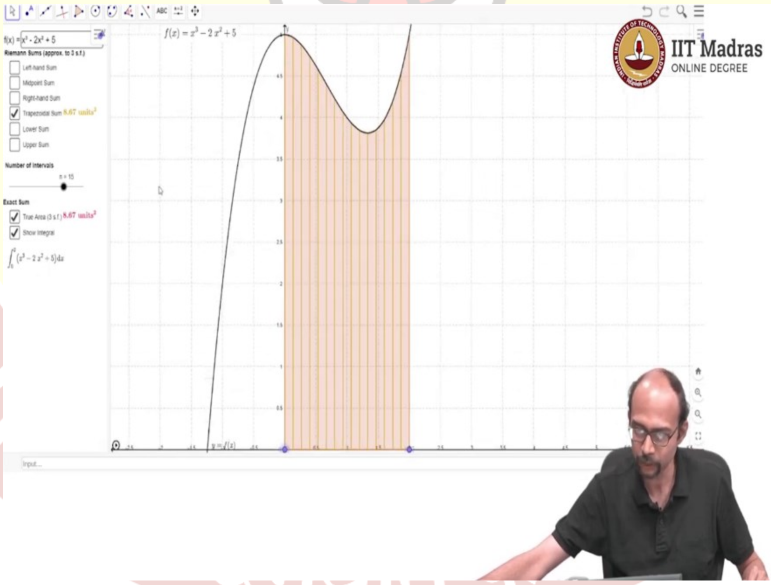
So, I hope the idea of Riemann sums is clear. So, we are going to use Riemann sums based on the rectangles, but you could have other kinds of Riemann sums. And it will not matter if your function is what is called Riemann integrable, it does not matter really how what kind of Riemann sum you take. In the limit, they all give you the same thing, which is the area fine.

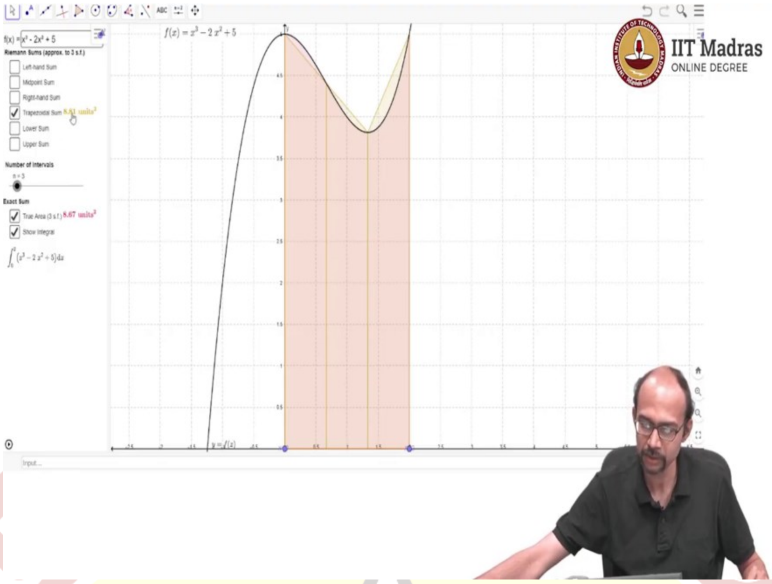
Now I used several words there. I used the word integrable, I use our integral already. And the title of the video had the word integral. So, let us go to what the integral is. Before that, let us for a second and ask two questions; one is, well over here, I said this is the true area, So, that in itself is a question. So, we will, of course aim to answer that question shortly, not this video, but in the next video.

And then there are something here called show integral. So, here is what the integral is. And that is what we are going to see in the next slide. Maybe before going there, let us change the function. So, just in case you think I am doing this for some special function. Let us change the function to some other function that that you might feel is more interesting.

(Refer Slide Time: 20:50)

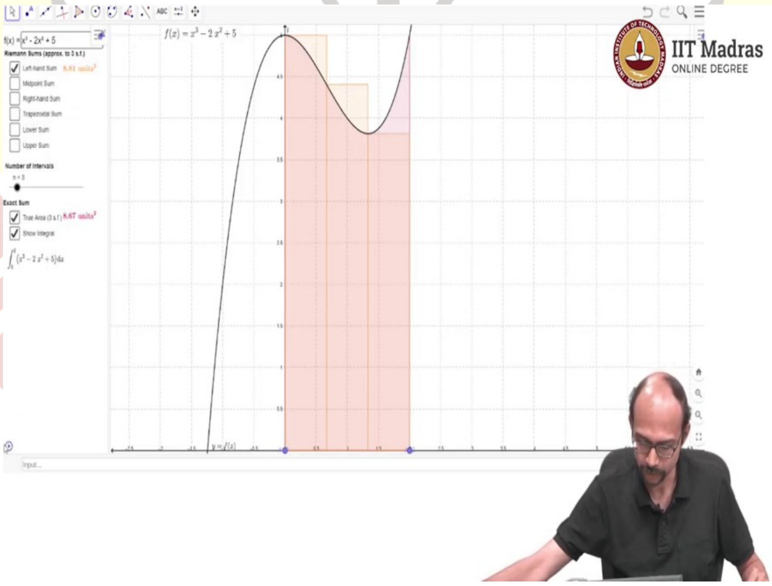


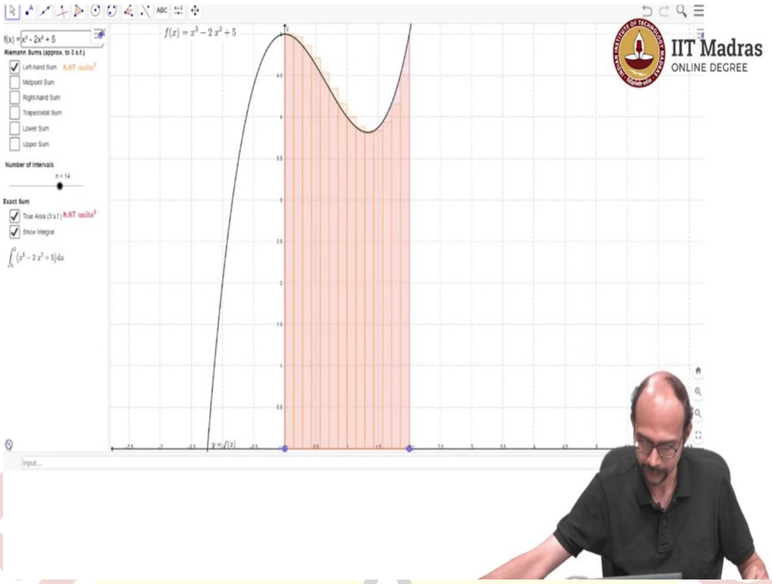




So, here is your function. And let us change our x values so that we have this in the region where we can actually see what is happening. So, so suppose I want to do it between $0 \leq x \leq 2$, I want to compute this. So, now we can see our various sums. So, this is the trapezoidal sum, you can see it is very, very closely approximating the area, even with 15 many things. So, here is with n=3. So, even for 3, the trapezoidal sum is very close. True area is 8.67, the trapezoidal sum is 8.81.

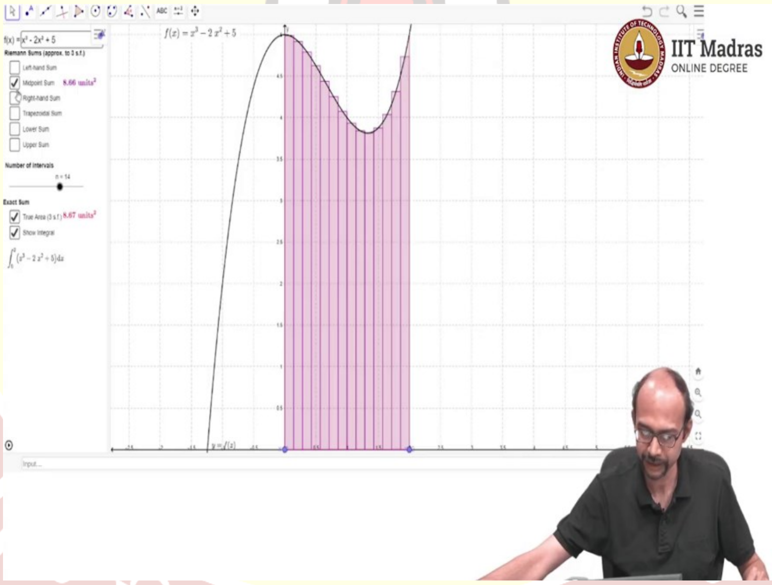
(Refer Slide Time: 21:34)

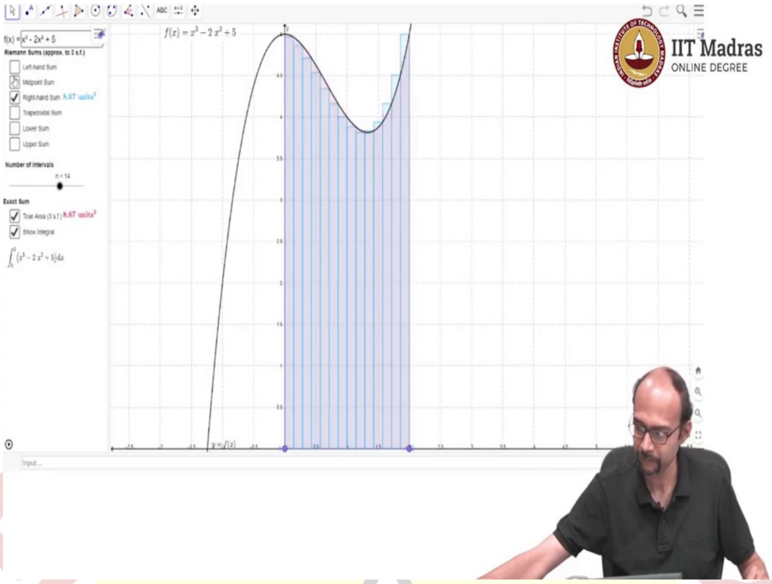


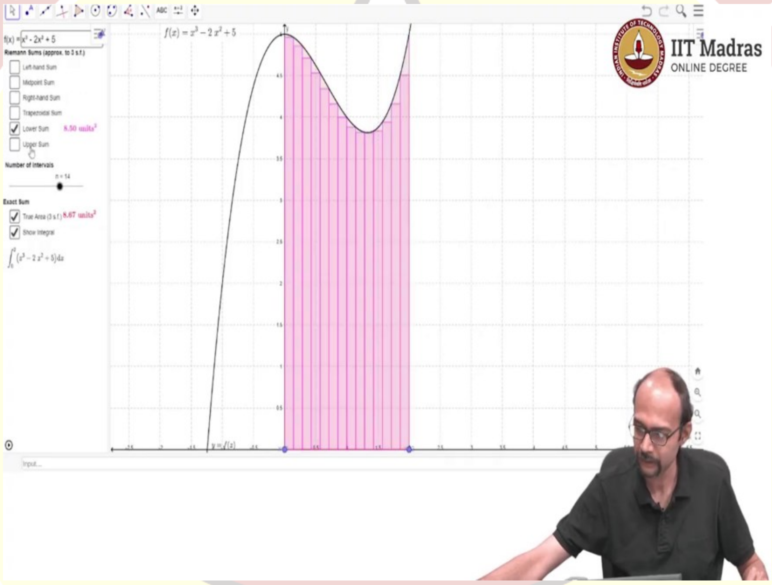


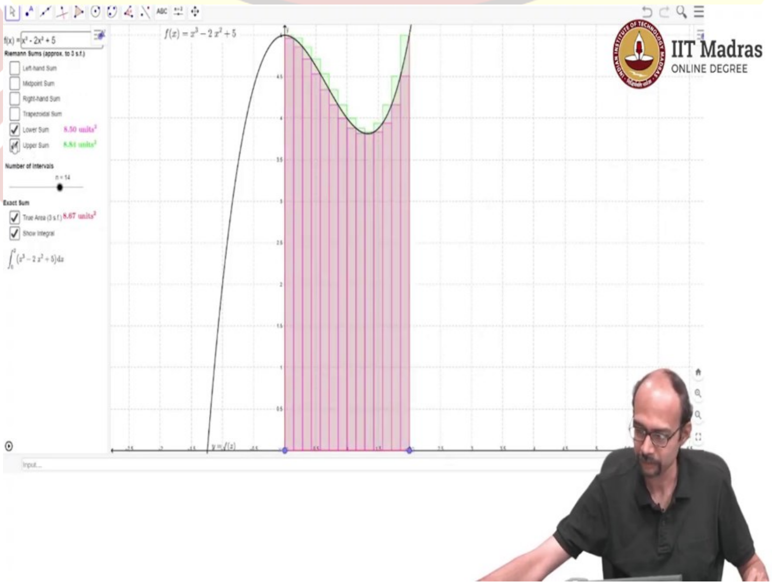
Let us do the left-hand sum instead. So, here is the left hand some still fairly close. And if you play this, you can see it is getting better and better and better. And it is quite close already.

(Refer Slide Time: 21:49)



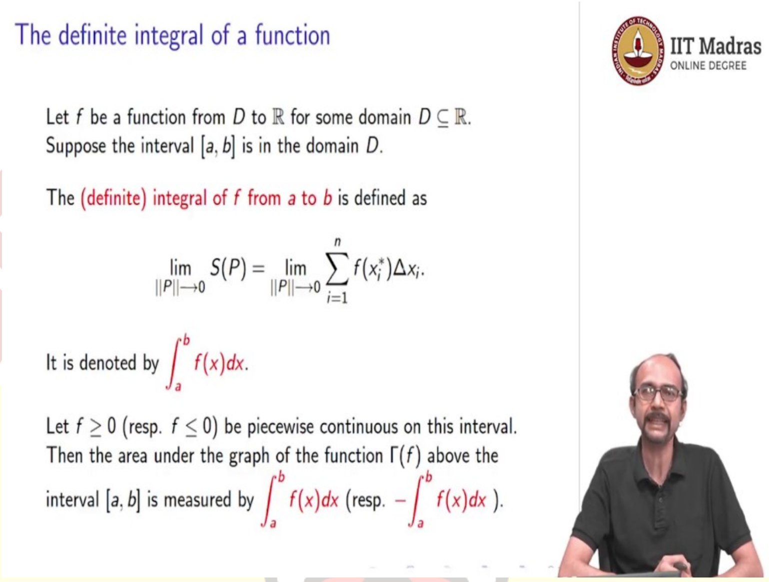

$f(x) = x^3 - 2x^2 + 5$
IIT Madras
ONLINE DEGREE

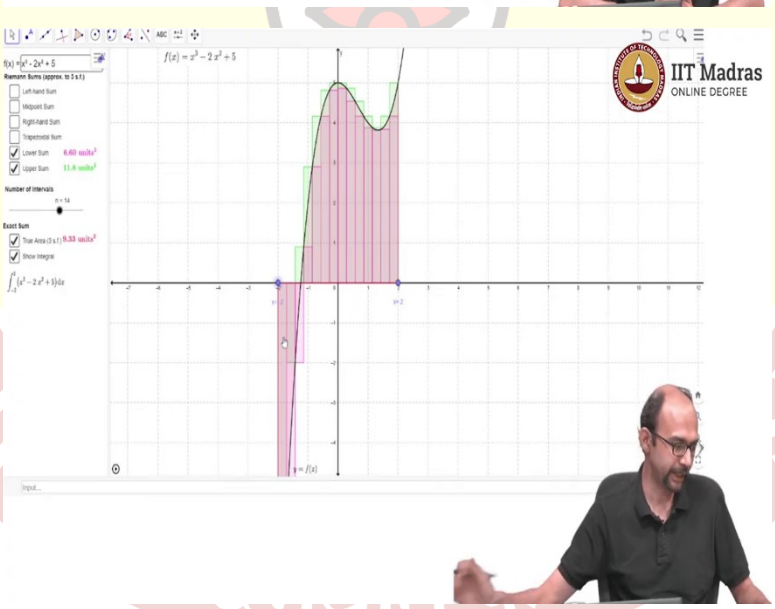

$f(x) = x^3 - 2x^2 + 5$
IIT Madras
ONLINE DEGREE


$f(x) = x^3 - 2x^2 + 5$
IIT Madras
ONLINE DEGREE

And here is your midpoint sum, again quite close. Here is the right hand sum. And here, your lower and upper sums. I highly encourage you to go to GeoGebra and play around with the Riemann sums to get a hang of what you are going to see next. So, let us go to the next slide, which will tell us what is the integral of a function.

(Refer Slide Time: 22:14)





Now, the title says the definite integral. The reason it says definite integral is because there is something else called the indefinite integral, which we are going to see in the next video. So, what is the definite integral, so let f be a function from D to $\mathbb{R}$ for some domain D, suppose the interval $[a, b]$ is in the domain D, same as what we had in the previous slide. The definite integral of f from a to b. So, if I say from a to b, I need not use the word definite integral.

But if I if I do not use from a to b, then I better say definite integral to indicate that it is over some interval. So, this is defined as you take the limit over these Riemann sums, and you allow your, the norm of your partition to shrink to 0. So, $\lim_{\|P\| \to 0} S(P)$, which is the same as $\lim_{\|P\| \to 0} \sum_{i=1}^{n} f(x_i^*)\ \Delta x_i$.

So, we have seen already this idea in the in animation we had, where the number of intervals were increasing. In other words, the interval sizes were decreasing and the interval sizes are exactly being kept track of by this $\|P\|$. So, this is denoted by $\int_a^b f(x)\ dx$, I want to warn you that this dx is very crucial. If you drop the dx, you may run into trouble somewhere else. So, I will qualify that when I get there in the next or the next to next video.

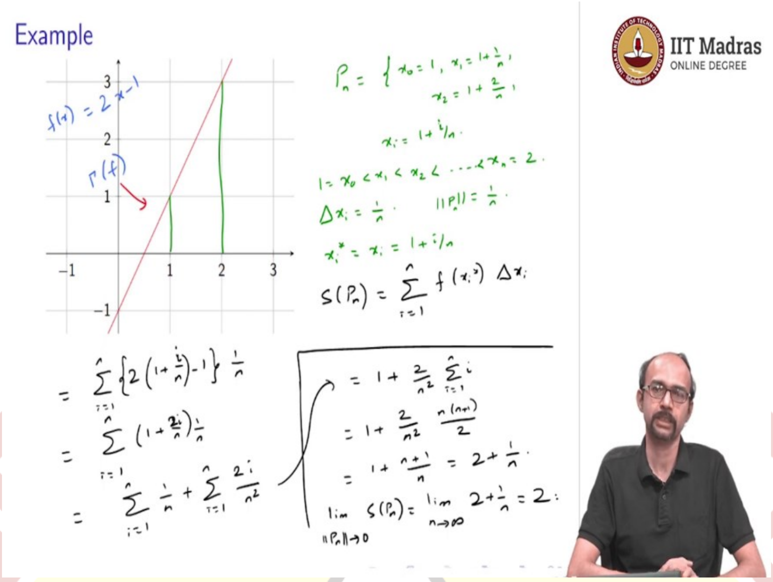
So, if f ≥ 0 and it is a piecewise continuous function on this interval $[a, b]$ then the area under the graph of that function above the interval $[a, b]$ is measured by this integral, $\int_a^b f(x)\ dx$. So, if f is piecewise continuous, which means, you allow some discontinuities, maybe at finitely many points. Then this integral, meaning the limit, indeed computes your area. That is what we have seen in the examples.

And if f ≤ 0, we have to qualify that as well, because remember that the definition that we have made never assume that f is positive or negative. it well for any function. So, if f is negative, what does this number tell you? Well, that is also the area, but it is the area above the graph below the interval $[a, b]$. Unfortunately, this number will then compute and something which is negative, if you because $f(x_i^*)$ will be negative, so the sum will be negative.

So, you have to qualify that with a minus. So, you have to put it, put a minus. So, if you take minus of the integral, that will indeed be the area. So, in other words, what the integral computes is signed area. So, if we go back to our example from here, and we let us say I choose a different interval, let us say I choose the interval here. So, x is, let us say −2 to 2.

And if I compute what is the integral, you can see it is computing this area minus this area, that is what it is computing. So, keep that in mind when you use an integral to compute area. Great. So, now we have a way to compute integrals or rather compute areas, based on integrals.

(Refer Slide Time: 25:41)



Let us do this example, just to fix ideas. So, this is the example this is an easy example. So, this is the graph of the function. So, this is $y = 2x - 1$, or $f(x) = 2x - 1$, is this line. So, this is the graph of f, the red thing here. And let us say I want to compute the area between 1 and 2. This is the simplest area I would want to do. In fact, I can read off from the figure here what the area is.

So, if I want to compute this area, let us first ask what is the area? Well, you have this part here. And you have this part here. And you can work out what the area is. So, you have a square of size 1, and then you have a triangle with base 1 and height 2. So, $2 \times 1 \times \frac{1}{2}$, so which is 1, so 1 + 1 is 2. So, this area is 2 square units.

Now let us try to do this from first principles using Riemann sums. So, let us take a set of partitions, I will call it P_n. So, what is P_n, P_n is the set of partitions $1 + \frac{i}{n}$. So, you have, so in other words, what I am saying is $x_0 = 1$, $x_1 = 1 + \frac{1}{n}$, $x_2 = 1 + \frac{2}{n}$ and so on. So, in general, you have $x_i = 1 + \frac{i}{n}$ and so you will have $x_0 = 1$, or rather $1 = x_0 < x_1 < x_2 < x_3 \dots < x_n$, $x_n = 1 + \frac{n}{n}$.

So, this is $1 + \frac{i}{n}$, just in case that was not visible. So, this is $1 + \frac{n}{n}$, which is 2. So, I am dividing my interval into 'n' equal parts, the interval $[1, 2]$, I am dividing into 'n' equal parts. So, each part has size $\frac{1}{n}$, so $\Delta x_i = \frac{1}{n}$. And so ‖P‖ is also going to be $\frac{1}{n}$. And then we will allow n → ∞, so ‖P‖ will go to 0.

So, what do we choose as our x_i^*? Let us say I choose $x_i^* = 1 + \frac{i}{n}$. So, which is x_i. So, let us say we do the right-hand sums. So, let us write down what this what this Riemann sum is. So, the Riemann sum corresponding to P_n is going to be S(P) = $\sum_{i=1}^{n} f(x_i^*)\ \Delta x_i$. So, I will continue over here. So, what does this give us? this gives us $\sum_{i=1}^{n}\{2\ (1 + \frac{i}{n}) - 1\}\frac{1}{n}$.

Well, let us see what we get. So, if I simplify the things in the bracket, I have $2 + \frac{2i}{n} - 1$. So, this is $1 + \frac{2i}{n}$. So, I have two sums here, so $\sum_{i=1}^{n}\frac{1}{n} + \sum_{i=1}^{n}\frac{2i}{n^2}$.

Now, I think you can see this and feel some more familiar, let me continue over here. So, the first sum is $\sum_{i=1}^{n}\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} 1 = \frac{1}{n} \times n = 1$, so the first sum gives you 1.

The second sum $\sum_{i=1}^{n}\frac{2i}{n^2}$, $\frac{2}{n^2}$ is constant. And what you get is $\frac{2}{n^2}\ \sum_{i=1}^{n} i$. But this is something we know, so $\sum_{i=1}^{n} i$, this is a formula that you may have learned before, this is $\frac{n\times(n+1)}{2}$. And so what we get is $\frac{2}{n^2} \times\ \frac{n\times(n+1)}{2} = 1 + \frac{1}{n}$. So total sum is $1 + 1 + \frac{1}{n} = 2 + \frac{1}{n}$.

So, this is the Riemann sum corresponding to this partition P_n. And now of course, if we let n tend to ∞, which is the same as saying allow $\|P_n\|$ to go to 0, it is the same as letting n tend to ∞, $2 + \frac{1}{n}$ and that is indeed 2. That is $\lim_{\|P_n\|\to 0} S(P) = \lim_{n\to\infty} 2 + \frac{1}{n}$. So so we had already computed by observation that this was 2 square units. And this rather complicated computation gave us the same thing.

So, I hope, both the idea of what is the Riemann sum, what is the integral, which is defined as this limit and the fact that it matches with what our intuition is, is clear. Now, the main problem as you can, you might be able to see is even for a simple function like linear function, this was a fairly non-trivial computation to compute actually what this limit is and find it.

So, if you have a difficult function, let us say like the functions, we were looking at say e^{-x} or some polynomial, like $x^3 - 2x^2 + 5$, if we try to use this method, it is going to be really difficult to compute what happens, you can but it is not going to be an easy process. So, we need a way to compute this integral in a more efficient manner. And indeed, that is what we will see in the next video. Thank you.