IIT Madras

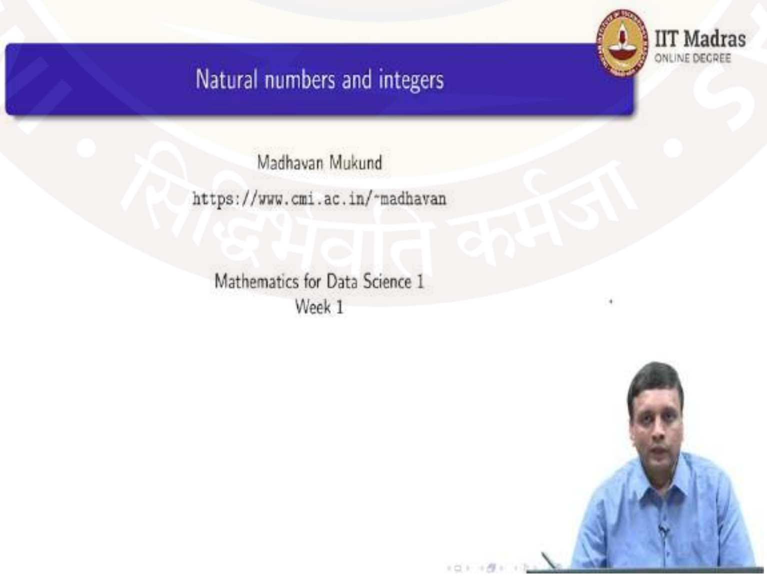
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture – 01
Natural Numbers and their operations

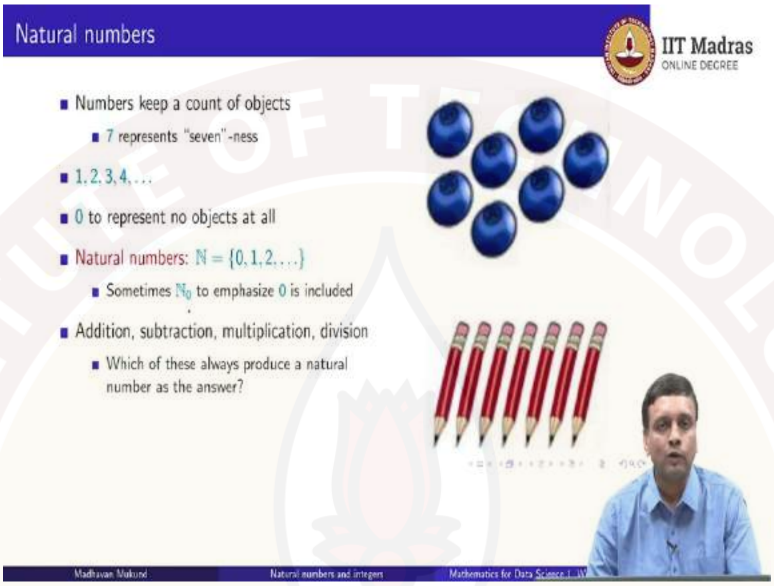
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, welcome to the 1st week of Mathematics 1 for Data Science. So, we are going to start with some very basic things which you probably know; right from the beginning we are going to start talking about numbers. So, in this 1st module what we are going to talk about is natural numbers and integers.

(Refer Slide Time: 00:30)



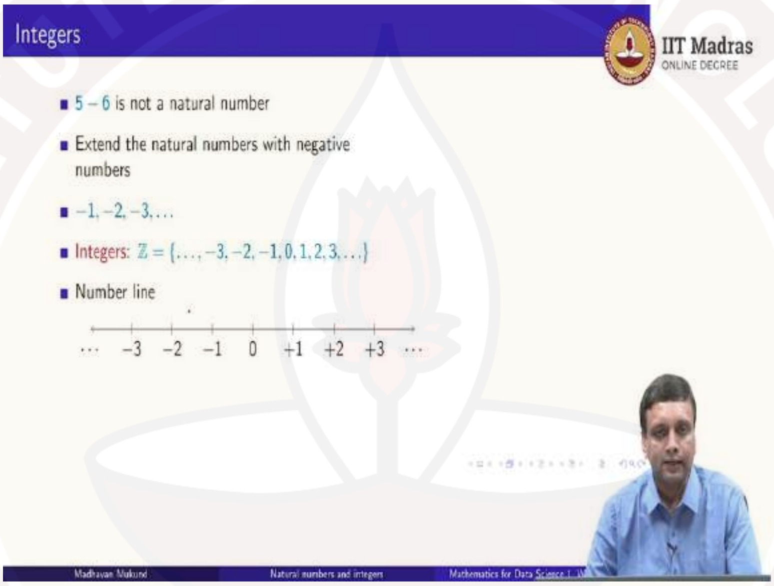
So, as you probably remember from as young as you were in school when you first came across numbers, we use numbers mainly for counting. So, for instance if we see 7 balls like this and then we see 7 pencils like this, then we need to know that these are the same number of things and for this we use this number 7. So, 7 represents what is common to these two objects that there are 7 balls and 7 pencils. So, 7 is an abstract concept in that sense and it refers to a quantity.

So, we all of course, know the numbers 1, 2, 3, 4 and all that. So, when we see a number of things, we can count them. But perhaps the most important number of all which is of Indian origin is 0. So, it is quite important to have a way to represent something when there is nothing to count because without a 0, we cannot use our place numbering system that we use to manipulate numbers.

So, these numbers starting with 0 are what are often called the natural numbers. Now there is some confusion in some books and many books will actually use only 1, 2, 3, 4 to represent the natural numbers. So, we use N to represent the set of natural numbers and in case there is any confusion whether 0 is included in this set or not, now sometimes people will not include 0 in the set of natural numbers.

So, sometimes to emphasize that we are using 0, we will actually put the subscript 0 below the N right. So, we will write either N or N_0, but whenever we are talking about natural numbers, it always includes a 0. Now what can we do with natural numbers? Well we can add them, we can subtract them, we can multiply them, we can divide them. So, these are the normal arithmetic operations which you have studied in school.

(Refer Slide Time: 02:43)



But what is really interesting from a mathematics perspective is, when we take natural numbers and we perform an operation on them, do we always get a natural number? So, if we add two natural numbers, do we get a natural number? If we subtract a number from another, we get a natural number? If we multiply them, do we get a natural number? If we divide one by another, do we get a natural number?

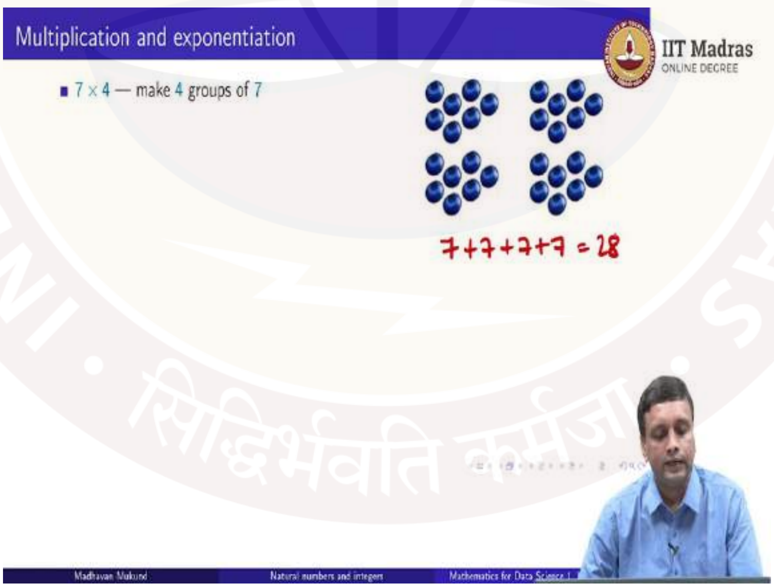
So, the first operation which fails this test is subtraction because if we subtract a larger number from a smaller number, so supposing we take 6 and subtract it from 5; then we go below 0 right. If you have 5 things and we take away 6 things, we will be cannot take away 6 things that is what subtraction means. So, we need to expand the scope of our numbers to allow these operations to work sensibly and this is how we get the negative numbers.

So, we had the positive numbers 0, 1, 2 the non-negative technically because 0 is neither positive nor negative. So, we had the positive numbers 1, 2, 3, 4. We added a 0 to account for the fact that we are counting nothing and now we add symmetrically on the other side negative numbers -1, -2, -3. So, this is just to illustrate why we get them of course, this is something that you should know from school.

So, this set which is the natural numbers extended with a negative numbers is what we call the integers and we use Z to indicate the set of integers. So, we have N the set of natural numbers which starts at 0 and goes forward 0, 1, 2, 3, 4 and we have the integers which start at no at minus infinity and go to plus infinity. So, these are both infinite sets, but the natural numbers have a starting point 0 and the integers extend to infinity in both directions.

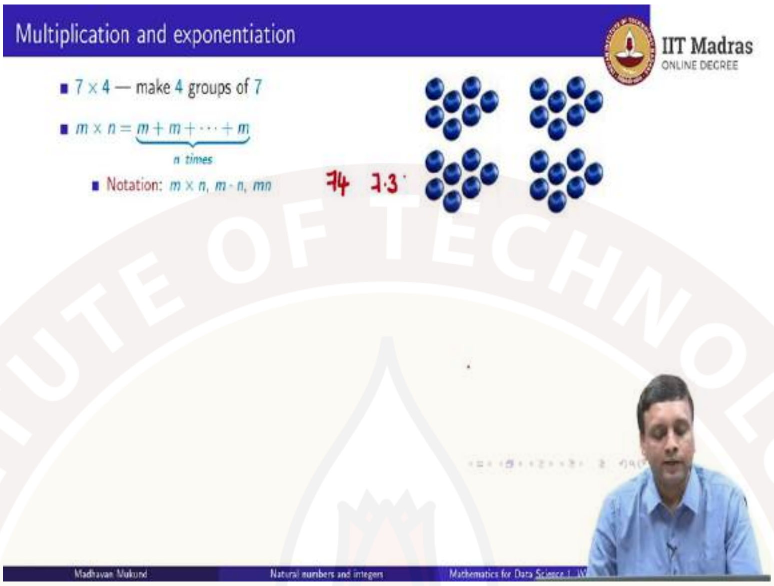
So, it is very convenient mentally to think of the integers as forming this kind of a sequence where on the left you have the very small ones and on the right you have the very long ones and this is normally called the number line. So, as you go from left to right, the numbers are increasing and this is how the integers are arranged.

(Refer Slide Time: 04:27)



So, we said that subtraction takes us away from natural numbers and we brought the integers. So, now, let us look at the other two operations that we talked about multiplication and division. So, let us start with multiplication. So, when we say 7 × 4 what we are really saying is take 7 objects and make 4 copies of them. So, for instance on the right, we have those 7 balls that we started with and then we have made 4 copies of them. So, if we want to know how many balls are here, then we have 7 from the first group, 7 from the second group and so on. So, we have 4 groups of 7 and this is if we add it up going to be 28.

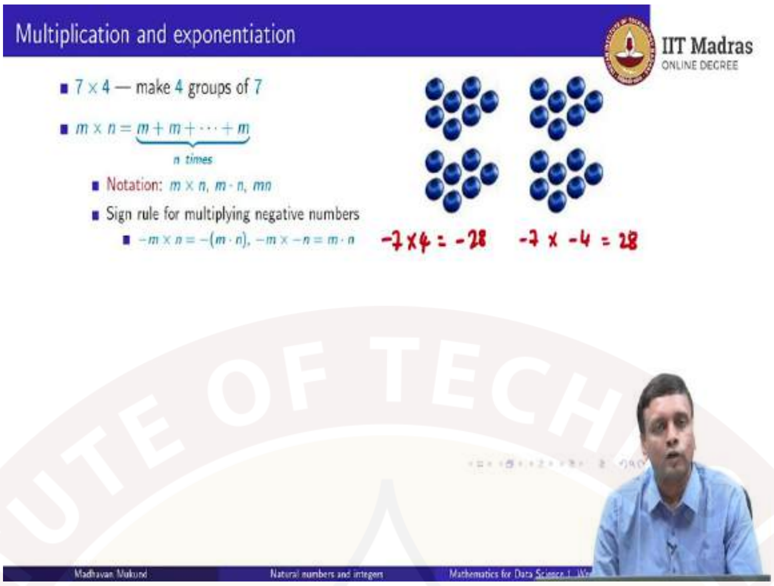
(Refer Slide Time: 05:02)



So, in general this is how we multiply when we take a number m and multiply it by n, what we are doing is we are making n copies of m. So, we are taking m + m + m… n times. So, in this sense multiplication is repeated addition.

So, we often use this time sign the × sign for multiplication, but this is often cumbersome when we write out equations. So, sometimes we replace this time sign by a . and sometimes we write nothing at all. So, if we just write two symbols together, we do not write this normally for numbers because imagine that if I write 7 4 like this, then you do not know whether it is a number 74 or its 7 × 4. So, if we have numbers, we will normally write a dot explicitly between them like 7 × 3. But when we have a names like m or n standing for numbers, then if we write mn; we assume that it is one number m multiplied by another number n.

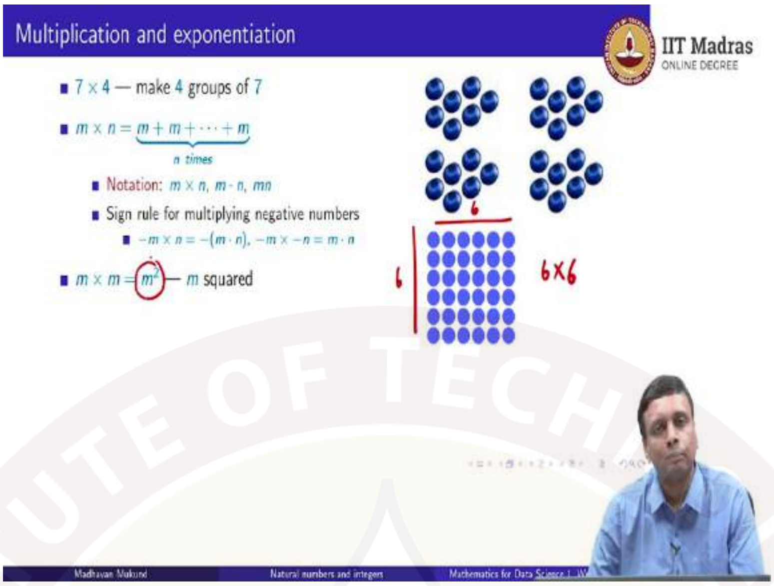
(Refer Slide Time: 05:56)



Now, we have integers, an integers have signs they are positive and negative numbers. So, we have to remember that when we multiply numbers with signs, the resulting number also has a sign and there is a sign rule which basically says that if we have one negative number multiplied by one positive number, then the result is a negative number. So, let us assume that m is a positive number so, - m is a negative number. So, say - 7 × 4 would be - 28. On the other hand if I take (- 7) × (- 4), then the two negations will cancel, and I will get 28.

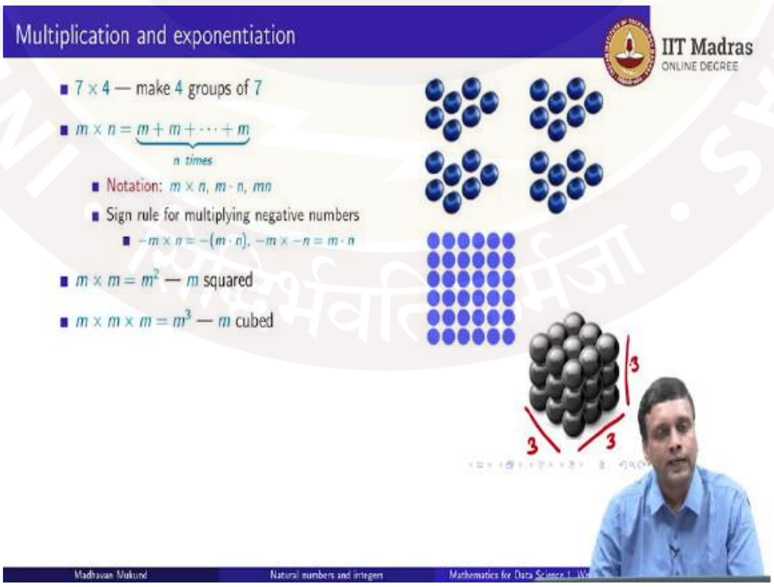
So, if you have an even number of minus signs, you get a positive number; if you have an odd number of minus signs, you get a negative number.

(Refer Slide Time: 06:40)



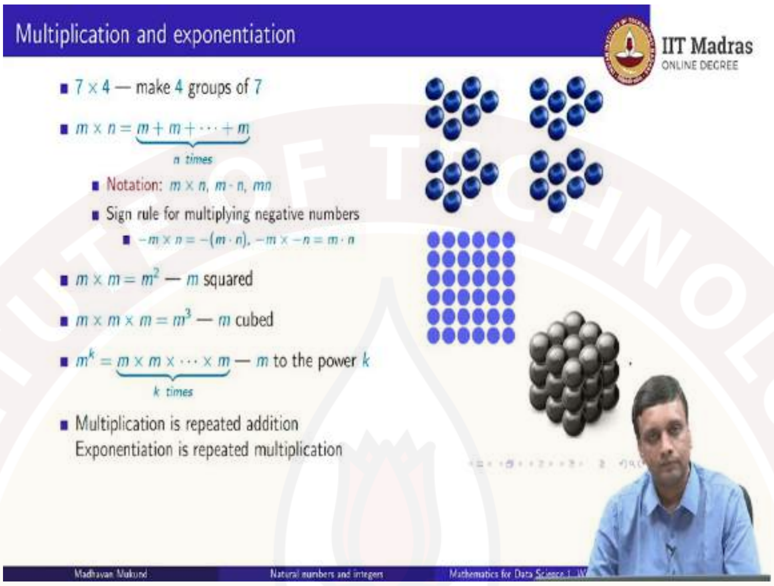
Now just like we have repeated addition, we can also do repeated multiplication. So, instead of doing m plus m, we can takes m times m and this is called m squared and the reason that it is called m squared is visible in the picture here. So, we have now here 6 balls and 6 balls. So, we have 6 × 6 right. So, this means that we can arrange these 6 times 6 balls in a square and this is why we call this square.

(Refer Slide Time: 07:15)



So, this notation m^2 stands to the fact that, m is multiplied by itself twice. Now if you multiply it by self 3 times, then we get a cube. So, here for instance we have 3 balls by 3 balls and then we have a height a stack of 3 such balls. So, we have a square of 3 by 3, 9 balls and we have 3 stacks of these one on top of the other. So, this naturally forms a cube so, m × m × m is written m^3.

(Refer Slide Time: 07:37)

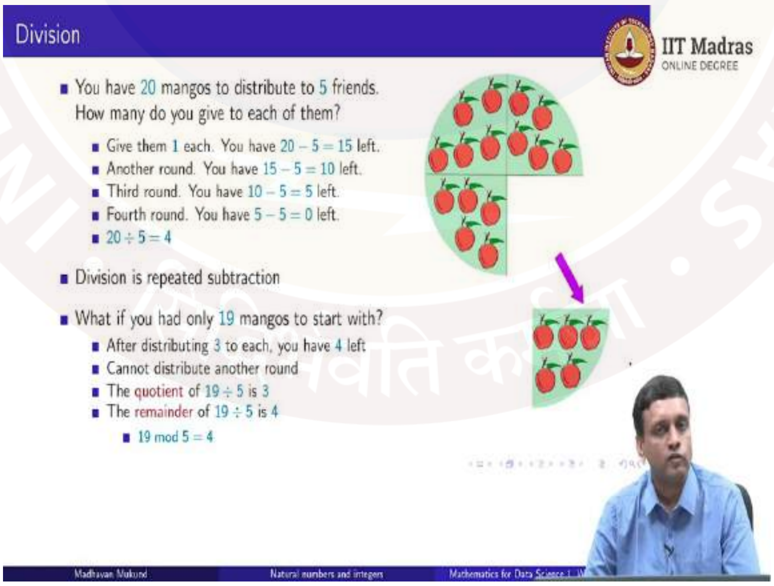


Now, unfortunately we live in a 3-dimensional world and we cannot imagine objects which have more than 3 dimensions. So, our vocabulary stops with cube. So, in general if we have m^k, then we write m × m × m…, k times and we just say it is m^k, we do not have a fancy name for it. We just say it is the kth power of m ok. So, to emphasize multiplication is repeated addition and exponentiation as we have seen is repeated multiplication.

(Refer Slide Time: 08:07)

So, now let us come to division. So, you would have seen this familiar problem in school. You have a certain number of objects and you want to divide them among certain number of people. So, for example, supposing you have 20 mangoes and you want to give them to 5 friends. So, how many mangoes does each friend get?

(Refer Slide Time: 08:31)



So, here on the right we have this picture and then, what you do is well you start by distributing one mango to each friend right. So, you take out 5 mangoes and you give them to each of your friends. So, now, you have given away 5 mangoes and you have only 15 mangoes left so, you repeat the process. Among the 15 mangoes, you give away 5 to your friends one each and now your 15 mangoes have become 10 and do it one more time and your 10 mangoes have become 5, do it a third time or fourth time rather and the 5 mangoes are now gone.

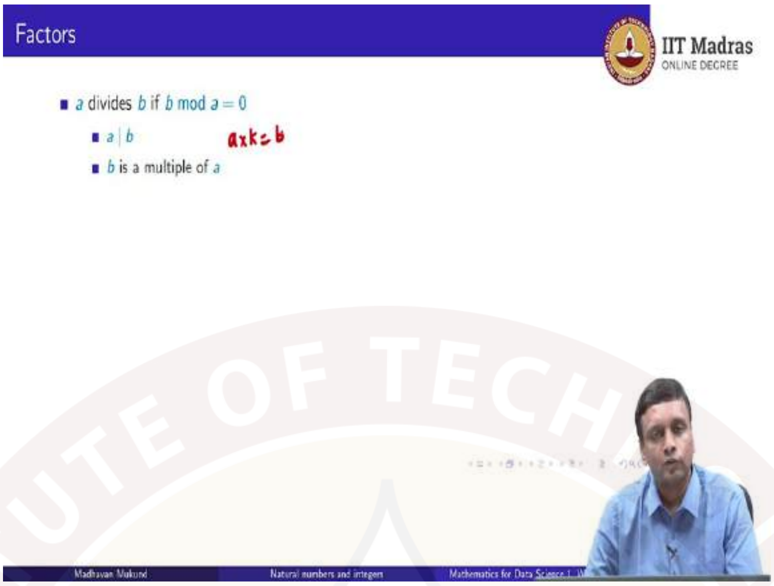
So, after 4 rounds of distributing mangoes, each time giving one mango each so, 5 mangoes per round, you have got rid of your 20 mangoes so, 20 ÷ 5 is 4. So, here as we have illustrated, division is actually repeated subtraction. You keep subtracting by the number you are trying to divide and finally, if you hit 0, then you have divided it exactly.

Well, what if you had only 19 mangoes? Now you know very well that 19 mangoes cannot be evenly divided into 4 into 5 groups. So, if you would start distributing like we had above the first three rounds would go fine; you would come from 19 to 14 from 14 to 9 and then you will come from 9 to 4 and now you have only 4 mangoes left and you have 5 friends so, you cannot give 1 each.

So, we have managed to distribute 3 times and we have 4 left over. So, formally this is written as saying that the quotient the number of times you can actually divide without getting into a fractional part is 3 and the remainder that is after you have a little bit left over which you cannot subtract one more time is the remainder is 4. So, for 19 ÷ 5, the quotient is 3 and the remainder is 4.

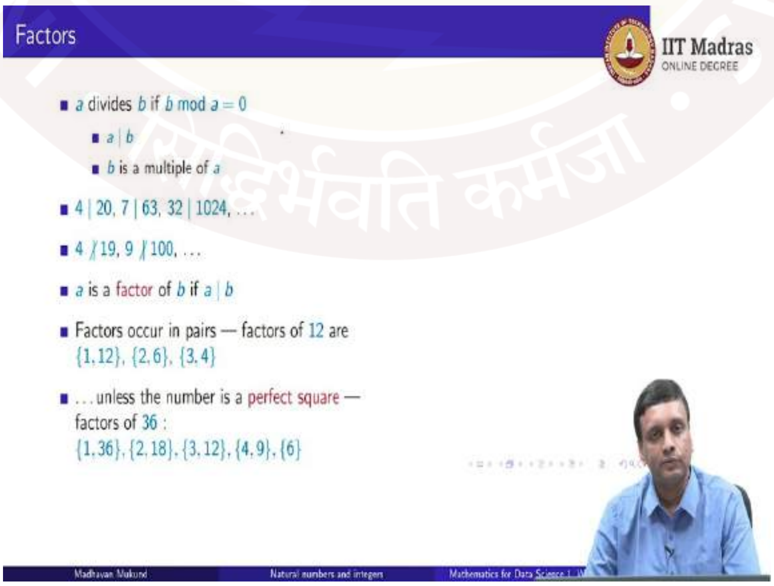
Now, very often we will need to use this remainder and there is a notation for remainder. So, this is this notation called modulus. So, modulus is another word for remainder and it is written as mod. So, 19 mod 5 is the same as the remainder when 19 is divided by 5. So, instead of saying the remainder of 19 divided by 5 is 4, we will often say 19 mod 5 is 4.

(Refer Slide Time: 10:38)



So, with this notation, we can now define what is a factor. So, a factor is a number which divides a bigger number evenly without any remainder. So, a | b, if b mod a is 0. Remember what this mean is means is that if b is divided by a, there is no remainder and we write this with this vertical bar |. So, on the left is the smaller number, on the right is the bigger number. So, a divides b this is what this is supposed to say and the other way of thinking about it is that b is some multiple of a. So, b if a | b then a.k=b ok. So, we have some multiple the some number of times that a goes into b. So, therefore, b is a multiple of a.

(Refer Slide Time: 11:27)



So, here are some examples we have already seen that 4 |20 because 4 × 5 is 20, 7 | 63 because 7 × 9 is 63, 32 |1024 because 32 × 32 is 1024 and so on.

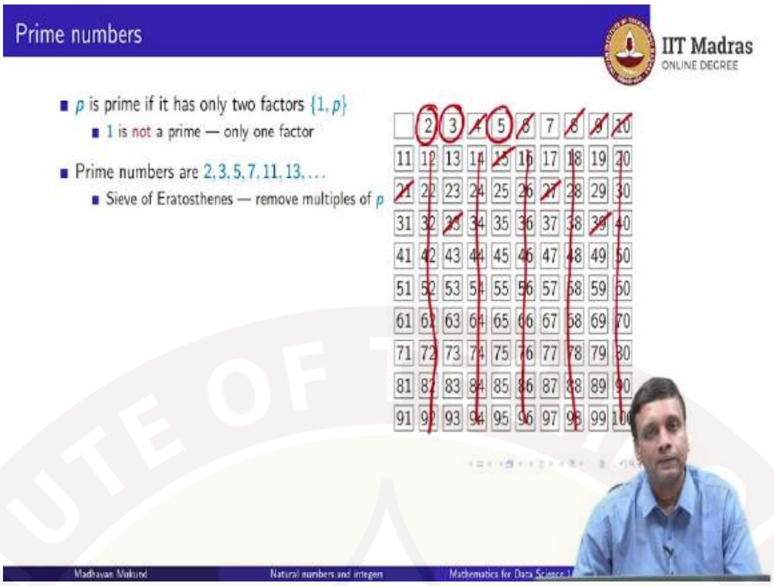
Now, the symbol that we use for not being a divisor is just to put a stroke across that vertical line. So, 4 does not divide 19 because there is no way to multiply anything by 4 and get 19. Similarly, 9 does not divide 100 evenly because we get 9 ×11 = 99 and then we go 108.

So, we say formally that a is a factor of b if a | b right. So, a | b is the same as saying that a is a factor of b and it is easy to see that factors must come in pairs because if a | b then, a goes into b some k times. So, k | b right so, k × a = b so, both k is a factor and a is a factor. So, for instance, if you take a number 12 then 1 is a factor because 1 divides everything and in fact, for every number n, 1 × n is n so, the pair for 1 is always the number itself.

Now in this case, 12 is divisible by 2 and 2 goes in 6 times. So, the pair 2, 6 form 2 factors 6 times 2 is 12, 2 × 6 is 12 and similarly 3 × 4. Now, of course, there is an important side condition which is that sometimes the pair is the same as the number itself and this happens when the number actually happens to be a perfect square that is, it is some number multiplied by itself. So, for instance consider 36 so, 36 is 6 × 6. So, if you look at the factors of 36 and group them in pairs, then we have 1 and 36, we have 2 and 18, we have 3 and 12, we have 4 and 9 and finally, we have the factor 6, but 6 is multiplied by 6. So, 6 does not produce a new factor as its pair, it is just itself.

So, another way of thinking about it is that, if you have something which is not a square you will have an even number of factors, you will have 2 + 2 + 2 +2. If something is a square, you will have an odd number of factors, you will have 2 + 2 +2 and finally, when you come to the number of which it is a square that number will come only once in the list of factors.

(Refer Slide Time: 13:45)



So, once we talk about factors, we come to a very interesting class of numbers which are the prime numbers. So, a prime number is one which has no factors other than 1 and itself. So, 1 is a factor always and 1 × n is n. So, for we try to usually write p for a prime number. So, a prime number has only two factors 1 and p.

Now, it is important that it must have two factors, two separate factors. So, one technically is not a prime because it has only one factor one itself because 1×1 is 1 and so, the only factor that 1 has is 1. So, the smallest prime actually is 2 because it has two factors 1 and itself 2 and no other factors. 3 is also a prime because it has only 2 factors 1 and 3, 2 does not go into 3 and so on.

So, we are all familiar with the smaller prime numbers. So, 2 is the first prime number, 3 is the next prime number, then 5, then 7. Notice that, after 2 no even numbers can be primes because they are all multiples of 2 and so, 2 divides them. Now we come to 9 and 9 is not a prime number because it is a multiple of 3, but 11 is a prime number and so on.

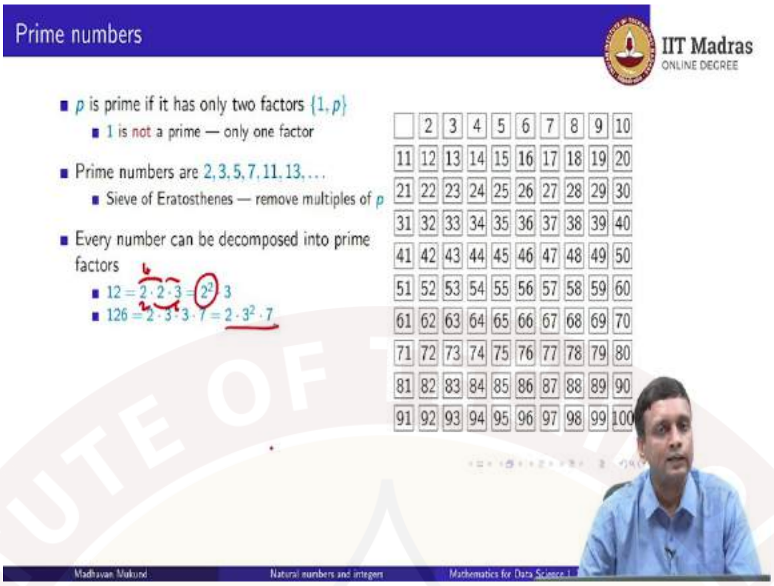
So, there is actually one clever way which is call the sieve of Eratosthenes to generate prime numbers which is whenever you discover a prime, you knock off all the numbers which are multiples of it. So, we can do this for instance to get all the prime numbers from 1 to 100. So, what we do is we first lay out a grid like this right, we know that 1 is not a prime so, the first prime that we have as a candidate is 2 right. So, this is how the sieve of Eratosthenes works, you lay out the numbers in a grid and now we can try and mark off all the prime numbers which are up to 100.

So, we know that 1 is not a prime so, we leave 1 off the grid and we start with 2. So, 2 is our first prime number and what the sieve of Eratosthenes says is you knock off all multiples of 2. So, you knock off all the even numbers and of course, now you can do it in one shot so, you can knock off this whole column, this whole column so, all these numbers are not prime ok.

So, now once you have you have a target so, we are looking only up to 100. So, up to 100 we have knocked off all the powers of 2 or all the multiples of 2. So, now, we look at the first number which is not been marked off and we notice that 3 is a prime because 3 is not yet marked off. So, now, we start mark off multiples of 3, some of them are already marked off because they are multiples of 2. So, 6 is already gone, but 9 is also gone, 12 is already gone, but 15 is also gone and so on.

So, we can mark off all the other multiples of 2 which are not multiples of 3 and so on right. So, we get this kind of a picture and now having done this assuming we have done it all the way, then we will come and find that 5 is a prime right. So, this is the process by which if you want to know count all the primes up to a certain number n, you can write out all the numbers up to n and starting at the left you can take the first unmarked number, call it a prime and mark all its multiples to the right as non primes and the next unmarked number will be the next prime and so on.

(Refer Slide Time: 16:56)

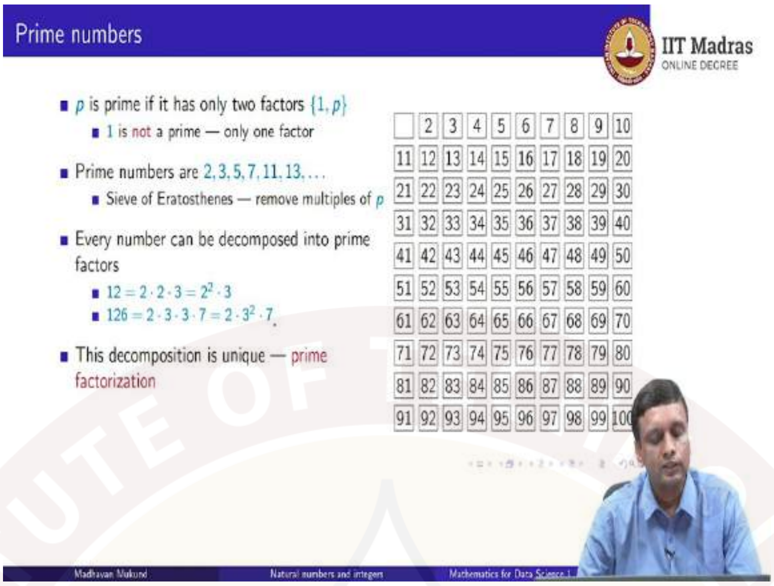


Now, this is not necessarily an efficient way to do the prime numbers, but this is a good way to generate them without missing out any. One of the important facts that we use all the time is that every number can not only be factorized as we have seen into a number of different pairs of factors it can actually factorize uniquely into the prime numbers that form it.

So, for instance if we look at 12, we said that 12 was 2 times 6, it was also 4 times 3, it was 1 times 12 and so on, but fundamentally it has 3 prime factors 2 2 again and 3. So, depending on how we combine them for instance, we can get 4 × 3 or we can get 2 × 6 and so on, but 2 × 2 × 3 is the absolute unique way of writing 12 as a product of prime numbers and using our exponentiation notation, we can condense this and put the 2 2's together and say it is $2^2 \times 3$.

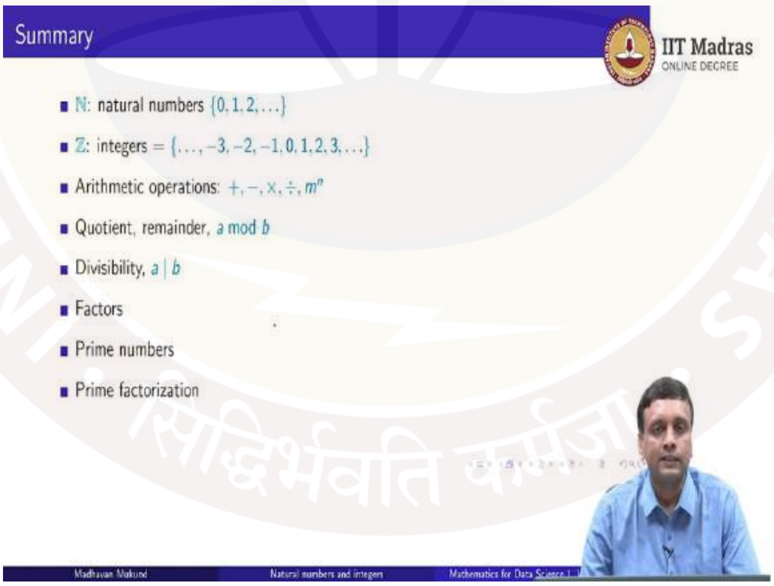
Similarly, if we take a number like 126, then it is 2 × 3, 6 × 3, 18 ×7 ok. So, the prime factors are precisely 2 3 twice and 7 and we can write this as $2 \times 3^2 \times 7$. So, this is very important because we use it implicitly along a lot and we will see later how we use this.

(Refer Slide Time: 18:10)



So, this is called the prime factorization right. So, every integer can be decomposed into a product of primes in a unique way.

(Refer Slide Time: 18:20)



So, to summarize we started with a natural numbers which we use for counting which are the numbers 0, 1, 2, 3, 4 and so on. Then, we extended these numbers with a negative numbers and this gave us the set of integers. So, the integers include all the natural numbers as well as the negative numbers 0, 1, 2, 3 and so on -1, -2, -3 and so on. We saw some basic arithmetic operations on these the usual addition, subtraction, multiplication, division and exponentiation.

We also looked at what happens when we divide integers and we do not want to look at fractions, then we talk about the quotient which is the integer number of times that the dividend goes into the number and the remainder is also written as a mod b. So, using this notation of a mod b, we can talk about divisibility which we write with a vertical bar. So, a | b if a mod b is 0. So, the factors of a number are those numbers which divide it and a prime number has exactly two factors 1 and itself and we can always decompose any integer uniquely into the list of factors, prime factors that multiply out to form that number.

IIT Madras

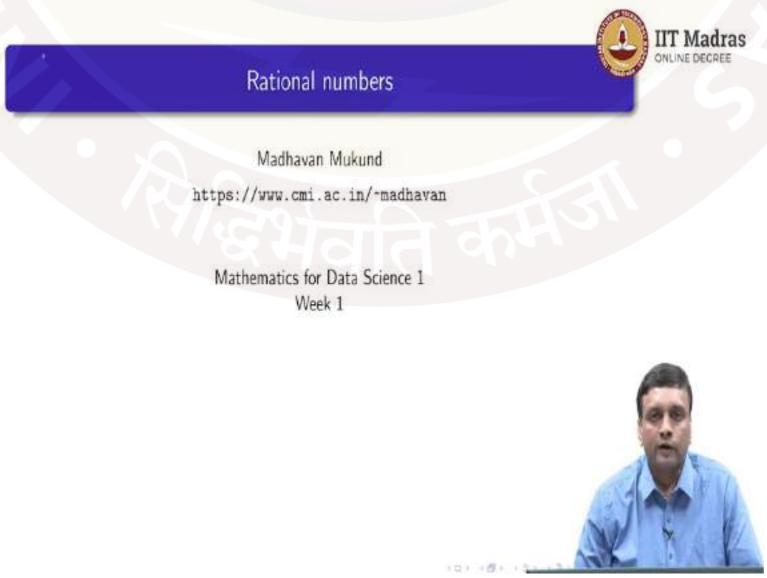
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture - 02
Rational Number

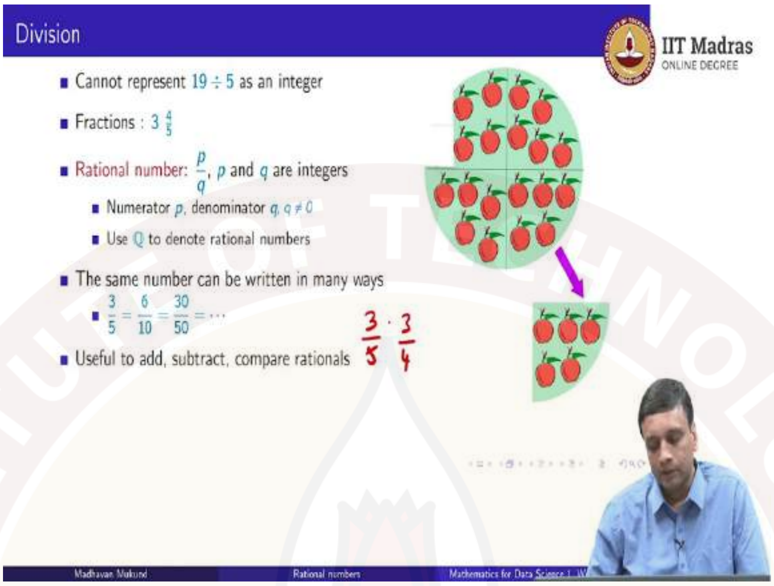
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, now first lecture on Numbers; we looked at natural numbers and integers. So, now, let see what happens when we try to divide. So, let us look at the rational numbers.

(Refer Slide Time: 00:23)



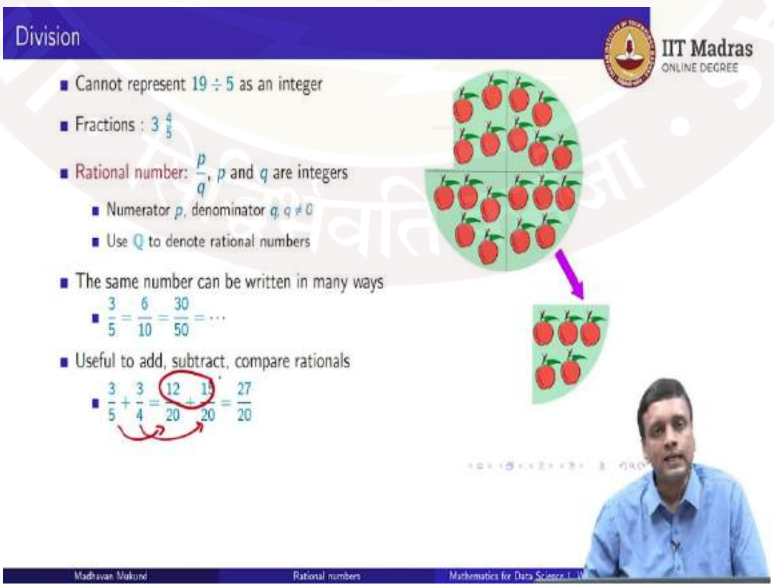
So, we said that we cannot represent 19 / 5 as an integer because we cannot find a number k such that 5 × k is 19. So, as we know the way we deal with this is to represent this quantity as a fraction. So, we say that 19 /5 is $3\frac{4}{5}$. So, this number is an example of a rational number. So, rational number what we usually called fractions in school, a rational number is something that can be written as ,$\frac{p}{q}$; where, p and q are both integers. So, as you probably remember from school, the number on the top is called the numerator. So for,$\frac{p}{q}$; p is called the numerator and q is called the denominator.

So, just like we had the symbols N and Z to represent the natural numbers and the integers, we have a special symbol which is somewhat unusual which is Q. So, Q stands for the rational numbers and again, to just say it is a special Q, we write these double lines on sides. So, this Q with these fat boundaries denotes the rational numbers. So, one thing about the rational numbers is that the same number can be written in many different ways. Now, this is not true of integers. Of course, we are not talking about changing base from binary to decimal or something.

But if you write a 7, there is only one way to write 7 fix, if you are fix the notation that you are using for writing numbers. With rational numbers, this is not true because there are many ways of writing $\frac{p}{q}$ such that $\frac{p}{q}$ is actually a same number. So, for instance if we take the number $\frac{3}{5}$, then we all know that $\frac{3}{5}$ is the same as $\frac{6}{10}$ and this is the same as $\frac{30}{50}$. So, when we take a rational number and multiply it by something the same quantity on the top and the bottom so, $\frac{3}{5}$, 3×2 and 5×2, we get the same number; $\frac{6}{10}$ or 3×10 and 5×10, we get the same number $\frac{30}{50}$. So, this is sometimes a nuisance, but it is also sometimes useful.

Now, there is no reasonable way to compare two numbers like say $\frac{3}{5}$ and $\frac{3}{4}$ or $\frac{2}{5}$ and $\frac{3}{4}$. If we have two fractions which have different denominators, there is no way to directly compare them. So, the only way to compare them is to somehow convert them into equivalent fractions such that they have the same denominator. So, the usual way is just to find a number such that both the denominators multiply into that number rather factors of that number. Now, you can find the smallest such number which is called the least common multiple; but you can find any number of this form.

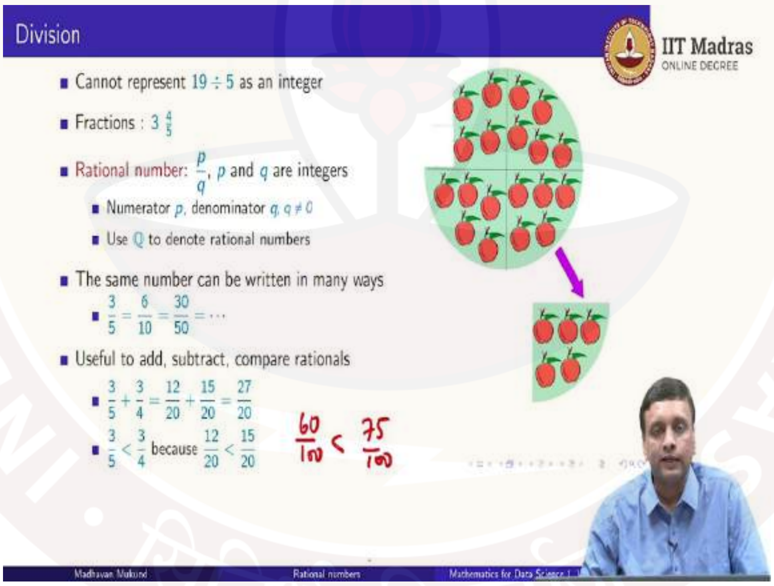
(Refer Slide Time: 03:05)



So, for instance, if you want to add $\frac{3}{5}$ and $\frac{3}{4}$, now you cannot do that directly; but you know that 20 is a number which divides which is a multiple of both 5 and 4. So, you can represent $\frac{3}{5}$ as equivalently as $\frac{12}{20}$; you can represent $\frac{3}{4}$ equivalent. So, this is equivalent and this is equivalent. So, you have converted these numbers into a different fraction of the same number; but this new representation has the same denominator.

And now once, the two denominator that the same, you can add the numerators and you can get (12 + 15)/20 is $\frac{27}{20}$. So, this kind of manipulation requires the denominators to be the same and therefore, it is actually extremely useful that we can write the same rational number in many different ways. The same is to we want to compare two numbers.

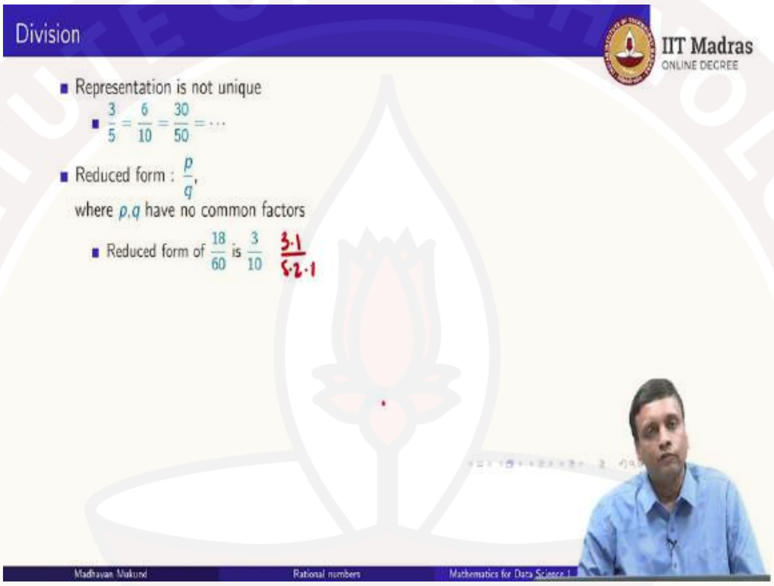
(Refer Slide Time: 03:54)



If we want to check whether $\frac{3}{5}$ is bigger or smaller than $\frac{3}{4}$, there is no way to do it directly. What we have to do is again take the denominators and make them the same and then, say that $\frac{12}{20}$ is less than $\frac{15}{20}$ because you are dividing something 20 parts and you are taking 12 of them that is less than taking 15. Now, as I said there is no reason why this must be the smallest one. So, for instance you could take a bigger number like 100, right. So, 5 goes into 100 and 5 goes 4 also goes into 100.

So, we could also say that $\frac{3}{5}$ is the same as $\frac{60}{100}$ and, $\frac{3}{4}$ is the same as $\frac{75}{100}$ and therefore, since 60 is less than 75; $\frac{60}{100}$ is less than $\frac{75}{100}$ and therefore, $\frac{3}{5}$ is less than $\frac{3}{4}$. So, it is not really important that the denominator is the smallest common multiple of the two denominators; but it must be some common multiple so that you can bring it all to a common number that you can then compare.

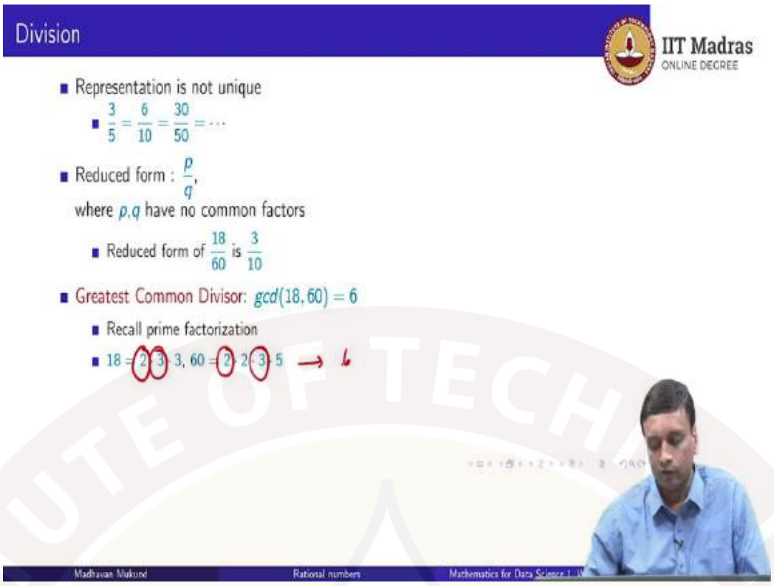
(Refer Slide Time: 04:49)



So, we saw that representation is not unique for rational numbers. So, how do we find actually the best way to represent a rational number? So, normally if you are not using it for some arithmetic operation of some comparison, we would prefer to have it in a reduced form. So, the reduced form of a rational number is one, where there are no common factors between the top and the bottom. So, $\frac{p}{q}$ is of the form, where we cannot find any factor f such that f | p and f | q.

So, for instance, if we take $\frac{18}{60}$, then its reduced form will be $\frac{3}{10}$. Notice that 3 is of the form 3 × 1 and 10 is of the form 5 × 2 × 1. So, therefore, there is no common factor between the top and the bottom and therefore, this is in reduced form.

(Refer Slide Time: 05:42)

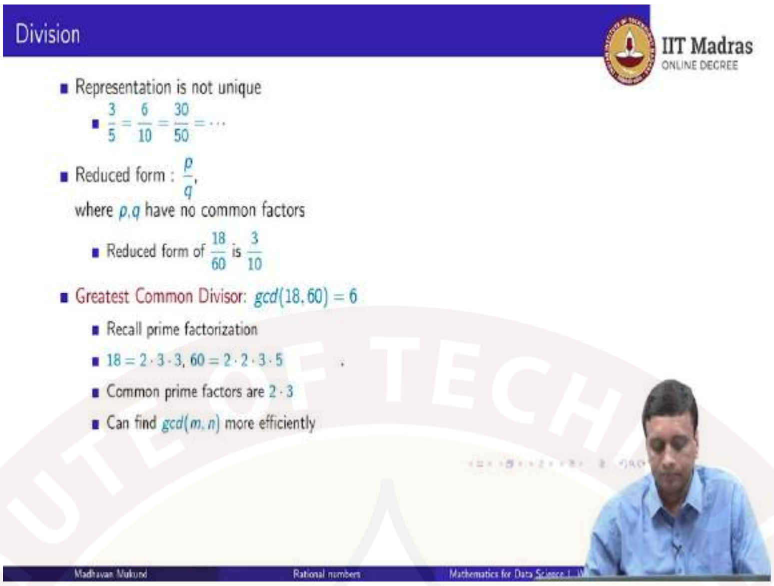


So, this is called the greatest common divisor problem. So, we want to find the largest number which divides both the top and the bottom; both the numerator and the denominator; divide them both by this and then come to something in the reduced form. So, in this case, what we are saying is that the gcd of 18 and 60 is actually 6 and we can do this using our prime factorization that we talked about before.

So, if we look at prime factorization for 18, then 18 is $2 \times 3 \times 3$ right; its 2×3 is 6 and 6×3 is 18 and the prime factorization of 60 is $2 \times 2 \times 3 \times 5$; its 4×3, 12 and 12×5. So, now, you can look at what are common. So, we have one 2 here and one 2 here. So, we can say that this is part of the same factor, we have one 3 here and another 3 there. The second 2 is not present in the first term.

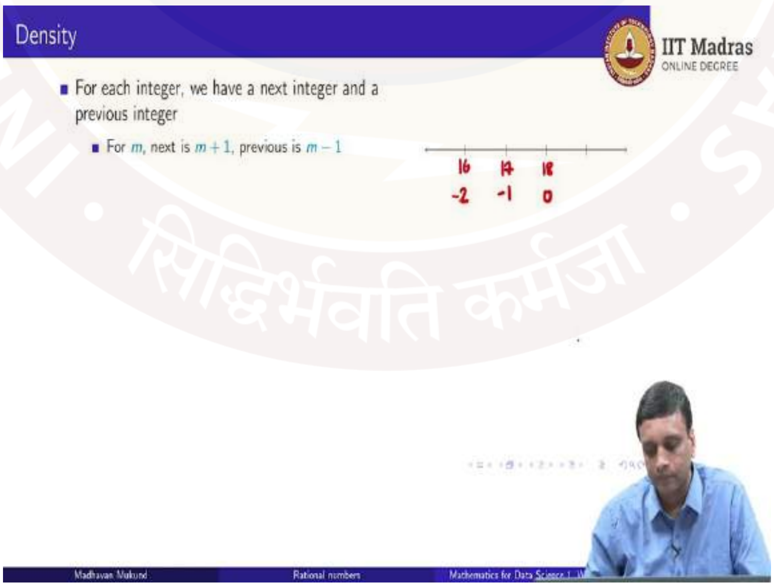
So, we have a 2 and 3 and 18 which are factors. We have a 2 and 3 in 60 which are factors and this gives us the fact that 6 is a common factor. There is no bigger common factor because we want to assemble a bigger common factor, we have to pull out one more prime from each side; but there is no prime left which is present on both sides. 3 is there in 18; 2 and 5 are there on 60, but we do not have a matching one of the other side right.

(Refer Slide Time: 06:59)



So, this way, the common prime factors are one 2 and one 3 and so, 2 × 3 equal to 6 is the gcd. Now, this is not the best way to find the gcd, there are more efficient ways to find the gcd. But this intuitively tells us what the gcd is. You take the prime factorization of both the numbers and you collect together all the primes that occur in both the numbers, the same number of times.

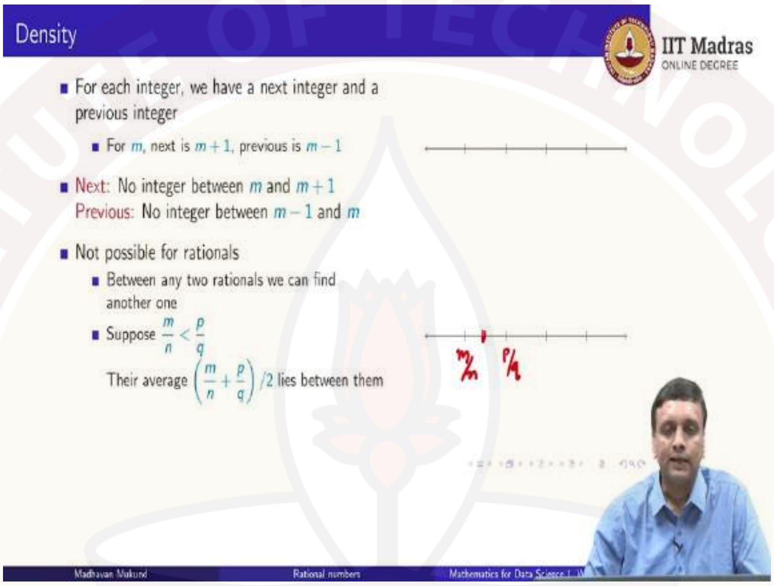
(Refer Slide Time: 07:23)



So, here is another interesting property about rational numbers. Now, for each integer, we know intuitively that there is something which is the next integer and the previous integer. If

I tell you 22 and ask you what is the next integer? Then, you will know it is 23. What is the previous one? It will be 21. So, for every integer m, the next one is m + 1 and the previous one is m - 1 and it does not matter, if this is positive or negative. So, for instance if I am at 17, then the next integer is 18, the previous one is 16; right. If I am at -1, then the next integer is 0 and the previous integer is -2. So, I can always take the integer that I am at, add 1 and get the next integer, subtract 1 you will get the previous integer.

(Refer Slide Time: 08:10)



So, the property of this next and previous is that there is nothing in between right. So, there is no integer between m and m +1, there is no integer between m and m -1. So, that is what next means, it is not some bigger integer or some smaller integer. It is the immediate neighbor in the integer of the in this number line. Now, what about rationals? Is it possible to talk about the next and the previous rational number? Now, it turns out that this is not possible for a very simple reason.

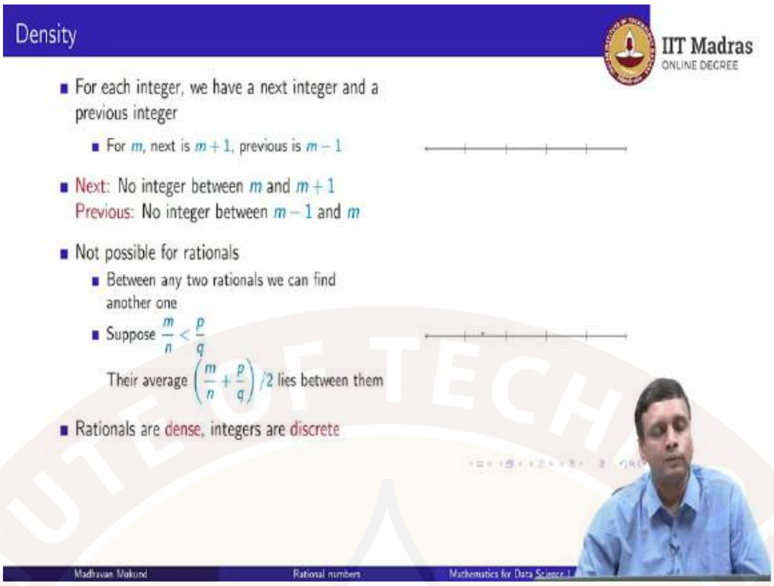
So, between any two rationals, we can always find another one because we can always take the average of 2 numbers. So, remember that if you take the average of any 2 numbers, then it must be between those 2 numbers right because it is the sum of the numbers divided by 2. So, the average cannot be smaller than both or cannot bigger than both. So, if the 2 numbers are not the same, then it must lie strictly between them. If the numbers are the same, then the average is the same.

So, if somebody has 37 marks and 37 marks, then their average marks is 37. But if they have 37 marks and 52 marks, even without calculating the average, you know that their average is bigger than 37, but smaller than 52; right. So, in the same way, if I give you 2 fractions $\frac{m}{n}$ and $\frac{p}{q}$ and I tell you that $\frac{m}{n}$is smaller than $\frac{p}{q}$. Remember that in order to do this, we would have to normally get the denominators to be the same and so on.

But supposing I know that $\frac{m}{n}$ is smaller than $\frac{p}{q}$. So, I know that say $\frac{m}{n}$is here and I know that say $\frac{p}{q}$is here and supposing you claim that $\frac{m}{n}$ and $\frac{p}{q}$ are adjacent, that is $\frac{p}{q}$is the next rational after $\frac{m}{n}$. Well, I will say no; let me take these 2 numbers and find its average right. So, this average now is also a rational number because you can also represent it as $\frac{a}{b}$ right. If you just workout this $\frac{m}{n}$ plus $\frac{p}{q}$ divided by 2, you can simplify this whole expression and you will get a new number which is also of the form $\frac{a}{b}$.

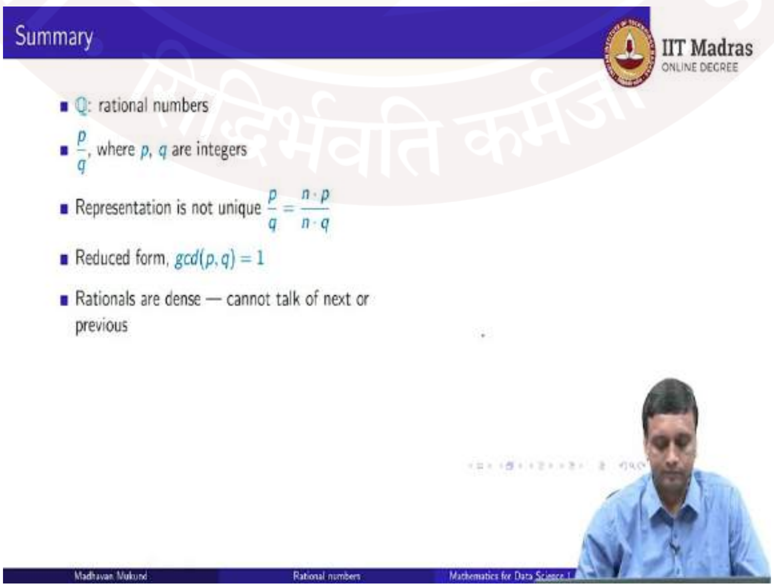
So, this is also a rational number and this rational number as we argued must be between the 2 numbers and therefore, between any 2 rational numbers by just taking the average of the mean of the 2 numbers, I can find another one.

(Refer Slide Time: 10:17)



So, in other words, the rational numbers are dense right. So, dense in the usual sense, so dense just means that they are closely packed together. So, basically you cannot find any gaps in the rational numbers because any between any 2 rational numbers, you will find another rational number and this is not true of the integers because we saw that in the number line, there is a gap between m and m + 1, there is no integer there right. So, we say that the rational numbers are dense and conversely, we say that the integers and the natural numbers are discrete. So, a discrete set has this kind of next property and a dense set has no next property between any 2 numbers, will find another number right.

(Refer Slide Time: 10:56)



To summarize, we use this funny symbol Q to denote the rational numbers and a rational number is just the ratio. So, that is where it comes from actually; so, ratio. So, rational number comes from the word ratio and so, it is a ratio of 2 integers p divided by q. Now, there is no unique representation of a rational number because we can multiply both the numerator and the denominator by the same quantity and get a new rational number which is exactly the same in terms of the quantity that it represents.

And we use this fact for things like arithmetic and comparisons, but if we really want to talk about rational numbers in a canonical way, in a unique way; then, we get this reduced form, where we cancel out the common factors using prime factorization. So, that we get a number whose gcd of the numerator and the denominator is 1.

And finally, we saw that we cannot talk about the next or the previous rational number because between any 2 rational numbers, there is another rational number. In particular, if you take the average of the 2 numbers, you will find a number that is in between. So, unlike the integers and the natural numbers which are discrete for which next and previous makes sense; for the rational numbers, there is no such quantity.

IIT Madras

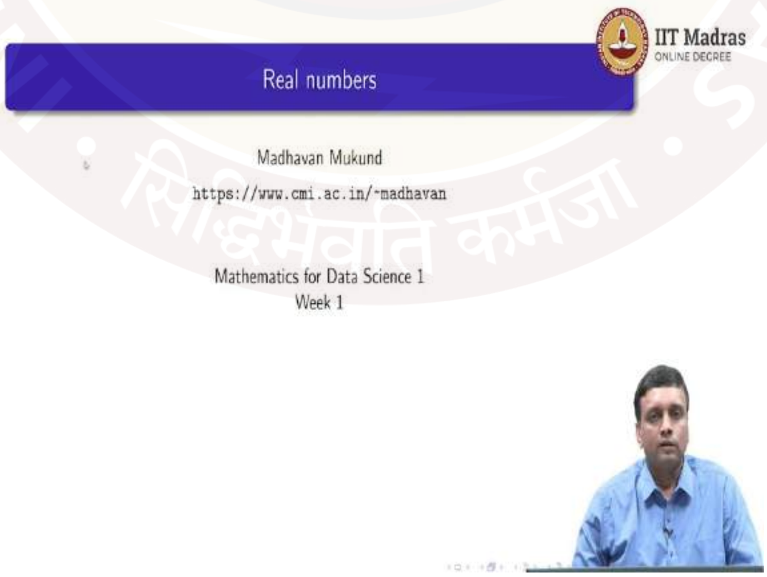
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture - 03
Real and Complex Number

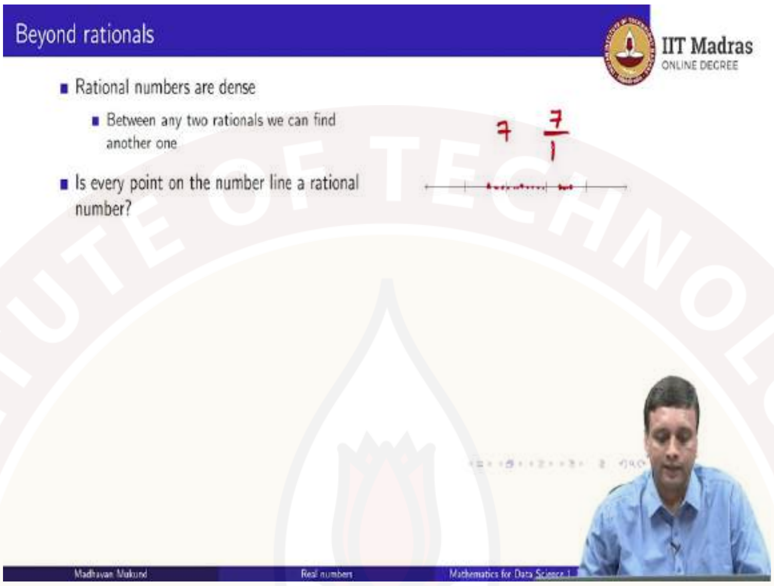
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, we started with the natural numbers and the integers and then, we moved on to the rational numbers which are defined as $\frac{p}{q}$; where p and q are both integers.

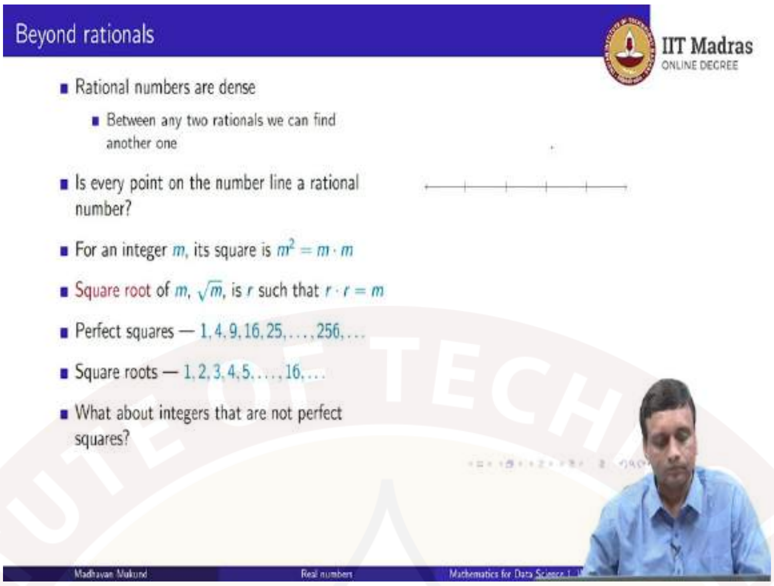
(Refer Slide Time: 00:23)



So, we decided that the rational numbers are dense right and that means that on this number line between any two rationals, you can find a rational. So, if I want to now talk about this number line, then I know that if I take any two positions, then I will find a rational between them and I will find a rational between them and so on. So, it makes sense to ask this question which is that if I take any two points and the rational between them any two points, then is this entire number line composed only of rational numbers. Of course, some of those rational numbers are integers.

So, an integer is a rational number because I can write 7; for instance, as $\frac{7}{1}$right. So, this is of the form $\frac{p}{q}$. So, any rational number which in reduced form as denominator 1 is an integer; so, an integer is a special case of a rational number. So, do all the rational numbers fill up this number line? That is the question.

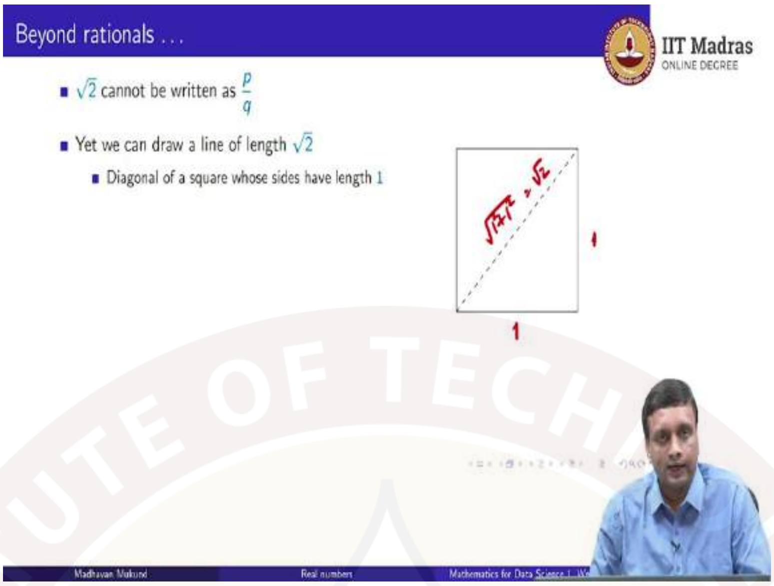
(Refer Slide Time: 01:12)



So, it turns out this is not the case. So, remember that a square of a number is the number multiplied by itself. So, if I take a number m and multiply it by itself, I get m^2 which is m × m and if I take this operation and turn it around, then the square root of a number is that number r such that r × r is equal to m right. So, I want to find out which number, I have to square in order to get m and that is called the square root.

So, if we take the so called perfect squares, like 1, 4, 9, 16, 25 and so on their square roots are integers. So, 1^2 is 1. So, the √1 is 1; $2^2$is 4, so the √4 is 2; $5^2$is 25, so √25 is 5; $16^2$is 256, so √256 is 16 and so on. So, some integers are clearly squares of other integers and so, you can get the square root and find an integer. Now, what happens if something is not a square right? So, supposing I take a number which is not a square like 10 and I take its square root, I know that the square root is not an integer, its somewhere between 3 and 4 because $3^2$is 9 and $4^2$is 16. Question is, is it a rational number or not?

(Refer Slide Time: 02:36)

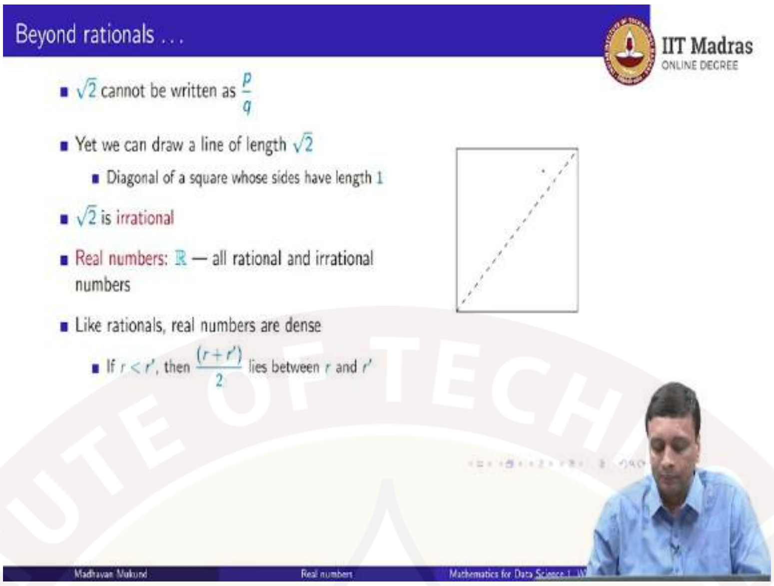


So, what happens to the square roots of integers that are not perfect squares? So, the smallest such number which is not a perfect square because 1 remember is a perfect square, 1 × 1 is 1. The smallest such number that is not a perfect square is actually 2 and it is one of the very old results that the $\sqrt{2}$ cannot be written as $\frac{p}{q}$. This was certainly known to the ancient Greeks, in fact, to Pythagoras and one way to do this is to see that you can actually draw a line of; so, this is not an unreal number in that sense right.

So, you can actually draw a line of this length because if you take a square, whose sides are 1 right. So, this is 1, then if you remember your Pythagoras theorem; then, the hypotenuse of this triangle is going to be $\sqrt{1^2+1^2}$,technically which is $\sqrt{2}$. So, I can actually physically draw a line whose length is $\sqrt{2}$. So, this is a very real quantity.

On the other hand, for reasons that we will not described here, but there will be a separate lecture explaining this for if you are interested. $\sqrt{2}$ cannot be written as a rational number $\frac{p}{q}$. So, here is a number which is a very measurable quantity, I can actually draw this quantity as a length. At the same time, it does not fit into this number line of rational numbers which seems to cover all the rational numbers, all the numbers because they are dense.

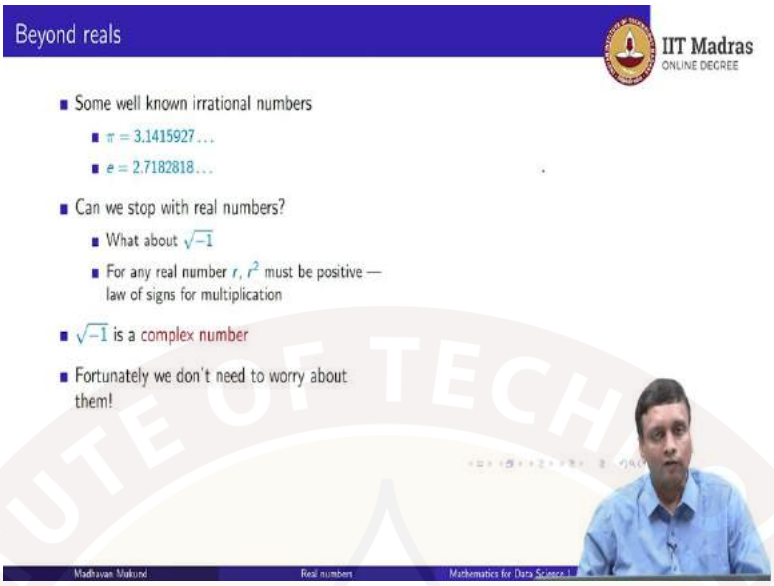
(Refer Slide Time: 03:56)



So, $\sqrt{2}$, since it is not a rational number right and yet it exists is called an irrational number and these numbers which constitute all the rational numbers and the real irrational numbers together are called the real numbers. So, the real numbers are denoted by this double line R. So, we had N for the natural numbers, Z for the integers, Q for the rational numbers and now, we have the real numbers R.

So, the real numbers extend the rational numbers by these so called irrational numbers which are very much on the number line, but which cannot be written on the form $\frac{p}{q}$. Now, it is not difficult to argue that like the rationals, the real numbers are dense for the very same reason. Because if you have two real numbers r and r' such that r is smaller than r prime, then you can just take their average r + r' divided by 2. This must be a number π which is bigger than r and it is smaller than r' and therefore, it must lie between them. So, between any 2 real numbers, you will find another real number. So, the real numbers are also dense.

(Refer Slide Time: 05:00)



So, there are some irrational numbers which we use a lot in mathematics and which you have probably come across; one of them is this famous number π which comes when we are talking about circles. Because it is the ratio of the circumference to the diameter and this is an invariant. π is always; the circumference divided by diameter for any circle is π ok.

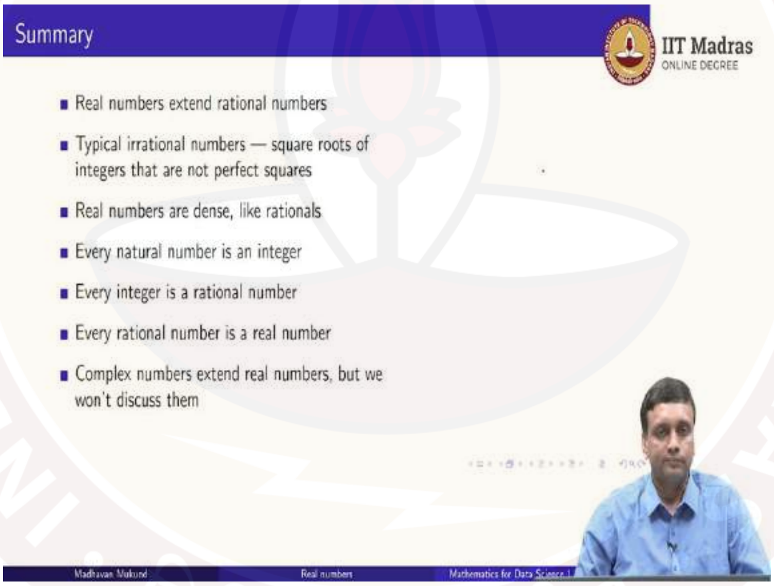
So, π is an irrational number. We cannot write it in the form $\frac{p}{q}$ and it has this. If you write it in this decimal form, it has this infinite decimal expansion. Another number which is very popular as an irrational number is this number e which is used for natural logarithms. So, it is 2.7182818 and so on right. So, there are a lot of irrational numbers. So, $\sqrt{2}$ as we have seen as an irrational number. It will turn out that square root of anything, $\sqrt{3}$ is also an irrational number, $\sqrt{6}$ is also an irrational number.

Anything which is not a perfect square, its square root is actually an irrational number. But many of these numbers are not very useful to us, but π and e are certainly very useful irrational numbers. So, now, we have seen that we can find more numbers on the line than just the rationals and these are the real numbers. So, do we stop here? Well, let us look at the square root operation which we use in order to claim that there are irrational numbers. So, what happens if we now take the square root of a negative number like -1?

So, remember that we had a sign rule for multiplication. The sign rule for multiplication said that if I multiply any two numbers, then if the two signs are the same that is their two negative signs or two positive signs, I will get a positive sign in the answer. Only if the two signs are different, if I have one minus sign and one plus sign, will I get a negative answer. So, if I want to multiply two numbers and get a -1, one of them must be negative and one must not be negative. But by definition, a square root is a number which is multiplied by itself, the same number has to be multiplied by itself. So, it will have the same sign.

So, any square root which multiplies by itself must give me a positive number. So, if I take a negative number, there is no way to find a square root for it. So, if we want to find square roots for negative numbers, we have to create yet another class of numbers called complex numbers. So, complex numbers extend the real numbers, just like real numbers extend the rational numbers and rational numbers extend the integers and so on. But the good news for you is that we do not have to look at complex numbers for this course.

(Refer Slide Time: 07:35)



So, to summarize, a real numbers extend the rational numbers by adding the so called irrational numbers which cannot be represented of the form $\frac{p}{q}$ and a typical example of an irrational number is the square root of an integer that is not a perfect square. So, $\sqrt{2}$ for example is not a rational number and this is also of the case was $\sqrt{3}$, $\sqrt{5}$, $\sqrt{6}$ and so on. So, except for the perfect squares, none of the square roots are actually rational numbers. Now, just like we said that the rational numbers are dense because the average of any two rational numbers is a rational number. Similarly, the real numbers are dense because the average of any two real numbers is a real number.

So, we have a progression in terms of numbers. So, every natural number that we started with is also an integer because the integers extend the natural numbers with negative quantities. Now, every integer is also a rational number because we can think of every integer as a ratio $\frac{p}{q}$; where, the denominator is 1. And finally, every rational number is a real number because we said that the real numbers include all the rational numbers plus all the irrational numbers. And finally, we said that there are even things beyond rational numbers like complex numbers, but we will not discuss them.

IIT Madras

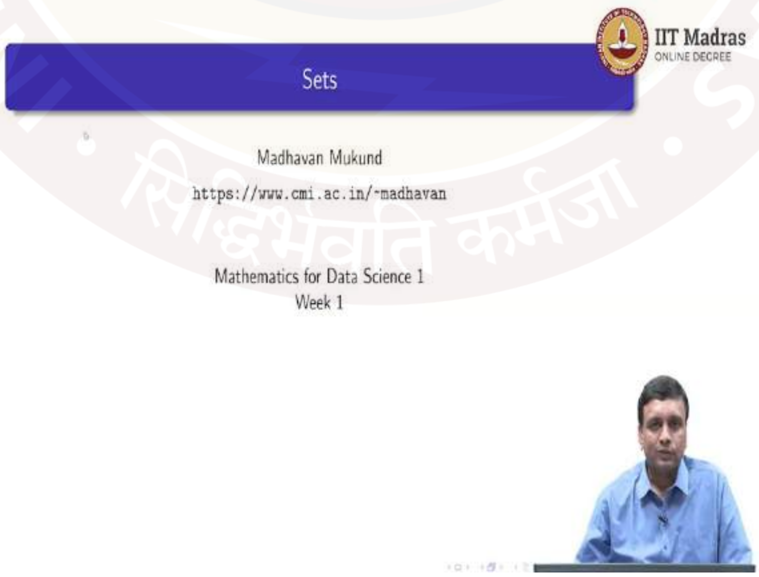
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture - 04
Set Theory

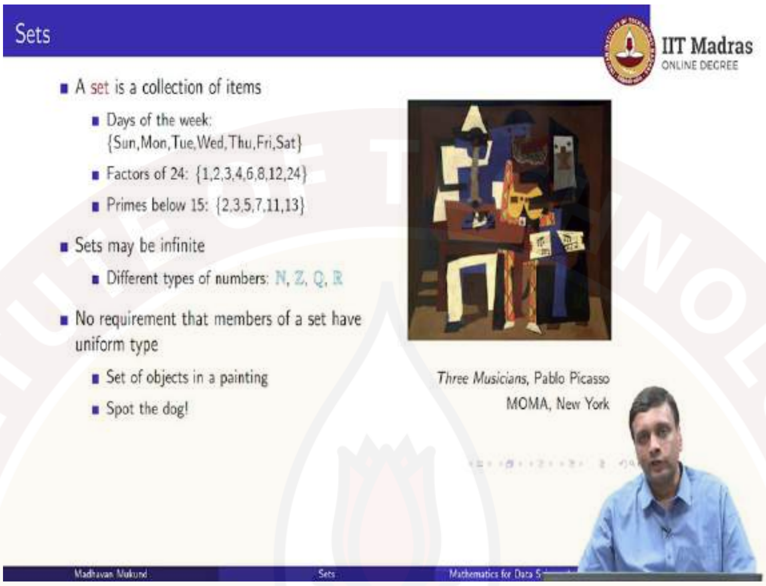
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, we have seen numbers; we have seen natural numbers, we have seen integers, rationals, reals and we have loosely talked of them as sets of numbers. So, let us try to understand little more clearly what we mean by a set.

(Refer Slide Time: 00:26)



So, at its basic level a set is a collection of items. So, for instance, we could have a set called the days of the week which has 7 members; Sunday, Monday, Tuesday, Wednesday, Thursday, Friday and Saturday or we could take a number like 24 and list out the factors of 24 and call this a set. So, we have 1, 2, 3, 4, 6, 8, 12 and 24.

So, if you count, there are 8 factors that 24 has or we could take all the prime numbers up to a certain limit. Supposing, we want to know the prime numbers below 15, then we know that we do not have 1; but 2, 3, 5, 7 are the single digit prime numbers and then, 7, 11 and 13.

So, 2, 3, 5, 7, 11, 13 are all the primes below 15. So, this is how we talk about sets informally as just collections of items. Of course, as we have seen sets can be infinite and in particular, the infinite sets that we deal with very commonly are those which consists of the different types of numbers.

Remember that N this funny N stands for the natural numbers that is 0, 1, 2, 3, 4. Z stands for the integers. So, that is the natural numbers along with the negative integers like -1, -2, -3 and so on. Q is a peculiar symbol for the rational numbers, these are the fractions those numbers which we can write as $\frac{p}{q}$; where, p and q are both integers.

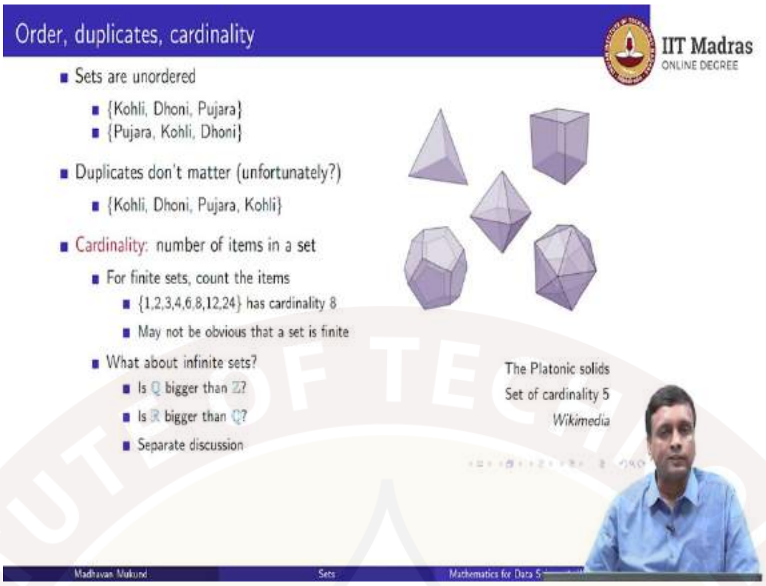
And finally, R is a set of real numbers. So, the real numbers includes all the rationals all the fractions, but also numbers that cannot be represented as fractions, such as the square root of 2 and other irrational numbers like π and e.

So, in all these things that we have seen above, it looks like there is some kind of condition which requires a set to have some uniformity; either a set consists of numbers or a set consists of days of the week or something like that. But actually mathematically there is no constraint on this. A set can have any kind of members, even a mixed membership; there is no uniformity of type.

So, for instance, we could enumerate the set of objects that appear in a painting. Now, here is a particularly famous painting, where it is not so easy to enumerate the objects because its drawn in a very abstract way. This is a painting called Three Musicians by Pablo Picasso. But we could see roughly that there are three people and that there are some musical instruments and so on and if you look very carefully, you will even find a dog.

So, notice that there is no commonality. There are people, there are musical instruments, there are chairs, there are tables, there are animals and so on. So, a set in particular can have any kind of members, it does not matter if they are mixed in type.

(Refer Slide Time: 03:00)



So, one of the important differences between say set and a sequence or a list is that the order in which we identify a set does not matter. So, normally when we talk of numbers, we tend to list them in a particular way; but as the set it does not matter. So, for instance, if you take the set of cricketers; Kohli, Dhoni and Pujara. If you reorder this set as Pujara, Kohli and Dhoni, it is the same set right.

So, the sequence in which you list the members of a set does not matter and for that matter, if you happened to accidentally write the same member twice, it does not change the set. So, in this particular set if we add Kohli a second time, as the set it does not matter. Though of course, if you are a cricket fan maybe you would like Kohli to bat twice for you.

So, when we look at a set, we might ask a basic question as to how many members it has. So, the cardinality of a set is the number of items in the set and if it is a finite set, we can just count the items. So, for instance, if you look at the factors that we listed of 24, then we can count them and say that this has cardinality 8.

Sometimes, it may not be obvious that a set is finite. You might remember from geometry that a regular polygon is one, where the all the sides are equal and all the angles are equal. So, the smallest regular polygon is an equilateral triangle in which we have 3 sides all equal and 3 internal angles of 60 degrees each. Then, we move to four sides we get a square, then we get regular pentagons, hexagons, heptagons, octagons and so on.

So, for any number of sides, you can draw a regular polygon with that many sides with equal angles on the inside. So, there is no limit. The set of regular polygons is infinite. But if we move to three dimensions, the corresponding notion to a regular polygon is what is called a platonic solid. In a platonic solid, first of all you have surfaces or sides each side is a regular polygon and all these regular polygons meet at the same angle in three dimensions.

Now, it turns out that though you might imagine that there are infinitely many regular polygons in two dimensions, there are only 5 platonic solids in three dimensions. So, this is an example of a set which turns out to be finite, even though there is no reason for it to be finite. So, these 5 platonic solids are the tetrahedron which has triangles.

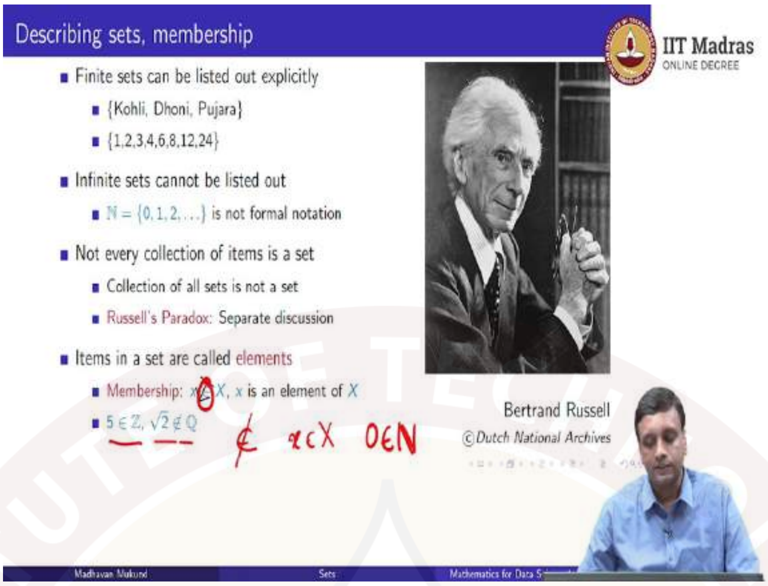
The cube which we have which has squares and then, we have an octahedron which has 8 sides which are triangles. Then, we have a dodecahedron with 12 sides and an icosahedron with 20 sides and there are no other regular solids, surprisingly it turns out.

Now, cardinality is quite easy to determine for a finite set, but what about for an infinite set? Remember that, we said that we wanted to go from integers to rational numbers because we want to talk about what happens when we divide 2 integers and the answer is not an integer and it is clear to us from our discussion that integers were discrete, we can talk about a next number and a previous number. So, there are gaps in the integers and rational were dense; between any 2 rational numbers, there is another rational number.

So, intuitively, it seems like we are adding things to the integers to get rational numbers. But can we make it formal in terms of cardinality? Are there more rational numbers than there are integers? And what happens, when we go from Q to R when we go from rational numbers to real numbers? So, remember that the real numbers, we had introduced because they were numbers such as the $\sqrt{2}$ which could not be represented as a fraction.

So, clearly there are some rational numbers which are real and some real numbers which are not rational and therefore, we have a bigger set; but again, R is really bigger than Q . So, this is a separate discussion, there will be a small separate lecture about this. But there is a way to measure cardinality of infinite sets, but it is not as straight forward as it is for finite set as you would imagine.

(Refer Slide Time: 06:58)



So, how do we describe a set? Well, we have already seen that for a finite set, we can just list out the members of the set explicitly. So, we can write out 3 numbers; Kohli, Dhoni, Pujara or 8 members the factors of 24. So, the normal notation for a list of items which form a set is to use these curly braces and to separate the items by commas.

Now, in many books and even in our lectures we will see notation like 0, 1, 2 …indicating that there is an infinite set of elements to be added which follows some kind of a pattern. So, this looks a way of listing out an infinite set, but you must understand that this is only an informal notation, this is not a formal notion.

So, you cannot write … and claim that you are listing out a infinite set. So, in fact, you need some other way of doing it and we will come to that as we go along in this lecture.

Now, it said seems reasonable that if a set is a collection of items, then we can collect anything and make it a set. It turns out that this is not quite true and this is particularly, a problem when we move to infinite sets. So, we have seen some infinite sets of numbers like naturals and reals and so on; but in general, if you take an infinite collection of objects, it may or may not form a set. In particular, Bertrand Russell showed that there is a problem, if we collect all the sets together and call it a set.

So, if we have a set of all sets, then we have a problem and this is something which is called Russell's Paradox which we will discuss in the separate lecture, but you must be careful to note that though the notion of a set is intuitive and it seems natural that any collection of objects is a set, we have to actually be a little careful in mathematics, if we are using sets in order to define what is a set and what is not a set.

But given that whatever we will see in our course, we will be fairly straight forward. So, whenever we see a collection of numbers or a collection of objects of mathematical description, we can safely assume that they are sets.

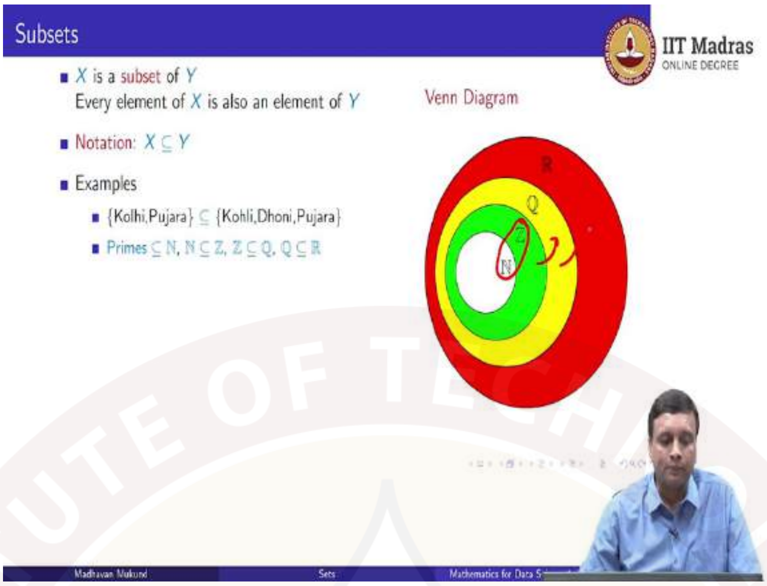
So, again some terminology. So, we have talked of different things items in a set, members of a set and so on. So, the most formal notation for the members of a set is an element. So, a set consists of elements and we write this membership of an element in a set using this $\in$ notion. So, we have this $\in$ notation which stands for element of. So, when we write $x \in X$, we mean that small x is a member of the set capital X.

So, example 0 is a member of the natural numbers right. So, $0 \in N$ is what we use. So, we can see for instance that 5 is an integer, but $\sqrt{2}$ as we claimed is not a rational number. So, an element of symbol with the line across it, means not an element of. So, 5 is an element of integer set and $\sqrt{2}$ is not a member of the set of rationals.

(Refer Slide Time: 10:02)

So, moving on from elements, we can compare sets by asking whether one set is included in another set and this is called a subset. So, $X \subseteq Y$, if every element of X is also an element of Y and this is written using this subset notation $\subseteq$. So, you have this familiar notation $X \subseteq Y$.

(Refer Slide Time: 10:27)



So, for example, if we take just 2 out of the 3 players were listed before saying Kohli and Pujara; then, this set forms the subset of our original set Kohli, Dhoni and Pujara. Similarly, if we take all the natural numbers and collect only the prime numbers. So, remember that the prime number is a number whose only factors are 1 and the number itself. So, it has exactly 2 factors; 1 and p and then, p is a prime number.

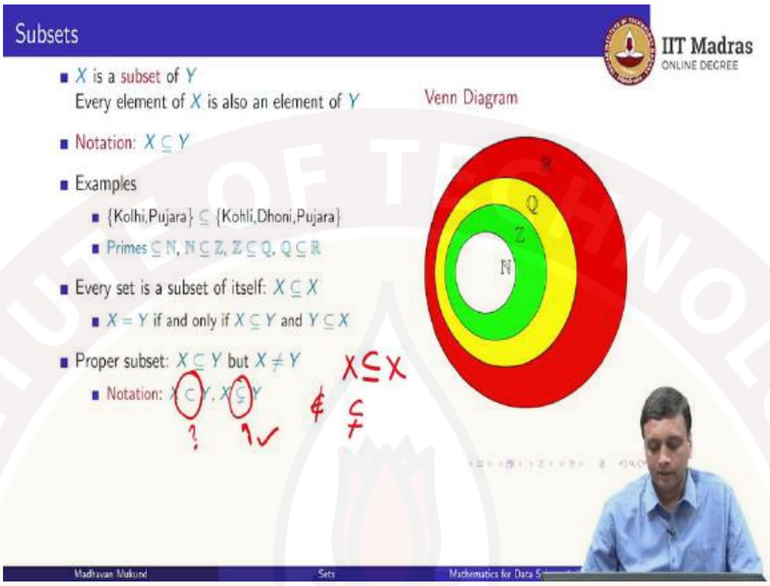
So, since some many numbers are not prime, primes is a subset of natural numbers. Since, the integers extend the natural numbers with the negative numbers, we can say that the natural numbers are included in the integers. So, $N \subseteq Z$. Similarly, we extended Z to Q. So, the set of integers is a subset of the rationals and the set of rationals is a subset of reals.

So, if you wanted to draw it, we could draw it in this particular way. So, we can raw a large circle representing the reals, a small circle inside right in the center representing the natural numbers and if one circle is included in another circle, it means that this circle is a subset of the circle outside it. So, here you can see that the natural numbers are a subset of the integers and then, from the integers, we can say that there are subset to the rationals and the rationals are a subset of the real numbers.

So, this kind of a diagram, where we represent a set by a boundary. So, this is a very abstract diagram. We are not in this case for example, listing out the elements of the set we are just indicating the extent of the set saying that the set are extends beyond Q and everything that is in Q is sitting inside R.

So, these are what are called Venn diagrams. So, a Venn diagram is a very useful way to picturize a set and relationships between sets; is one set a subset of another, is one set not a subset of another and so on.

(Refer Slide Time: 12:20)



So, we often use Venn diagrams pictorially in order to represent sets. So, notice that every set is a subset of itself because remember the definition of a subset set that X ⊆Y, if every member of X is also a member of Y. So, since every element of X is also an element of X, trivially as a extreme case of this definition, every set is a subset of itself.

So, this in fact, gives us an important notion which looks obvious; but it is not so obvious, when are two sets equal. So, two sets are equal if and only if, they are actually the same set of elements. So, one way to check that two sets are equal is to check that everything in the first set belongs to the second set. So, X ⊆Y and everything in the second set belongs to the first set. So, Y ⊆X.

So, often this happens when we have two different ways of looking at the same set of objects. We have two different descriptions of the same set of objects and we want to check whether they are equal or not. Then, using the first description, we argue that everything which satisfies the first description also satisfies the second description and vice versa.

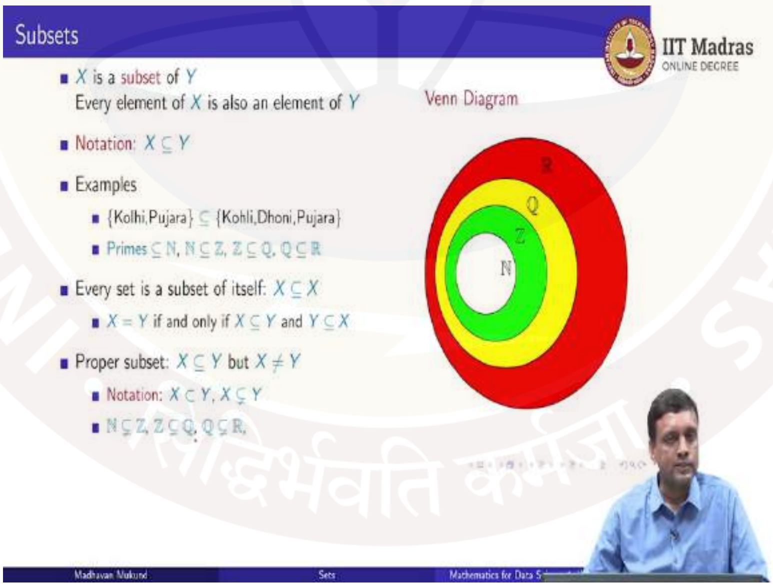
So, though this looks fairly obvious for finite sets, when it comes to infinite sets we have sometimes have to argue in an indirect way. So, this although it is an obvious statement is very important that X = Y provided X ⊆Y and Y ⊆X. So, sometimes we want to distinguish between the case, when X is really a proper subset of Y; that means, it does not include all of Y and that it is possibly equal to Y.

So, the subset equal to notation that we have right allows both. When we write X ⊆X, what we are saying is that it is a subset, but it is actually equal. So, we are allowing both cases. So, if you want to talk about proper subsets, sometimes we use a different notation.

So, we might either drop the equal to sign , just write the subset sign ⊂or we might explicitly like we said not element of right. So, we are saying that this is not equal to. So, we are dropping the equal to from below the subset.

Now, this is a bit dangerous. Second symbol this not equal to this is always correct. This is sometimes used both ways. So, you have to be bit careful when we look at books when you see the single subset without the equal to whether they mean subset and equal to or proper subset.

(Refer Slide Time: 14:45)



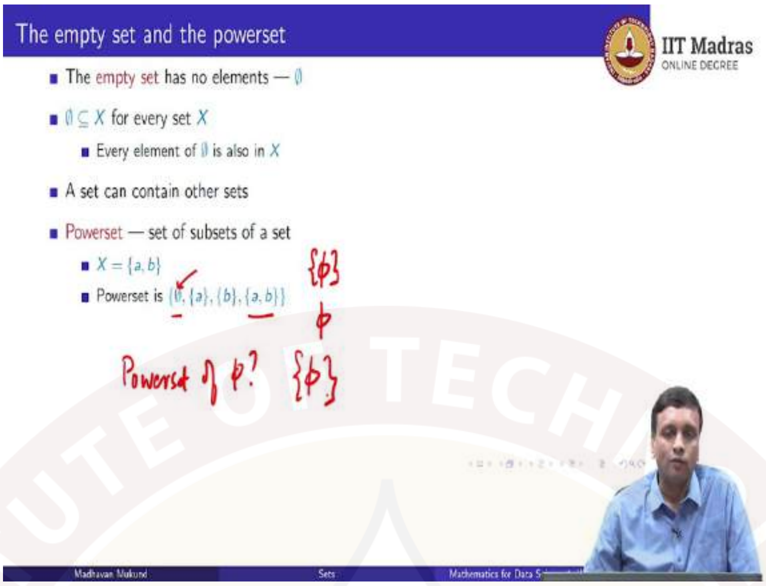
So, we know for instance that the natural numbers is a proper subset of the integers because the negative numbers are not there. Similarly, the integers are clearly a proper subset of the rationals and because the irrational numbers are not rational, the rational are a proper subset of the real numbers.

So, in most interesting cases, we will be looking at proper subsets. Sometimes, we will emphasize it by adding this cross against the equal to and sometimes, we will not and very often from context we will know whether we are talking about proper subsets or we are talking about subset which allow the full set.

(Refer Slide Time: 15:16)

Now, there is a very important set just like the 0 is very important in numbers, there is a very important set which is important set theory. It is the equivalent of 0. It is the set which has no elements. So, the set which has no elements is called the empty set and is written ∅. It is basically you can think of it as a 0 with a line across it. So, this Greek letter phi, symbolizes the empty set; so, it has no elements.

(Refer Slide Time: 15:44)



Now, what may not be very obvious is that this empty set is actually a subset of any set. Remember that we said that X ⊆ Y, if every element of X is also every is also an element of Y. Now, you might argue that an empty set has no elements. So, why is this true? Well, when we say for every and there is nothing in the set, then for every something is true right.

So, if I say that all birds with 3 legs have pink beaks, then this is actually true because we can imagine that there are no birds with three legs and therefore, every bird which actually has 3 legs will have a pink beak. But since, there are no birds with 3 legs this is actually true.

So, these kinds of vacuous statements as they are called will hold for sentences which use the word all where the set is empty. So, in particular, every element of the empty set because there are none. So, every element that could be in the empty set is also an any set X that you build. So, this empty set is a subset of every possible set. Now, though we have talked about elements and sets.

So, they are two different categories of objects. So, we have numbers and the numbers belong to a set of the type N or Q or R or Z; a set can clearly contain other sets. So, there is no restriction saying that the members of a set or the elements of a set must be some kind of discrete and indivisible objects.

So, one of the important sets of sets that we would like to look at is what is called the Powerset. So, we talked a subset. So, supposing we want to enumerate all the subsets. So, here is a two element set a comma b. So, what are all the subsets? Well, we already saw that the empty set is always a subset. So, that is one subset.

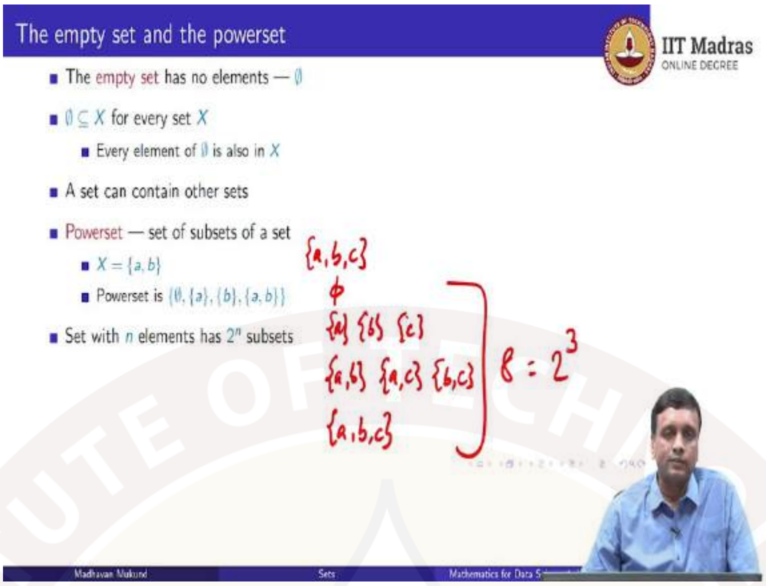
The set itself for any X, $X \subseteq X$. So, X equal to $\{a, b\}$. So, we have these two subsets which come just from the fact that empty is the subset of every set and the set itself is a subset. And then, we have two proper subsets either we can include the a and exclude the b or include the b and exclude the a. So, there are four subsets of X and if we group together these four subsets into a larger set, then we get the Powerset.

Now, notice that this itself is the set right. So, we do not write. So, this is different from this. The first is a set consisting of one element, namely the set consisting of the empty set. The lower thing is the empty set alone which is the set with no elements. So, if we put a brace around the empty set symbol, then we create a set with one element.

So, for instance, if you ask what is the power set of the empty set right. So, we know that the empty set has a power set which contains the empty set. So, we have at least one empty set as one element of the power set and there is nothing else right.

So, the full set is also the empty set, but if you duplicate an element, it is a same thing. So, in fact, the power set of the empty set is a set consisting of just one element, namely the empty set itself. So, just remember this, that the empty set on its own denotes a set with no elements, but an empty set with the brace around it is not the same thing. It is a set consisting of one element, namely the empty set.

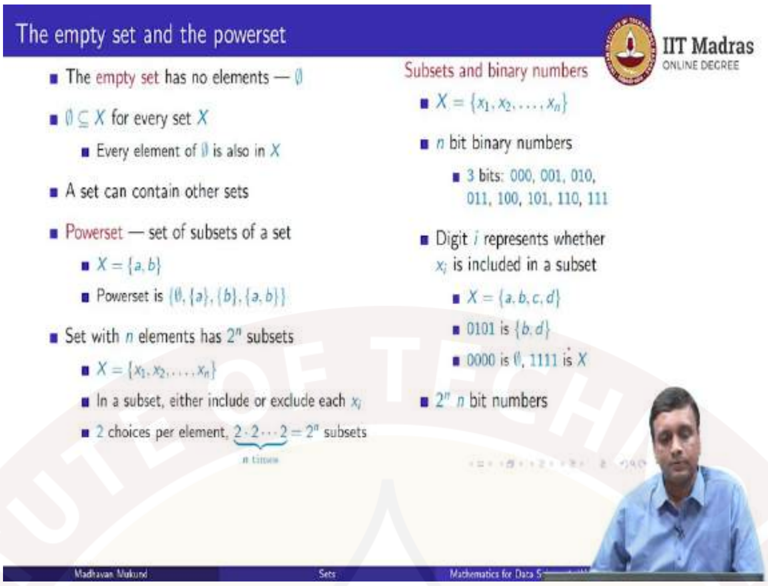
(Refer Slide Time: 19:03)



So, we saw above that if we have two elements, then the power set had four elements. So, in fact, one can generalize this and say that if we have n elements, then we would have 2^n subsets. So, for instance, if we had a, b, c right, then we would have 1 subset which is empty. We would have 3 subsets which are one element each and then, we would have 3 more subsets which are 2 elements each a, b a, c and b, c and finally, we would have the set itself right.

So, these are the only subsets. If you add these up, this is 8 which is 2^3. You can check that if you do it for a, b, c, d; then, you would have 2^4 , 16. So, why is it that a set with n element should have 2^nsubsets, no more no less?

(Refer Slide Time: 19:55)



So, here is one argument. Supposing we have n elements in the set. So, let us just call these without describing what they are specifically as x_1, x_2 up to x_n. So, we have n distinct elements x_1 to x_n. Remember these must be different because you cannot duplicate elements in the set. So, now, we want to construct a subset.

So, how do you construct a subset? Well for each element x_i, we have to either include the set include x_i in the subset or exclude x_i subset. So, we have to make a choice for each x_i right.

So, overall, we have to make n choices right. For each x_i we have to decide whether to include it or exclude it from the subset. So, we have two different choices for each elements. So, we have two ways to decide whether to do something with x_1, keep it or leave it; x_2 keep it or leave it. So, then we have two times two choices for x_1 and x_2 together; two times two choices for x_1, x_2, x_3 together.

So, in general, if we have n such choices where each choice involves two options, then we have 2 into 2 into 2, n times 2^n choices. So, each of these choices gives us different subset. So, whenever we make a different choice, we will either leave out i from the set or put an x_i. So, it will differ from the choice, where we do the other thing. So, each choice generates a separate subset. So, there are exactly 2^n subsets.

Here is another way of looking at subsets and getting to the same result. So, we can actually think of subsets in terms of binary numbers. So, let us again think of our n element set $x_1$to x_n right. So, now, supposing we look at n digit binary number. So, digit actually is comes from decimal. So, we say bit for binary digit. So, n bit binary number. So, remember in a binary number system, we have 0's and 1's and the place values represent powers of two.

So, we have the unit digits is units as usual. The next digit 2 to the power 0 is a is number of twos, number of fours, number of eights. So, it is like the decimal thing is in base 10. This is in base 2. So, now, if we look at n bit binary numbers, then for instance, if we look at 3 bit binary numbers, then we have 8 of them.

We can start with 0 0 0, then 0 0 1, 0 1 0 and so on up to 1 1 1 and again, the reason that there are 2 to the n, n bit numbers is because for each bit we can choose to put 0 or 1. So, we have two choices for the first bit, two choice for second bit and so on.

So, it is not surprising that an n bit binary number can represent 2 to the n different values from 0 to 2 to the n minus 1, if we think of them as numbers. Now, we are interested in n bit binary numbers as representing subsets. So, what we will look at is the ith bit and say that the ith bit represents the choice that we made.

If we chose to keep $x_i$in our subset, we will call it 0. If we chose to we will call it 1 for example. And if we choose to omit $x_i$from our set, we will call it 0. So, 0 represents the choice, where we leave outx_i; 1 represent the choice, where we keep x_i

So, supposing we have this four elements set a, b, c, d; then, if we look at the binary sequence or the bit sequence 0 1 0 1, the first 0 corresponds to a, so it says leave out a. The second 0 corresponds to c, so it says leave out c and for b and d we have put a 1. So, it says keep b and keep b. So, it says leave out a, keep b, leave out c, keep d. So, this 0 1 0 1 as a binary sequence corresponds to the set b comma d.

What does 0 0 0, the all 0 sequence say? The all 0 sequence says every $x_i$in the set is omitted from the subset. So, this is precisely the subset which is the empty set because it has no elements and what about the all 1 sequence? Well, the all 1 sequence says every x_i that we have is included in the final subset. So, this is the set itself. So, remember that these are the two extreme subsets; the empty set and the set itself and all the other ones come in between.

So, from this, we can see that every n bit number represents one sequence of choices. So, this gives us 2^nchoices because there are precisely2^n, n bit numbers. So, hopefully with this, you are now clear about the fact that any finite set with n elements has exactly 2^n subsets.

IIT Madras

ONLINE DEGREE

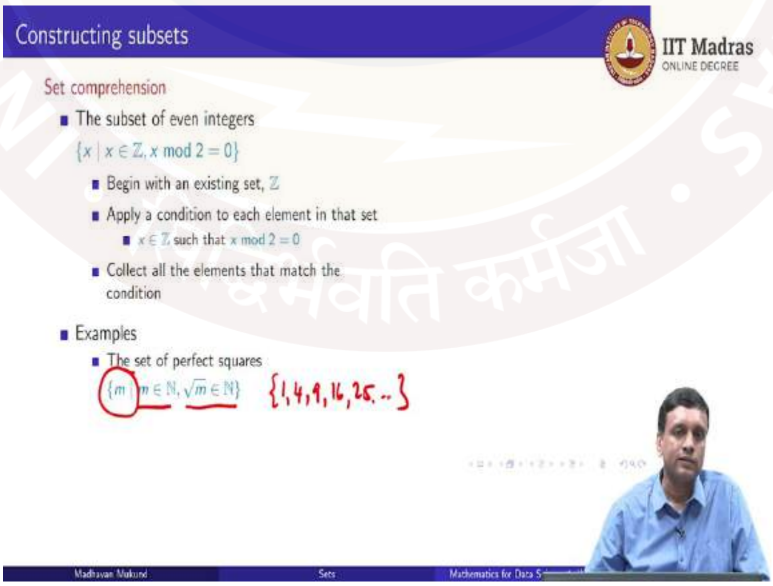
Mathematics for Data Science 1

Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Lecture - 05
Construction of Subsets and set operations

(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



Now, let us talk about subsets in the infinite context. So, how do we talk about subsets of the numbers in a precise way? So, this is something called set comprehension. So, this is just some jargon. So, a set comprehension is just a term used for this which we have sometimes seen and which we will now review. So, if we want to talk about the set of even integers, the set of even integers are those integers which when divided by 2 have a remainder 0. So, remember that the remainder is called mod. So, x mod 2 is the remainder when divided by 2.

So, if x mod 2 is 0 it means that when we divide x by 2 there is no remainder. So, any such x is an even number. So, this notation that we have written is actually the set comprehension notation. So, let us try and separate out the different parts and understand what is going on.

So, when we use set comprehension first of all we can only do set comprehension when we have a starting set. So, we have to begin with a set and construct a subset of that set. So, the first thing says that we want to take all x in Z. So, this here says that we are looking at elements from an existing set in this case this set is a set of integers. Then it says I want to take all elements and apply some condition to decide whether to keep that number or not. So, that is the second part of the right hand side.

So, we have the first part which tells us which set we are looking at the second part which tells us what condition we want. So, we are really saying x in Z such that x mod 2 is 0 and finally, with this bar and this left hand side we are saying collect together all the x which satisfy this. So, this overall this notation says collect all the x for which x is in Z such that x mod 2 is 0 or in other words x is even. So, this is set comprehension notation and this is formally how you define a subset of an infinite set. Remember that we cannot list out the elements in an infinite set.

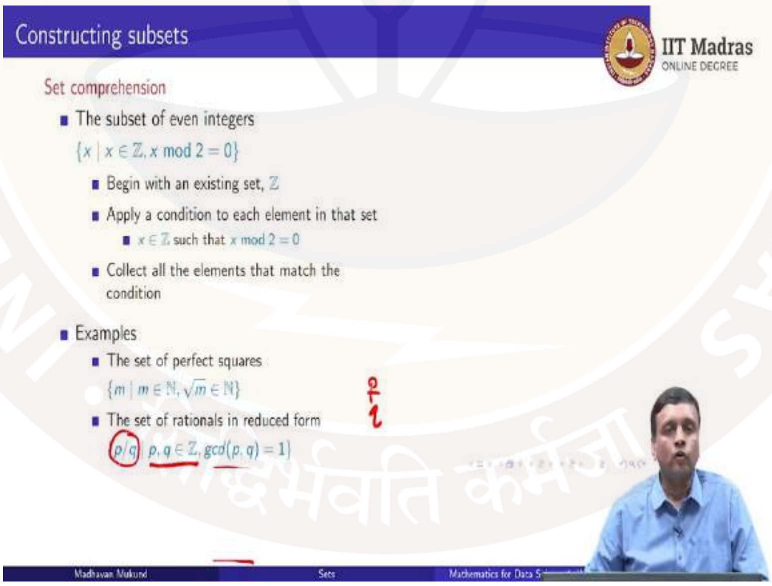
Now we assume that we already have a set like Z or N or Q or R for which we know what elements are. So, we do not have to describe how to pick out element we know what those elements are. What we are now giving is a description of how to choose elements which satisfy a given property. So, let us look at some more examples. So, for instance let us look at perfect squares.

So, remember that we said an integer is a perfect square if its square root is also an integer. So, for instance 25 is a perfect square because the square root is 5, but 26 is not a perfect square because there is no integer which multiplied by itself is 26. So, here is a set comprehension notation of the perfect square.

So, first of all remember square number has to be positive. We already discussed that negative numbers cannot be squares because when we multiply 2 numbers by them to the same numbered by itself the, 2 numbers will have the same sign. So, either it will be minus into minus is plus or it will be plus into plus is plus because the multiplication rule says that if the 2 numbers you are multiplying have the same sign the outcome is always positive. So, first of all we can only have positive numbers. So, instead of looking at integers, it suffices to look at the natural numbers.

So, we say for all m which are natural numbers such that the square root of m is also a natural number. So, this is that the square root of m also belongs to the set N collect all such m right. So, we are collecting all the m. So, this will give us if we write it out explicitly 1 will fall into this set, the next number that will fall into the set is 4, then 9 and then 16 and then 25 and so on right. So, the notation in blue is a succinct way of writing this informal infinite list which starts with 1 and goes on. So, we are pulling out the numbers from N one by one; checking if they are perfect squares and if so we are enumerating them.

(Refer Slide Time: 04:07)

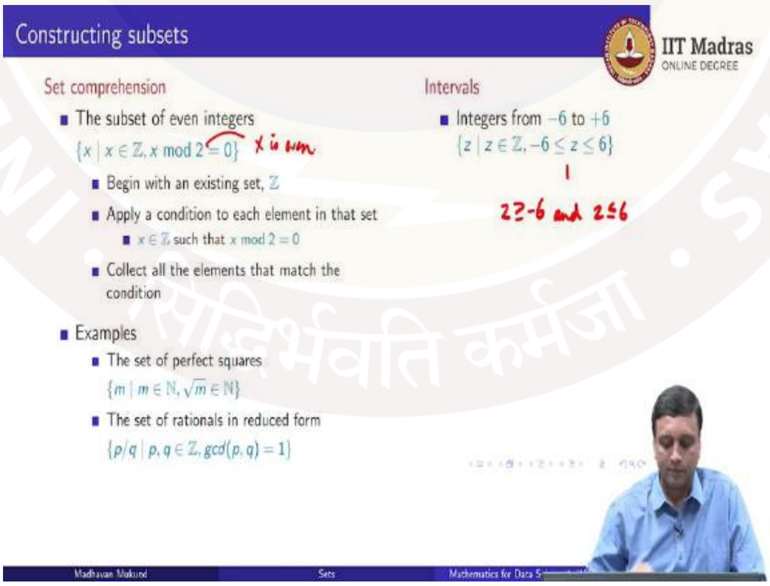


We also talked about rationals in reduced form. We said that there are many different ways of writing the same rational number because if we multiply the numerator and the denominator by the same quantity, the number we are representing does not change. And we use this fact in order to make denominators same when we did comparisons or arithmetic like addition and subtraction. So, what are the actual rationals in reduced form. So, this is a subset of the rationals. For example, $\frac{3}{5}$ is in reduced form, $\frac{6}{10}$ is not in reduced form because I can; cancel the 2 and get $\frac{3}{5}$.

So, if we want numbers and rationals in reduced form first of all we pick up any 2 numbers which are integers. Remember that a rational is actually a pair a numerator and a denominator which are integers. So, every rational looks like this $\frac{p}{q}$ right, but we do not want any such $\frac{p}{q}$. We want $\frac{p}{q}$ such that they do not have any common divisors other than 1. So, recall the gcd is the greatest common divisor; it is the largest number that divides both p and q and what we want is that p and q have no numbers which can be divided into them other than 1. And if the gcd of p and q is 1 then $\frac{p}{q}$ is a rational and it is in reduced form because the gcd is 1 right.

So, this is another example of set comprehension in order to define an interesting subset of the rationals.

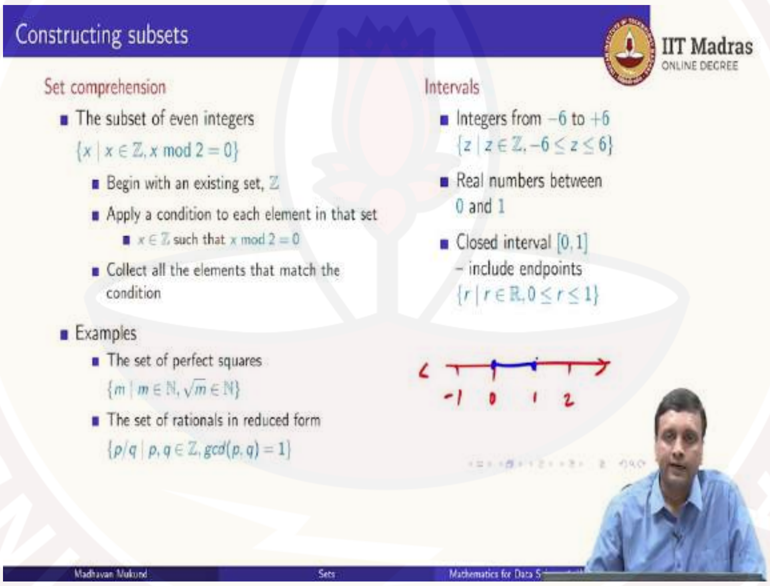
(Refer Slide Time: 05:34)



One of the things that we will often use with respect to numbers is to define intervals of numbers between something and something else. So, for instance if you are looking at the integers; we might want the integers from some lower limit to some upper limit. This for example, is an expression which describes the integers between -6 and +6 right. So, it says I want all z which belong to the set of integers such that z is above -6 greater than equal to -6 and less than or equal to 6. Now, we could split this for instance into two conditions. We could also say z is bigger than -6 and z is smaller than 6 and so on

So, the way in which we write this condition which applies to the thing may vary and all of them could be equivalent to each other. So, we will not be very pedantic about what syntax we used to write there. So, for instance in the previous case here, we could have just read written x is even instead of x mod 2 is 0 ok. So, we will not worry too much, but it is just that we have this format where we take the underlying set, we pick out all elements, make it satisfy condition. If it satisfy the condition, it belongs to the subset.

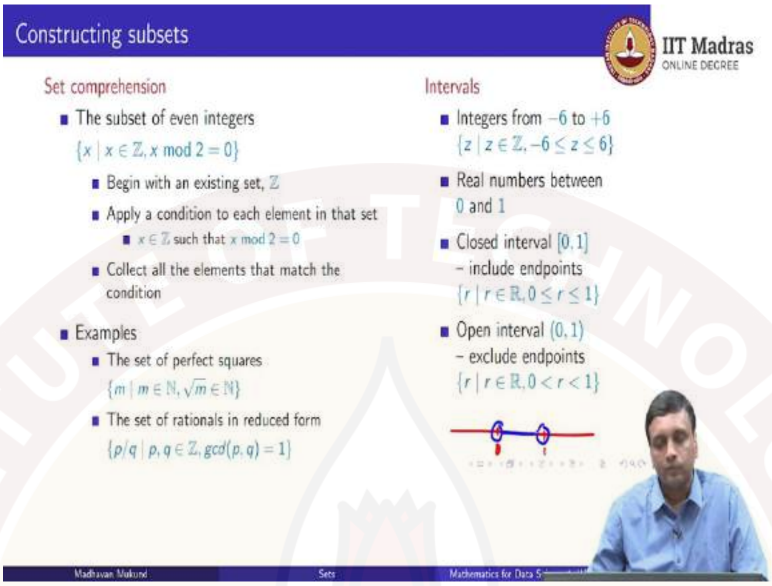
(Refer Slide Time: 06:44)



So, intervals are more interesting when we talk about real numbers and one of the intervals that we really often want to talk about is the interval between 0 and 1. So, 0 to 1 is quite interesting because we will often talk about probabilities for instance and probabilities range between 0 and 1. So, what can we do between 0 and 1? Well first of all we can take all the real numbers between 0 and 1 including both 0 and 1 and this is called the closed interval.

Closed interval means in this case, it includes the endpoints. So, if I draw this as a number line for instance. So, normally I have 0 1, 2 ,-1 and so on. So, this is my number line. So, then this closed interval says I want all the numbers from 0 to 1 including 0 and 1. So, this is my closed interval right. So, what we write is take all r in the set of reals such that $0 \leq r \leq 1$.

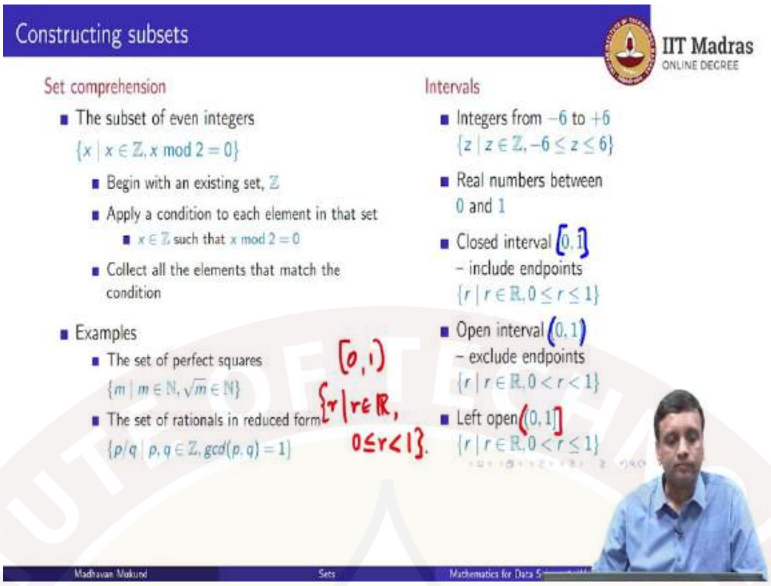
(Refer Slide Time: 07:46)



So, r must be between 0 and 1 it could be 0 and it could be 1. If we want to exclude the endpoints, then we get what is called an open interval and the way we draw an open interval; if we want to draw it in a pictorial way is to emphasize that the endpoints are missing by drawing a circle there.

So, we draw a circle to indicate that those are not included. So, if we so I have to fill in the circle corresponding to the endpoints that endpoint is included in our interval. If we do not fill it in it is not included, but formally it is just a set defined using set comprehension and whether it is open or closed depends on whether the inequality has an equal to or not whether it is strictly less than or it is less than equal to whether it is strictly greater than or greater than equal to.

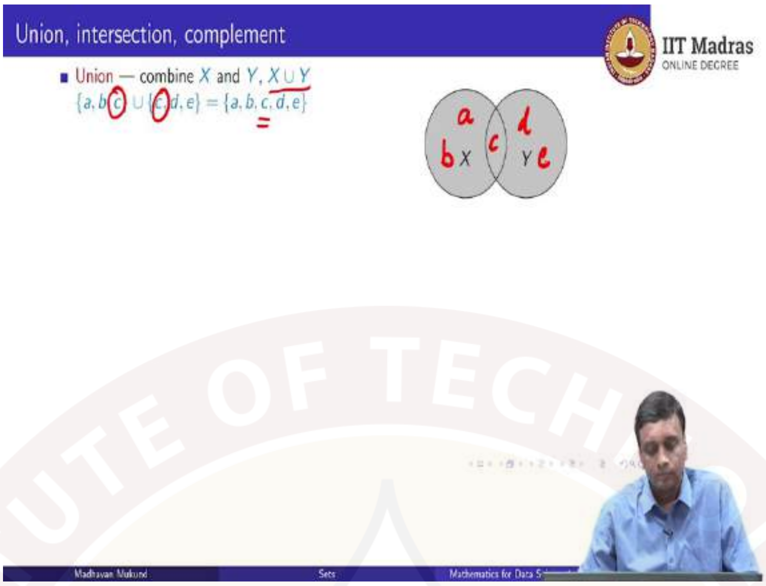
(Refer Slide Time: 08:29)



Now, there is nothing to stop us from including one endpoint and not including the other. So, we had an closed interval which had both endpoints, we had an open interval which had both endpoints missing. And we could say for instance that an interval is left open. So, it is all numbers between 0 and 1; it does not allow us to use 0, but 1 is included. So, in notation we will use this. So, the notice that we use this round bracket for open and we use the square bracket for closed. So, here obviously we will use a round bracket for the open end and a square bracket for the closed end. So, the left is open. So, we call this a left open interval.

So, left open interval has all numbers which are strictly bigger than 0, but less than equal to 1. So, correspondingly you could have a right open interval. And what would this be? This would be all the r such that r belongs to a set of reals. Now, $0 \leq r$ we are allowed to include 0, but we should not include 1 right. So, this is the right open interval. So, this will be an important part of many discussions. So, you should be aware of these intervals as representing sets of points in particular a subset of the reals which can be defined using set comprehension.

(Refer Slide Time: 09:41)



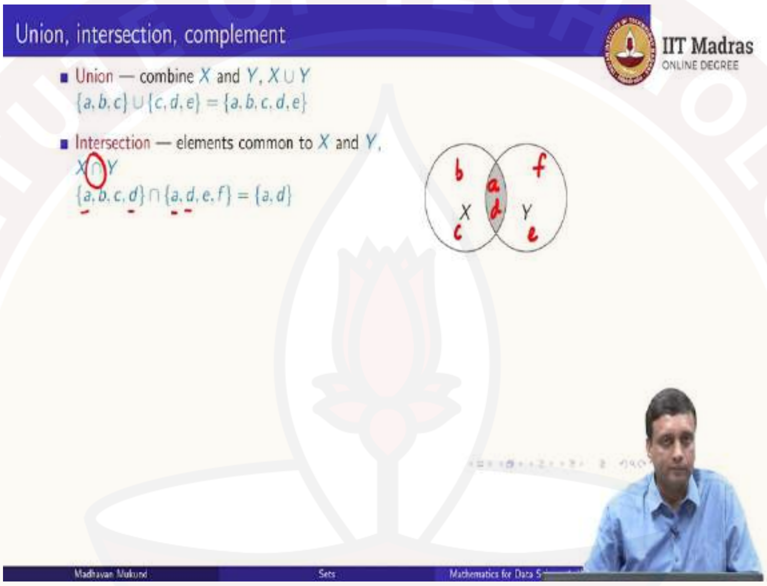
So, finally, let us look at some simple operations on sets which we are all familiar with. So, the first one is union. So, the union of two sets just combines them into a single set. So, suppose we have a, b, c as one set and we combine it with c, d, e then we get a single set. And notice that we have some elements which may appear in both sets and they appear only once in the final set because remember that a set has no duplicates right. So, in the union if we take sets which have some common elements across the two sets, they get represented exactly once in the final set.

So, therefore, the cardinality of the union will in general be less than the cardinality of the two sets put together. So, here we have two-three element sets, we take the union we get a five element set not a six element set because there are some elements which are common and the symbol for union is this ∪ right. So, $X \cup Y$ and if we go back to our Venn diagram; so, remember that we used when diagrams in order to informally look at sets and we talked about subsets. So, here we have a Venn diagram which represents the left hand side set is X, the right hand set is Y and the picture suggests that X is not a subset of Y and Y is not a subset of X, but there may be some overlap. So, this is the general case right.

Generally speaking if I give you two sets, there will be some elements which belong only to X some elements should belong only to Y and some which belong to both. So, this kind of a picture with two overlapping circles or ellipses is a particularly general picture of two sets represented as Venn diagrams. Even though we are not specifying what the elements are this is a picture. So, here for instance if we wanted to write out these elements in this particular set if you wanted to write we have a here, b here, c here, d here and e here.

So, what this means is that if we look at the circles a, b, c belongs to the left circle c, d, e belongs to the right circle, but we put c in the portion which is covered by both circles to indicate that it is in the common portion. So, this grey shaded area in this particular case represents the union of two sets.

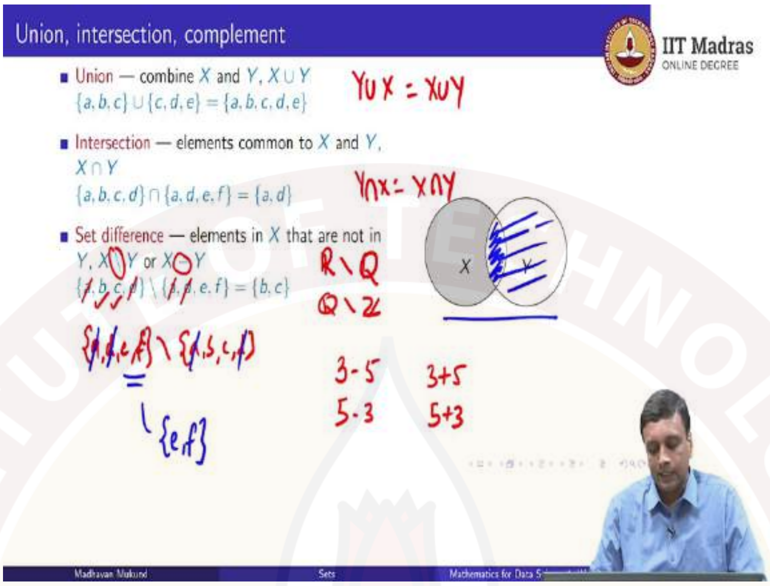
(Refer Slide Time: 11:43)



So, the corresponding thing which takes up only the elements which occur in both sets as you know is called intersection. So, intersection is written with the upside down version of the union sign right. So, X intersection Y is written like $X \cap Y$. So, here for instance we look at elements which are on both sides. So, we have a, b, c, d intersection a, d, e, f. So, a is common to both, b is not there on the right hand side, c is not there on the right hand side, d is common to both and if you go to the right hand side e is not there on the left hand side f is not there. So, only a and d are surviving intersection.

So, again if we draw this out as a Venn diagram on the right, the shaded portion which is the area which is overlapped by both the circles is the intersection. So, in this particular case we would write a here because it is in both b here. Notice the order is not important and in an Venn diagram if we actually put the elements the position is not important. So, I can put them anywhere and then I put e here and f there for instance. So, this is a pictorial representation of the two sets on the left. The shaded area corresponds to the intersection and the non-shaded portions are those which are in one set, but not in the other.

(Refer Slide Time: 12:53)



Another operation on sets is called set difference. So, in set difference we take two sets and we want to know what is there in the first set that is not there in the second set. So, for instance we want to know which are the real numbers which are not rational right. So, then we would write in this notation which are the real numbers which are not rational right or which are the rational numbers which are not integers. So, this is a common thing that we might want to do.

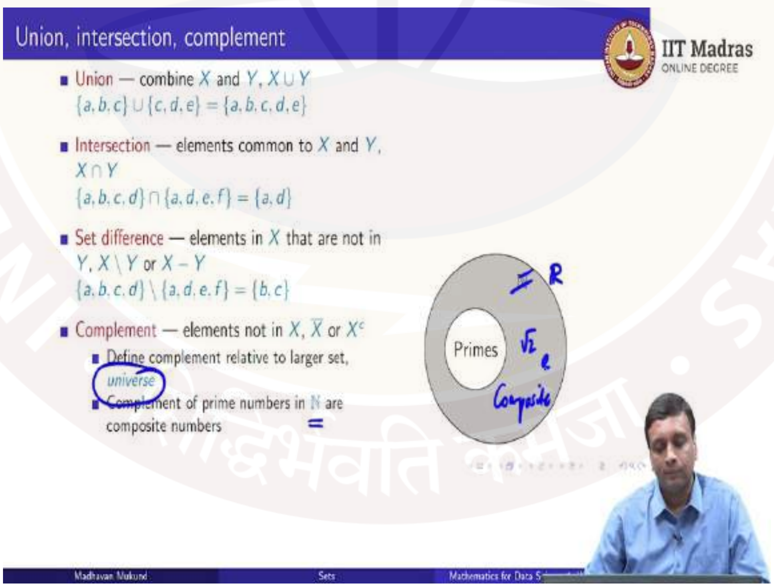
So, we write either this direct subtraction which is the normal minus sign or we write this back slash kind of notation \ to indicate the set difference. So, it is all elements in the first set which are not in the second set. So, here for instance if you look at the first set a is there, but a is also there in the second set. So, a is not counted, b is there, but b is not there in the second set. So, b is in the set difference c is there c is not there in the second set. So, c is in the set difference, but d for instance appears here. So, d is not counted.

So, here we have that the first set minus the second set has b and c because those are the 2 elements in the first set which are not in the second set. Now, this is like subtraction not symmetric in the sense that you know that 3 - 5 is not the same as 5 - 3 unlike 3 + 5 right. So, 3 + 5 is the same as 5 + 3, but 3 - 5 is not the same as 5 - 3. So, if I take union for instance, then Y ∪X = X ∪Y right and Y ∩X = X ∩Y because this it does not matter which side you take from.

Because finally, you are going to look at all elements which I has a common to both side or included in both sides. Now here if I take the reverse if I take a, d, e, f right and I subtract out the elements from a, b, c, d; then I would see that again a would disappear. So, the same elements disappeared because the common part is the same. So, a would disappear and d would disappear because these are the parts which are on both sides, but what survives now is e, f right.

So, when I do it in the other way around, I get the elements on the right hand side which are not on the left hand side. So, in the set difference the order of the sets in the expression matters. X \ Y is not the same as Y \ X just like in subtraction and here we have a picture right. So, this shows us this picture. It says that you take everything in X and you remove everything that all includes. So, in particular you remove all these elements which are in the intersection and that gives us X \ Y.

(Refer Slide Time: 15:19)



And finally, we often talk about the complement. We say those numbers that are not prime. So, those numbers that are not prime in particular are called composite numbers. So, composite number is defined to be a number which has factors other than 1 and same. So, any number which is not prime has more than 2 factors. So, such a number is called a composite number. So, clearly a number is either a prime or it is not a prime.

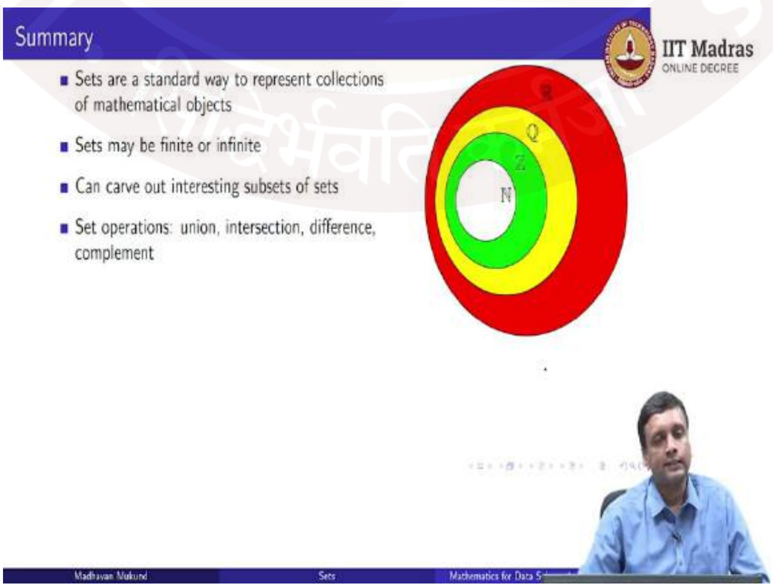
So, either it is prime or it is a composite. So, the composite numbers are disjoint from the primes and they are all the numbers that are not prime. So, this is what we mean by complement. Complement means the opposite side it means everything else, but complement is not very straightforward in set theory because complement with respect to what.

So, if I say numbers that are not prime, but I do not tell you in what set I am talking about this thing. If I look at complement in for example, in the reals; it will include all numbers like π and e and $\sqrt{2}$and so on and that is not what you mean right. When I say the complement of the primes; you are not thinking of rational numbers, irrational numbers and so on. You are thinking of integers or in particular you are talking about natural numbers which are not primes right. So, we would always want to define what is called a universe ok.

So, we need a universe with respect to which we are going to complement. So, if we say that the complement of prime numbers in the universe of natural numbers, then we get the composite numbers. So, when we say primes for instance we see this Venn diagram on the right, we see primes as a subset of the natural numbers. So, then the grey shaded area is all the composite numbers right. But if this was not this, but R then we would have various thing we would have $\sqrt{2}$, e and so on sitting here which is not what we intend.

So, whenever you use the word complement, you must make sure that you have specified complement with respect to what. What is the overall set with respect to which you are negating the set that you have and that is very important.

(Refer Slide Time: 17:18)



So, let us wrap up this lecture. So, we are all familiar with sets as an informal term which we have come across from school level and a set is a standard way to represent a collection of mathematical objects. So, it is very important to be familiar with the terminology of sets element of subset of and so on and also the notation the curly brace listing out the elements set comprehension and so on. So, sets may be finite or infinite. An infinite sets are actually quite tricky and interesting and most of the interesting sets that we are going to look at will be infinite because very often we will be thinking of sets in terms of numbers, but we will also be thinking in terms of finite things.

For instance we talked about we could talk about for instance a time table then we might want to know the set of stations at which the train stops or we might want to look at a shopping list and we might want to look at the set of items that the store has in its inventory. So, sets are a very useful way to talk about collections of objects infinite collections are important because numbers are infinite, but other finite collections are also important from a computational and data science point of view.

So, we saw that we have some useful notation like set comprehension which allows us to define subsets of infinite sets and we have these standard operations on sets like union, intersection, set difference and complement which allow us to take sets and combine them in many different ways. So, it is important that you get used to all these notions as I said because these notions are used implicitly throughout mathematics and these are not difficult notions is just a question of understanding the notation and understanding exactly what happens when you apply each of these operations.

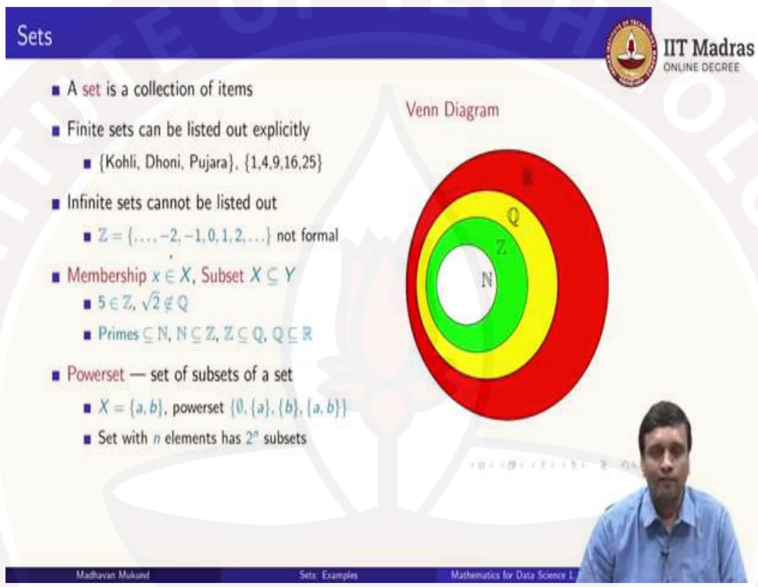
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute
Lecture- 5A
Sets: Examples

So, we have seen some definitions of Sets and some operations on them. So, let us look at more examples to get familiar with the notation and the terminology of sets.

(Refer Slide Time: 00:25)



So, remember that a set is a collection of items and when we write out a set, if it is a finite set, then we can just enumerate the items in the set by writing them within curly braces. On the other hand, if we have an infinite set, we really can not write out all the elements even though informally, we put dot, dot, dot to indicate a sequence, if that sequence is not very regular. For example, supposing it is a set of prime numbers, which does not have a clear pattern, then it is not very easy to represent it explicitly like this.

So, we saw that there will be another notation called set comprehension that we will come to. But, before that let us talk about the two basic relationships between sets and membership of a set. So, membership is denoted by this element of relation. So, small x typically denotes a member or an element of a set, and capital X usually denotes a set itself. So, when we write x belongs to X like this, what we mean is the element x belongs to X.

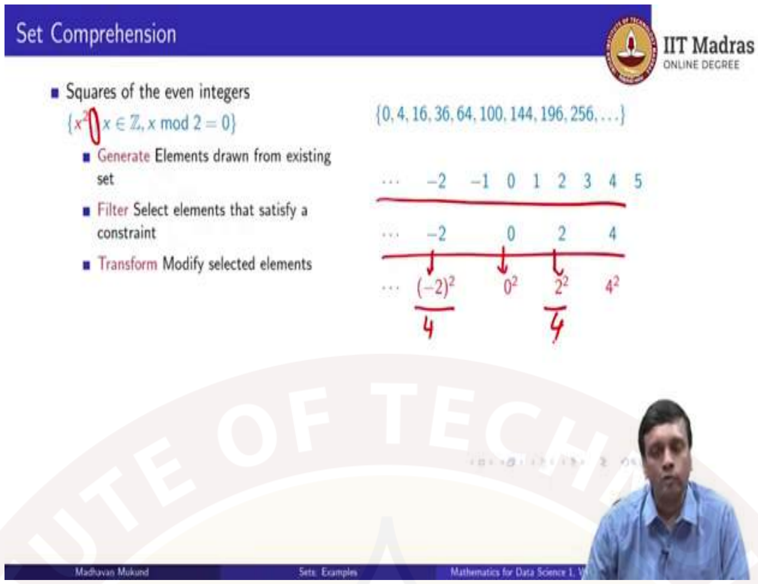
So, for example, the number 5 belongs to set of integers, and $\sqrt{2}$ does not belong to the set of rationals for instance. Subset on the other hand, says that one set is included in another set, so everything that belongs to X belongs to Y. So, for instance, all the prime numbers are natural numbers, so the primes are a subset of the naturals. Every natural number is an integer, so the natural numbers are a subset of the integers. Similarly, the integers are a subset of the rationals and the rationals are subset of the reals.

And we draw this using these Venn diagrams where we draw these ovals or circles or boxes representing the extent of a set, it is a picture of a set. And then depending on whether a box intersects another box or it sits inside a box, it indicates whether the first set is a subset of the other one or they overlap and so on. So, in this particular diagram which also has colors, we have indicated the subset relationship between the different types of numbers that we have studied, the naturals, the integers, the rationals, and the reals.

And finally, one very useful thing to know about sets is the power set. So, when we take a set, we can enumerate all its subsets. So, remember that we have just defined a subset. And in particular, we have this special subset called the empty set, which is a subset of every set. The empty set has no elements in it, but we needed it for technical reasons, and it is a subset of every set. And in addition, if you have 2 elements set $\{a,b\}$, then the subsets could be the individual elements, the set containing a and the set containing b or the entire set itself.

So once again, just like the empty set is a subset of everything the set itself is also a subset of itself. And we argued that for a finite set with n elements, we will always have 2^n subsets. So here for instance we have 2 elements, so we have $2^2 = 4$ subsets. So, this is just a review of what we have already seen.

(Refer Slide Time: 03:06)



Now, let us look at this set comprehension notation, which is what we said we would use when we have to describe infinite sets which cannot be written down explicitly. So, this was a typical example. So, supposing we want to write down the set of all the squares of the even integers. So, the even integers are -2, +2, 0 as an even, -4, +4, and so on. But if we square them, then we know that $(-2)^2$ is the same as 2^2 is 4.

So, the set on the right which is written in this informal dot, dot, dot notation has 0^2, 2^2, 4^2, 6^2 and so on. So, how would we write this out? Well, this is that notation on the left, which says that we take every x which belongs to the integers, check whether it is even, whether (x mod 2 = 0), and then square it. So, let us just break this up into parts so that we remember exactly what is happening.

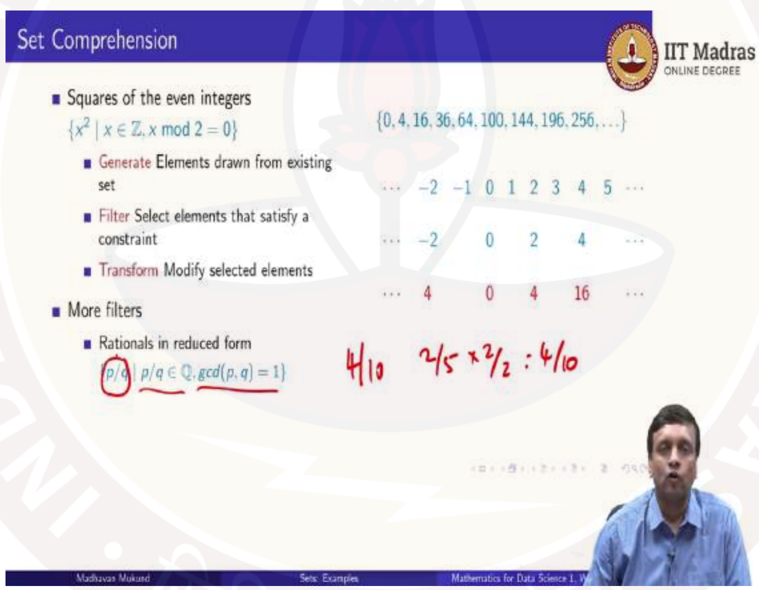
So first, in the set comprehension notation, we have a generator. A generator says that we are taking elements from an existing set, so we can only build new sets from old sets. So, we already have a set of integers, and we are going to try out every integer in the set, so, that is what x element of Z says, is try every x ∈ Z, so Z generates this set. Now, all the x's that come out are not interesting to us. So, we want to filter out those that are useful, that satisfy a given property.

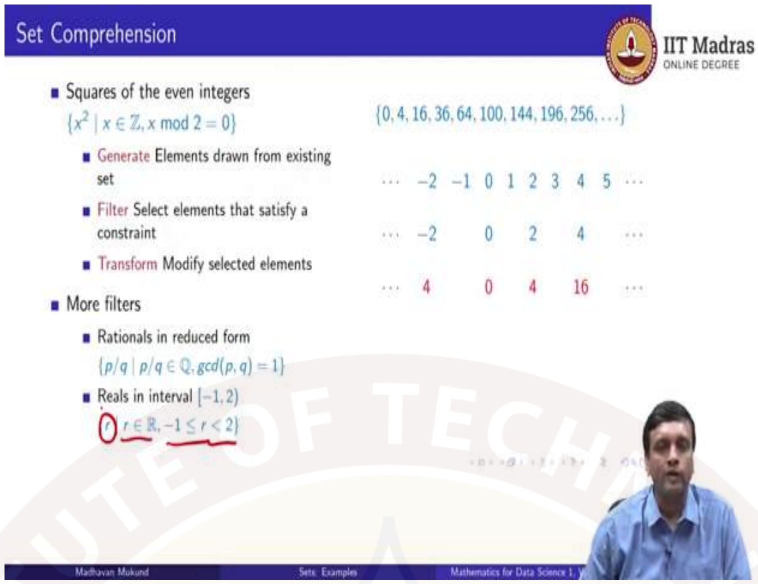
In this case, the property that we are looking for is that the number is even. So, we want those x which come out of Z through the generator, such that they satisfy the property that x when divided by 2 has remainder 0, which is the property that x is even. And finally, with these x, we do not want to keep them as they are, we want to transform them. So, on the left-hand side of this vertical bar, this is the left-hand side are the actual elements of the set.

The elements of the set are generated right, then filtered through some conditions, which rule out the ones we do not want, and when the ones we keep, we can transform them. In this case, we want the squares, we do not want the even numbers, we want their squares. So, if you look on the right, this is what happened.

So, when we started the generating process, we had all the integers, then we filtered out, and we got only the even ones, and now we transform them. So, for each even number, we produced its square. And now in this process, you will notice that $(-2)^2 = 4$ and $2^2 = 4$ also. So, some elements will disappear because we do not keep duplicates.

(Refer Slide Time: 05:35)





So finally, when we go through this, we end up with this sequence 4, 0, 4. And then in this, we will throw away all the elements on the left, and we get the number sequence on the top. So, this is how set comprehension works.

So, we can write filters in many different ways as long as it is unambiguous, we will not be very particular about the language we use so long as there is no question about what we mean. So, for instance, we looked at this example, we have rational numbers, but some rational numbers are not in reduced form. For instance, if I write $\frac{4}{10}$, then I should actually think of this as $\frac{2}{5}$, because it is $\frac{2}{5} \times \frac{2}{2} = \frac{4}{10}$.

So, I have actually multiplied both the numerator and the denominator by 2, to go from $\frac{2}{5}$ to $\frac{4}{10}$, but it is the same rational number. So, we want the numerator and the denominator to not have any common divisors, which is the same as saying that their greatest common divisor is 1, that is nothing other than 1 divides both the top and the bottom of the fraction. So, if we take all the rational numbers, so we generate all the possible rational numbers $\frac{p}{q}$, which belong to the set of rationals.

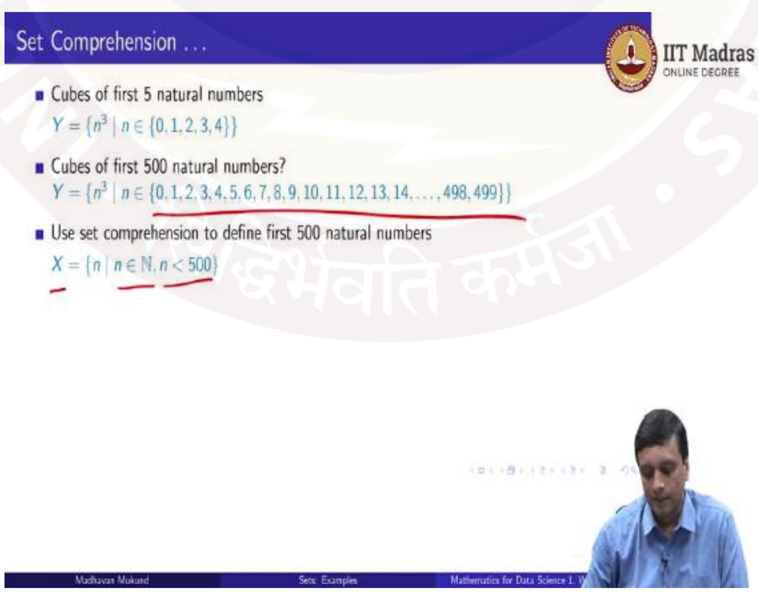
Then, we filter out those which have no common divisor between the numerator and the denominator and we keep only those, we do not transform it in any way, we just keep it here. So, here the transformation is just to keep it as it is, this is sometimes called the identity transformation. The identity just takes an input and produces the output the same as the input.

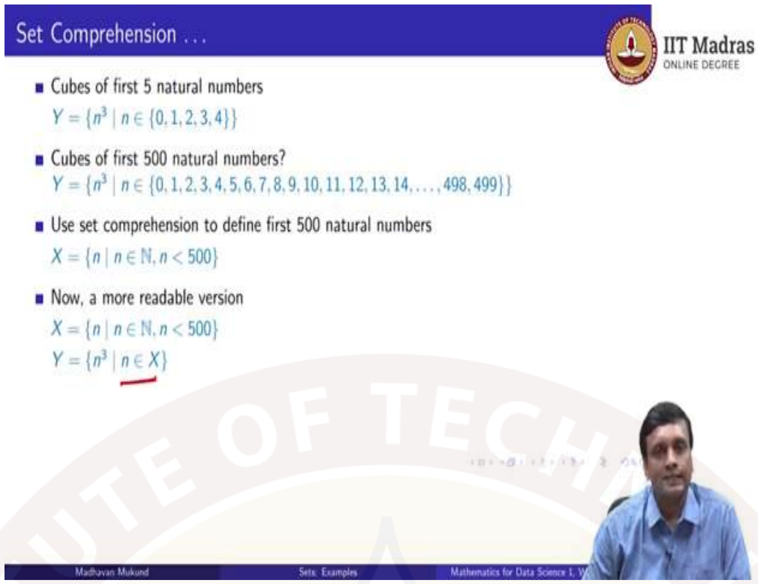
So, this gives the set of rationals in reduced form. So, here we have used a function, GCD. Even though we have not formally defined it here, we assume that people understand what GCD means. So, this is what we mean by saying that we can write the filter in any reasonable way, as long as people understand what it means.

Another example, we looked at are intervals. So, here we want the real numbers, which start from -1 including -1, and go up to but not including 2. So, in this case, we will use less than and less than equal to, so we will take all the reals. So, we take every possible real number, but we are not interested in all the reals, so we check whether it is greater than or equal to -1, so it includes -1 and everything above it. So, it cuts off everything which is strictly smaller than -1.

But, we also do not want it to cross 2, so we stop below 2, so it should be greater than equal to -1 or and less than 2 and if so, again we keep it without any transformation. And this notation on the top, the square bracket, and round bracket are indications of whether the endpoint is included or not. So, -1 endpoint is included, +2 endpoint is not included.

(Refer Slide Time: 08:14)





So, let us see why we would actually want set comprehension notation. So, let us extend our first example of squares of the even numbers to cubes. So, cube is just a number multiplied by itself 3 times. So, square is x×x, a cube is x×x×x, 3 times. So, if we want the cubes of the first 5 natural numbers, we can write it out explicitly like this, we can take this generator and generate the first 5 natural numbers as 0, 1, 2, 3, 4. Remember that, in our terminology natural numbers start with 0, even though in some books, you will find that natural numbers start with 1, we always assume natural numbers start with 0.

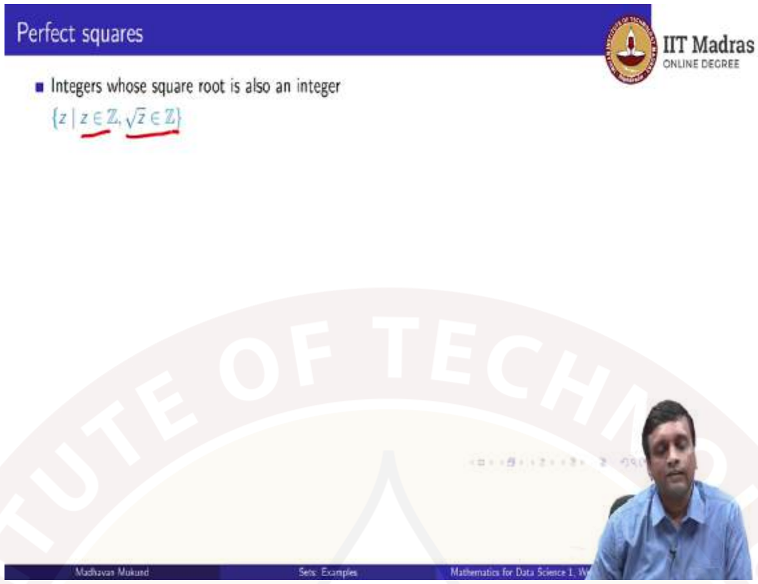
So, the first 5 natural numbers are 0, 1, 2, 3, 4. So, this is our generator, take every n in this and transform it to n^3 without doing any further filtering. We are not asking for the first 5 odd numbers or the first 5 numbers which have some other property, we just take, taking the first 5 numbers. Now, imagine that we change this question to the first 500 natural numbers, then though we can write it out explicitly, it is rather tedious.

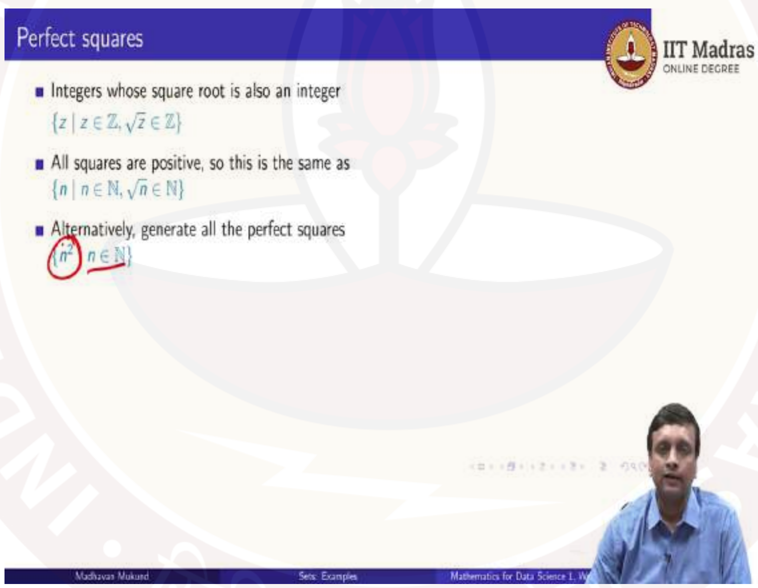
So, we have to replace the small list of 5 numbers by a long list of 500 numbers. And remember, we are not really allowed to write dot, dot, dot if we are being mathematically precise. So, we actually have to physically write out these 500 numbers. Now, this is not terribly convenient. On the other hand, we can define the first 500 numbers quite easily using set comprehension.

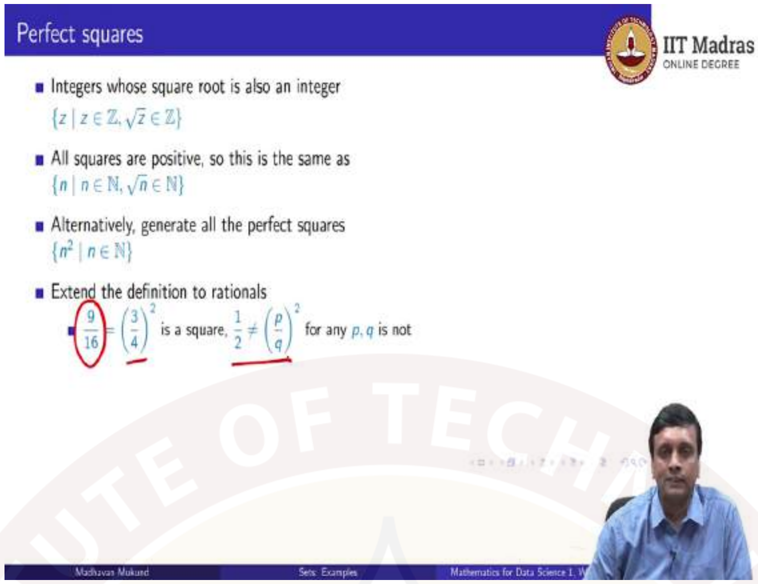
So, we can say, give me all the natural numbers, that is the generator, but restrict the natural number to be less than 500. So, remember that the first 500 natural numbers are going to be 0 up to 499. So now, this says that this set X is actually this long set here which we have written explicitly. So, we have replaced that very long and tedious expression by a much more compact expression, which captures exactly the same set. So now, we can have a much more readable version of these cubes of the first 500 natural numbers.

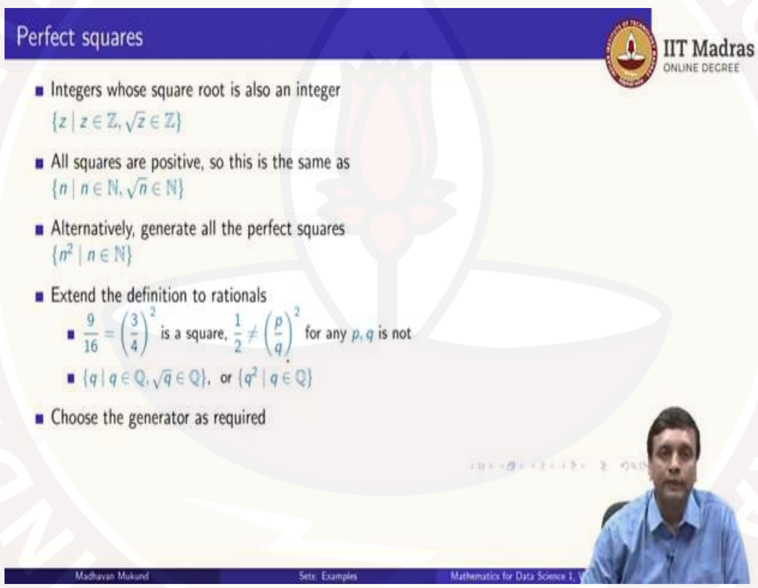
As an intermediate set, we generate the set X, set $X = \{n \mid n \in N, n < 500\}$. And then we take this as the generator and we say, okay, take every n which belongs to this X. So now, we know that x is restricted to 0 to 499. And then, take the cubes of these numbers, so we get n cubed in this range. So, this is one other use of set comprehension, which is to make our definitions more readable and understandable and less tedious to write.

(Refer Slide Time: 10:37)









So, let us look at one more round of examples. So, we saw this before, we talked about perfect squares. So, we said that some integers are squares of other integers and some integers are not squares. In particular, those which are not squares, their square roots are actually irrational. We proved for instance, in our supplementary lecture, that the $\sqrt{2}$ is irrational. So, perfect square is an integer such that its square root is also an integer. So, this is what this says, give me all the integers, which satisfy the condition that their square root is also an integer.

So, the square root of small z also belongs to a set of integers, give me all set Z and call it a perfect square. Now, notice that the square must be positive, we have already discussed this because you multiply 2 negative numbers, you get a positive number, you multiply 2 positive numbers, you again get a positive number. So, in fact, a perfect square must always be non-negative, it could be 0.

So, we could as well assume that the target set is generated by the set of natural numbers. And that, we are only interested in the positive square root, so remember that, 4 has 2 square roots, the number 4 is either $(-2)\times(-2)$, or 2×2, but it is sufficient to know that one of its square roots is an integer because the other one will just be the same with a minus sign.

So, we can as well define the same set of perfect squares in terms of the natural numbers, we generate all the natural numbers whose square roots are also a natural numbers. Now, we can turn this around and replace the filter by a condition. So, we know that every natural number when it is squared will give us a natural number. So, all the perfect squares will be generated in that form, take a natural number, square it. So, instead of looking for those numbers whose square root is a natural number, we can just take every natural number and square it.

So, we just generate all the natural numbers and without filtering them, we just take the output square. So, this also gives us 0, 1^2, 2^2, 3^2 and so on. So, these are all different ways of writing the same thing. In one case, we replace the generating set from integers to natural numbers because of the property of perfect squares. In another case, we transformed the filter into a transformation. So, instead of putting a condition on the numbers that we are generating, we took all the numbers and then squared them to get the actual perfect squares.

Now we could extend the notion of perfect squares to other sets of numbers. For instance, rationals can also admit a definition of a perfect square, so a rational will be a perfect square if it is a square of another rational. In particular, a rational could be an integer, but we will now integers can also be above and below the line, so we could have $\frac{9}{16}$, for instance as a rational number, which is $\frac{3^2}{4^2}$, so, $\frac{3}{4} \times \frac{3}{4} = \frac{9}{16}$.

So, we might want to say that this is a perfect square in the world of rationals. And not everything is a perfect square because since $\sqrt{2}$ cannot be represented a rational, it is easy to check that half cannot be represented with a form $(\frac{p}{q})^2$. So, not every rational in this sense is a perfect square, some are, some are not. So, we can again change the definition above and replace Z and N by Q and get a reasonable definition of perfect squares in a different domain of numbers.

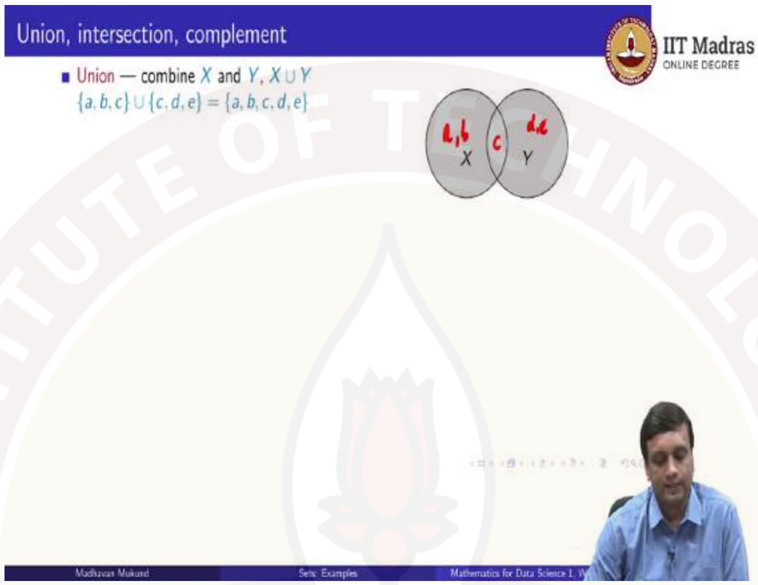
So, we can say give me all the rationals q such that $\sqrt{q}$ is also a rational. Or using the second form, we can say take all the rationals and square them. So, take every q, which is a rational q and give me q^2. So, this says that depending on how you choose the generator, you might generate the same set, or you might generate a different set. So, it is important to specify all the parts of a set comprehension correctly, so that there is no ambiguity and so that you get the set that you mean to get.

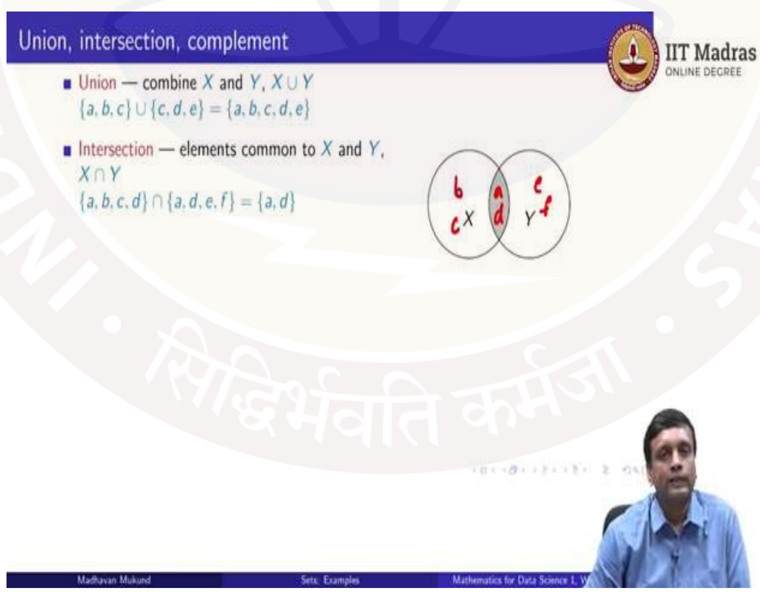
IIT Madras

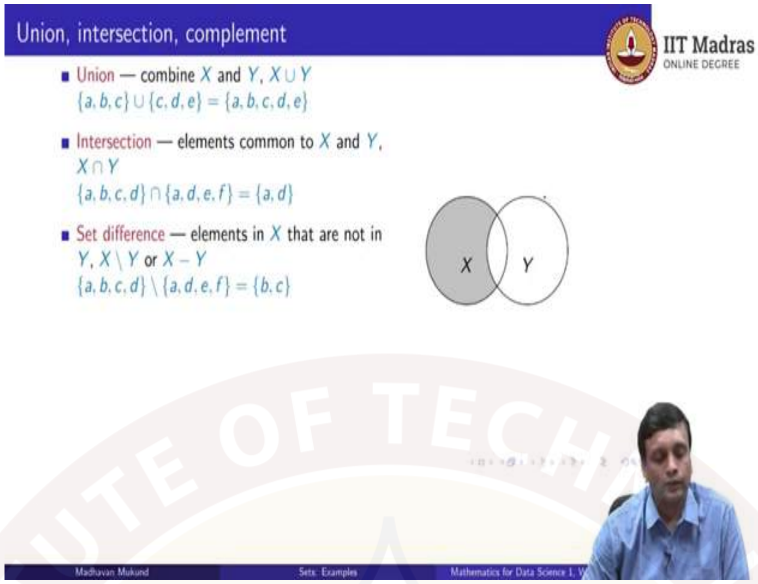
ONLINE DEGREE

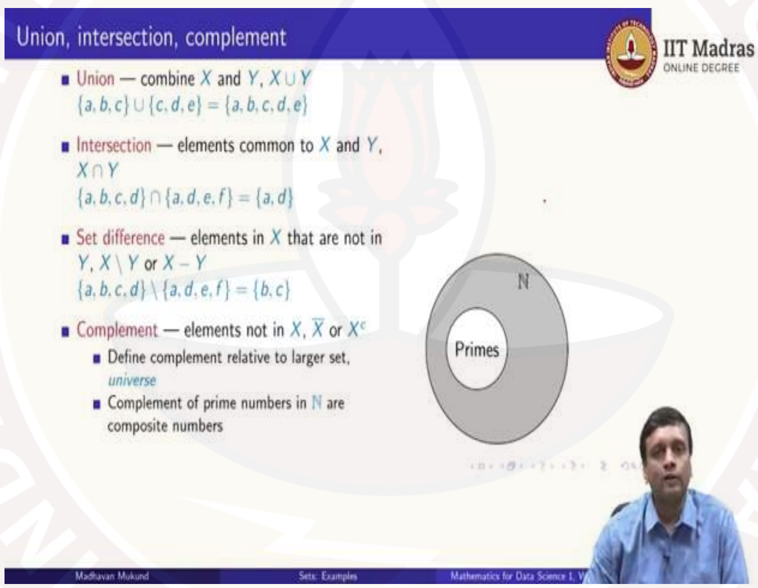
Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute
Lecture- 5B
Examples of Set Operations and Counting Problems

(Refer Slide Time: 00:15)









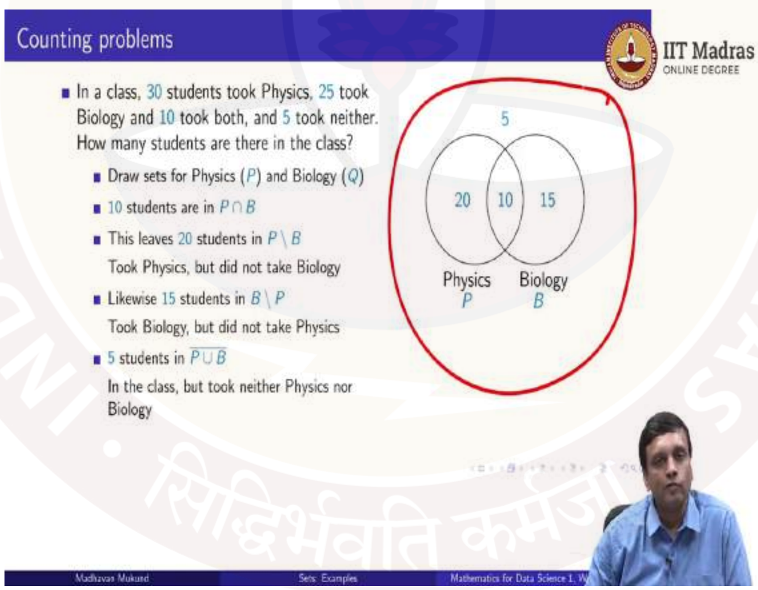
So, the other operations that we saw on sets are union, intersection, and complement, which we represented using Venn diagrams as shown here. So, the union takes two sets and combines them and removes the duplicates. So, the overlapping part between the two diagrams represents the common element. So, in this case, we would have this common element c over here, and then we have had a and b over here, and we would have d and e over here because d and e belongs only to Y, a, b belongs only to X.

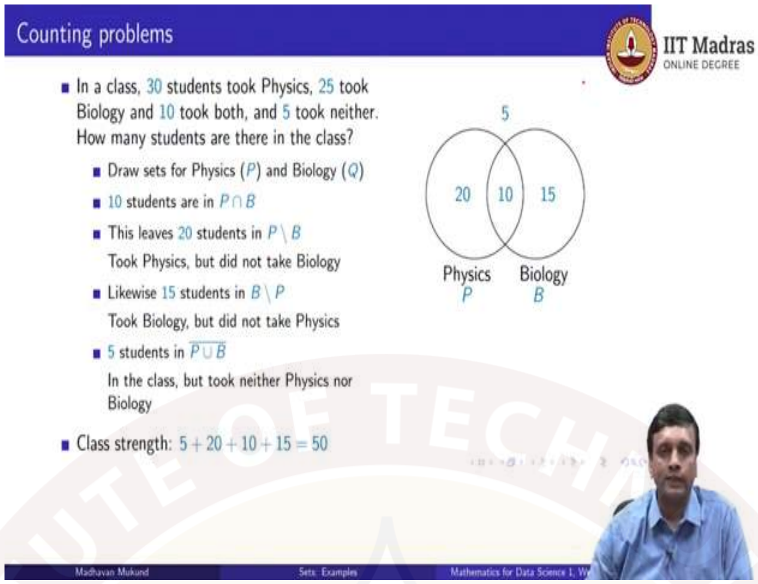
Conversely, we can take only those things which are common to the two and in this case, we have a and d over here, and then we know that b and c are only on the left and e and f are only on the right. So, the intersection tells us the elements which are common to the two sets. Set difference tells us what is on the left but not on the right.

And finally, the complement can be taken if we have an overall universe that is a full set to talk about. And with respect to that set, we can ask which elements are not in the set that we are looking at. So for instance, if we are looking at the natural numbers as a whole, the primes are a subset of the natural numbers, the complement of the primes are all those natural numbers that are not primes.

Now, remember that the complement matters, because if we take the complement of the primes, for example, with respect to the real numbers, we will get all sorts of other numbers which are not even integers. So, whenever we define the complement, we need to define the universe that we are talking about.

(Refer Slide Time: 01:37)





So, this leads us to a class of problems that you might come across, which can be solved nicely using these Venn diagrams. So, these Venn diagrams are not just pretty pictures, they are actually useful ways to reason about these problems. So, here is a typical problem that you could come across. So, you have a class in which 30 students have taken physics, and 25 students have taken biology, but 10 have actually taken both physics and biology, but there are also 5 who have taken neither of these two subjects.

So, these are the facts that are given to you. There are 30 students taking physics, 25 taken biology, 10 take both, 5 take neither, the question is how many students are there in the class. So, using Venn diagram notation, you can represent the fact that there are two sets of students, those who take physics and those who take biology by representing them by two sets, say P and Q. And we know that some take both, so, there is an intersection so these two sets overlap.

Now, from the data that we are given, we know that the overlap has 10 students, so we can write a number 10 in the intersection to indicate that there are 10 students who take physics and take biology. Now, we know that 30 students took physics overall and we have already accounted for 10 of them because they have all taken both physics and biology. So, there are 20 students who have taken physics, but have not taken biology.

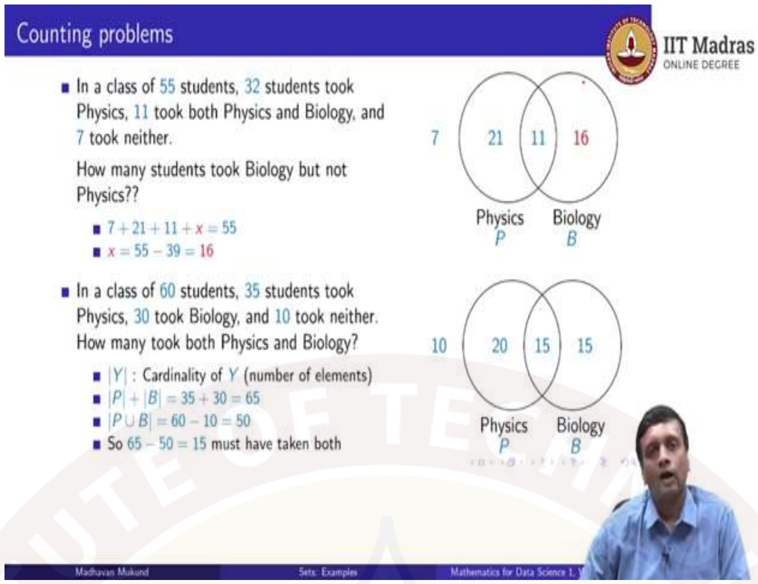
So, this in our set notation is the set difference, it is the difference between P and B, how many elements are in P which are not in B, how many students have taken physics who have not taken biology. And we have a symmetric thing on the right hand side. So, we know that there are 10 students who have taken both but 25 students take biology. So, there must be 15 students who are in $B \setminus P$, these are students who took biology and did not take physics.

So, in this way, we can populate the three regions of the Venn diagram with numbers indicating how many students are in each of these regions at 10 in the intersection, 20 on the left hand side, 15 on the right hand side. But, this is not the entire class because with respect to the entire class we have to take the number who are in the complement, those who have taken neither physics nor biology, and these are 5 students who are outside $P \cup B$.

Now, technically one should draw outside this the complement to indicate the entire class but just for convenience, I have not done that, but this entire complement outside this contains 5 elements. So, totally from this, we can see that there are 4 regions of interest. We have the $P \setminus B$ region physics but not biology, we have the $B \setminus P$ region, biology but not physics, we have the $P \cap B$ region taking both, and we have the complement, taking neither, and these are all disjoint from each other.

So, now if we add up the students across these, we get the exact number of students. And in this case it is 5 + 20 + 10 + 15 = 50. So, there are actually 55 students taking physics and biology together, but the total class strength is only 50. And actually only 45 students are taking these subjects because 5 are not taken either of them.

(Refer Slide Time: 04:43)



So, here is a variation where the data for the problem is given in a different way. So now, you are told the class strength 55, you are told that 32 students took physics and of them 11 took physics and biology and you are also told that 7 took neither. So, the question is how many took biology but not physics. So again, we draw a Venn diagram and from the previous question, we know that we can put 11 in the intersection, because that is the number who took both.

And since there are 32 who took physics, we can subtract out these 11 and say that $P\setminus B$ is 21 and in the complement, we have 7. So, the question now is how many are in $B\setminus P$, which I have marked by x, but now we know the total. So, we know that the four numbers together, add up to the total which is 55. So, $7 + 21 + 11 + x = 55$. So, if we solve for x, we get that $x = 16$. So, we can deduce that 16 students have taken biology but not physics in this situation.

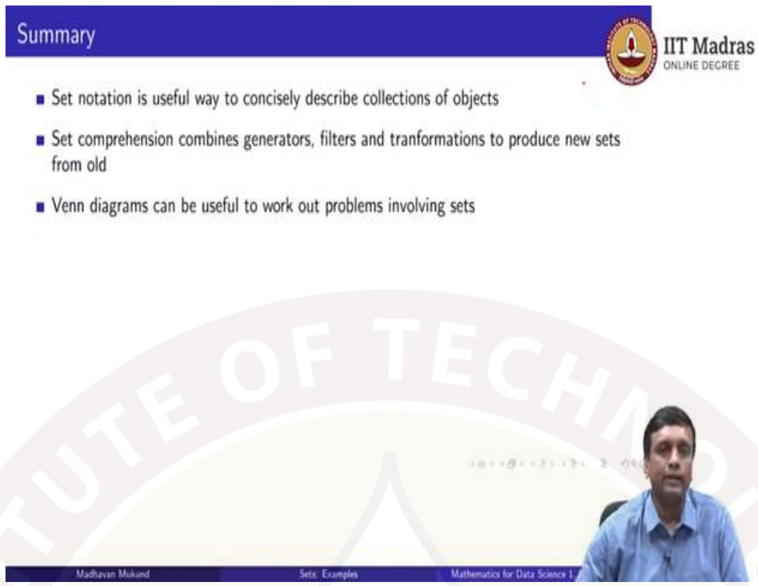
So here is yet another version of this. So, we have 60 students in the class. So again, we know the total number of students in the class, we are told that 35 students took biology, 35 students took physics, and 30 took biology, and 10 took neither. So now, we are trying to calculate the intersection, how many people took both subjects. So again, let us use this notation which we introduced when we first introduced sets.

So, this perpendicular bar on the side of a set indicates the size of the set. So, this is the cardinality of a set, cardinality is the number of elements, so the cardinality of Y is denoted by putting Y inside these bars. So, what we are told is that the set P has cardinality 35. That is a set of students who have taken physics overall, including those who have taken both, set B has 30 and 35 plus 30, there are 65 students who have taken in the union, of these I mean, have taken these together.

But we also know that there are 60 students in the class of whom 10 have taken neither. So, the actual union has only 50 elements. So, there are totally 65 people who are taking either physics or biology or both, but this total number actually spans only 50 students, so some of them must be taking both and are being counted twice. So, this must be the difference of the two.

So, 15 of these people must be counted twice, otherwise we would not have this mismatch. So, if we draw the diagram for this, this is how it comes out. We have 15, that we calculated for the intersection by taking the total number, realizing the 10 have taken neither, and then computing the difference between the number who should have taken both the subjects from those who are actually registered for either one or both of the subjects. So, these are three different examples using Venn diagrams to indicate how you can solve these kinds of counting problems.

(Refer Slide Time: 07:37)



So, to summarize, we use set notation because it is a very useful and precise way to talk about collections of objects. And if we use it nicely, it is also a concise way sometimes instead of writing out a long sequence of values, we can actually describe it using a condition. So, this is typically where we use set comprehension.

So, remember that set comprehension has three parts, some of which may not be used. So, you always have a generator, a basic set from which you are creating new sets, you may have a filter which takes out some elements from the generated set and throws them away and keeps only those that satisfy the condition.

And finally, you may have a transformation which takes these filtered elements and does something to make them into the elements that you want, for example, the squares of the even numbers. And then we also saw that Venn diagrams are not just simple doodles that you draw to indicate sets, Venn diagrams can actually be very useful for calculating properties about sets, especially numerical problems about sets. So, it is important to be able to draw the proper Venn diagram to indicate which groups of sets overlap, how they overlap, and which parts are empty, and so on.

IIT Madras

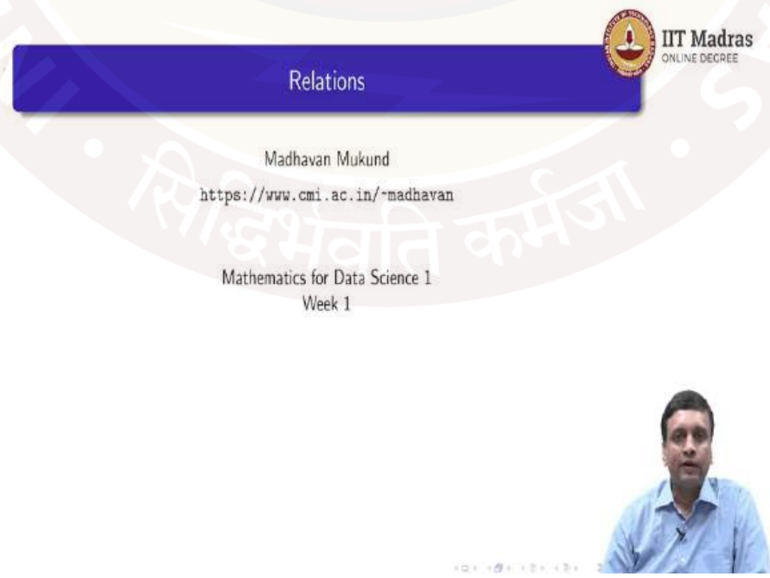
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture – 06
Relations

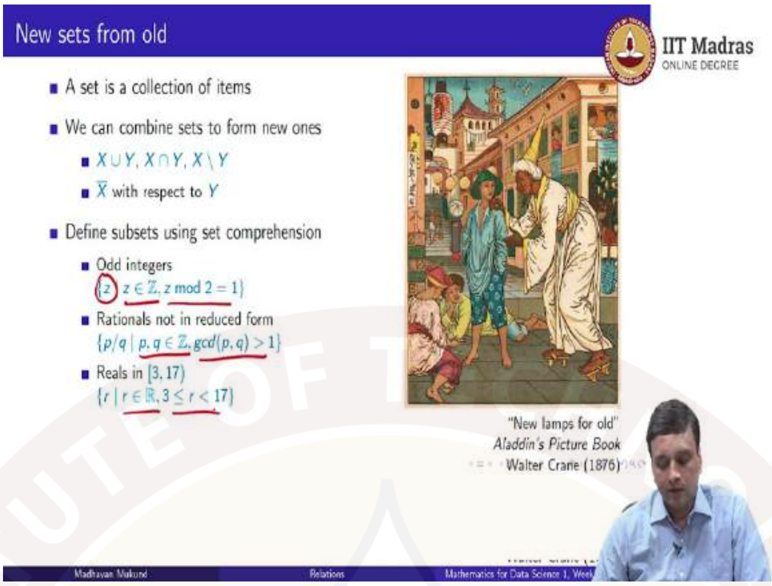
(Refer Slide Time: 00:05)

(Refer Slide Time: 00:14)



So, we have seen Sets, now let us move on to Relations.

(Refer Slide Time: 00:17)

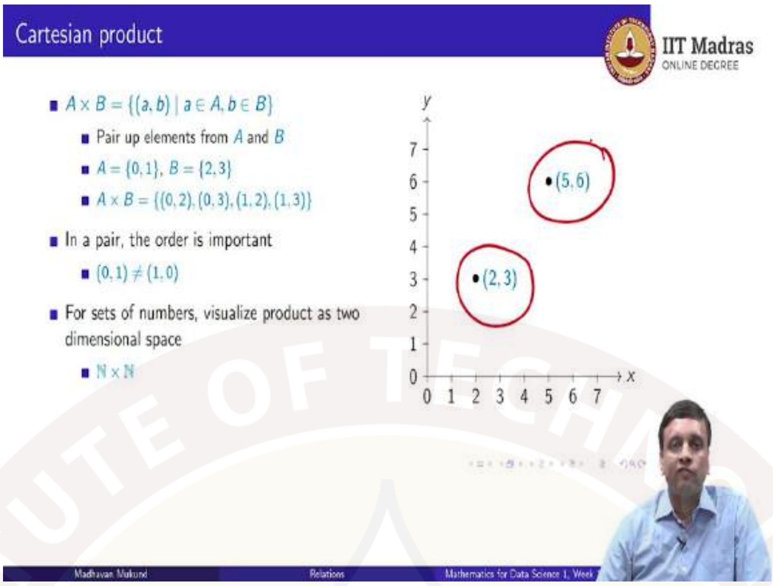


As we saw a set is a collection of items and we can construct new sets from old sets. So, we can take unions combine two sets into one. We can take intersections, take the common elements. We can take the difference that is take the elements of X which are not in Y and if we define the universe with respect to which we are working, we can define the complement those elements that are not in X.

Now, in general we are interested in carving out subsets of a set and so, we use the set comprehension notation. So, what this does is it takes a base set and takes elements of that set, then it applies some condition those elements we are interested in and then it collects them all together. So, we can take all the integers which are divisible by 2 or not divisible by 2 in this case, so we get the odd one; so, those where the remainder is 1.

Or we can take all fractions in which the numerator and the denominator have no common divisor or we can take for instance the real numbers which lie in an interval with [3,17).

(Refer Slide Time: 01:17)



So, now, we will see a new way to combine sets to form new sets and this is called the Cartesian product. And, in the Cartesian product basically what we do is we take two sets and we take one element from each and form a pair. So, A × B as it is called is the set of all pairs which we write with this normal bracket notation (a,b) such that the first element a comes from the big the set A and the second element comes from the set B.

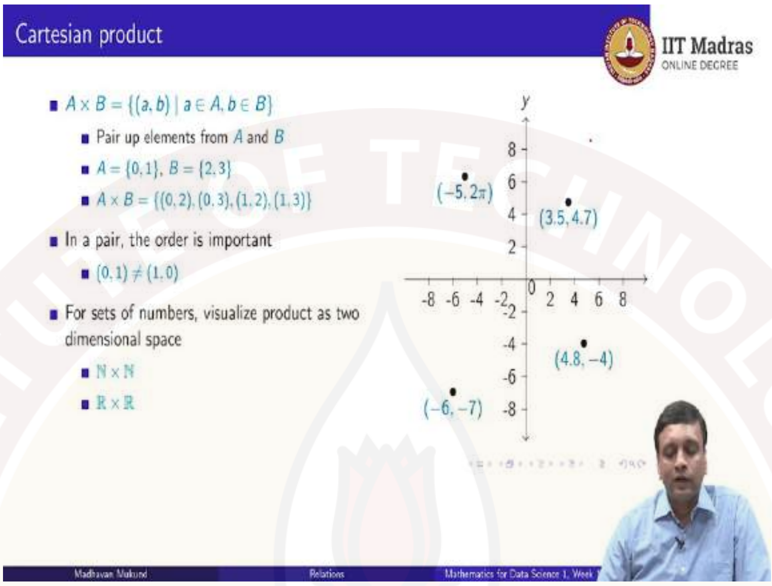
So, for instance, if A is the set {0, 1} and B is a set {2, 3} then all possible pairs we can form in the Cartesian product a 0 combined with 2. So, (0, 2), (0 ,3) and then 1 combined with 2, (1 , 2) and (1,3). So, we have four possible pairs.

Now, in sets we said that the order of the element is not important, but of course, when we are doing this kind of a pairing, then we know that the left set comes from the left part of the product and the right element comes from the right part of the product. So, for example, (0 ,1) is not equal to (1 , 0). So, here we have to respect the order when we talk about a pair.

Now, if we have sets of numbers right, then we normally visualize the product as a space which we draw familiarly as a graph. So, for instance if we take $N \times N$ then we draw $N \times N$ as this grid, where on the x-axis you have one copy of N, on the y-axis you have another copy of N. And, for example, if you want to look at the pair (2 , 3), then such that the x-coordinate is 2 and the y-coordinate is 3 and you get this point and similarly, if you look at the point (5 ,6); you get this point right.

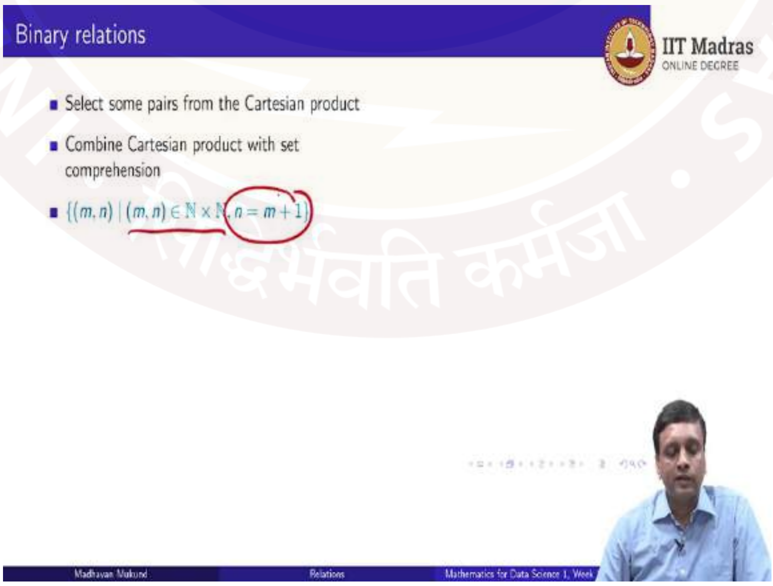
So, you can take the first coordinate plot it on the x-axis; take the second coordinate plot it on the y-axis and where those two points meet in the grid is the point that we are interested in. So, this is one way of visualizing a binary relation on numbers.

(Refer Slide Time: 03:00)



And, we can do the same thing if you are using say the reals, in which case the grid points that we are going to plot will have real coordinates and not just natural number coordinates.

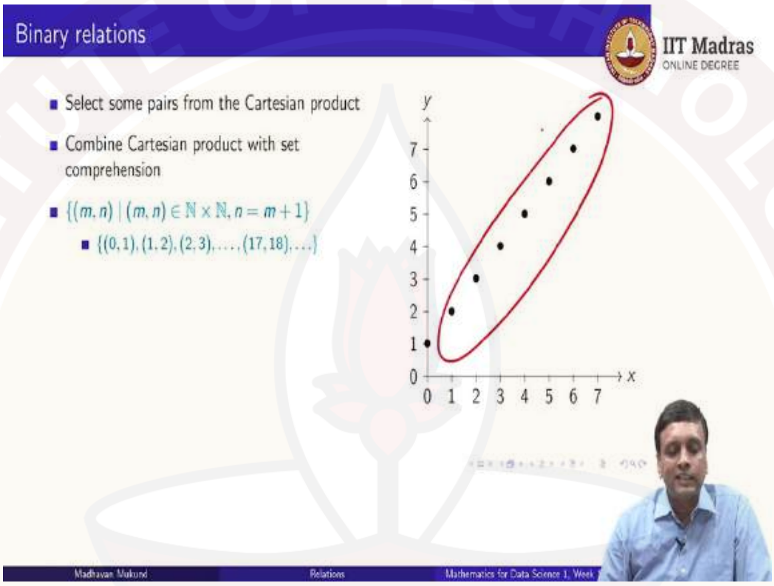
(Refer Slide Time: 03:10)



So, now we have this Cartesian product which consists of all possible pairs of the two sets and as we did with set comprehension we might want to pick out some of these sets some of these pairs and this is what we call a relation.

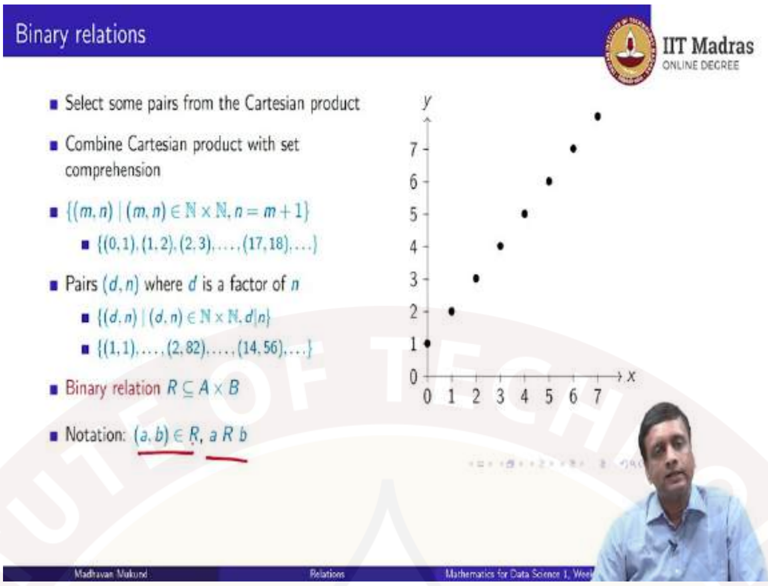
So, we combine this Cartesian product operation with set comprehension. So, for instance, we can take all pairs of numbers which are natural numbers (m ,n), but we want to insist that the second number is 1 plus the first number.

(Refer Slide Time: 03:40)



So, we get for instance (0, 1) because the second number one is 0 + 1; (2, 3) because 3 is 2 + 1, (17, 18) and so on. And, if we plot these points alone on the right then we get these so, we get a subset of the overall points and this these points satisfied this set comprehension condition.

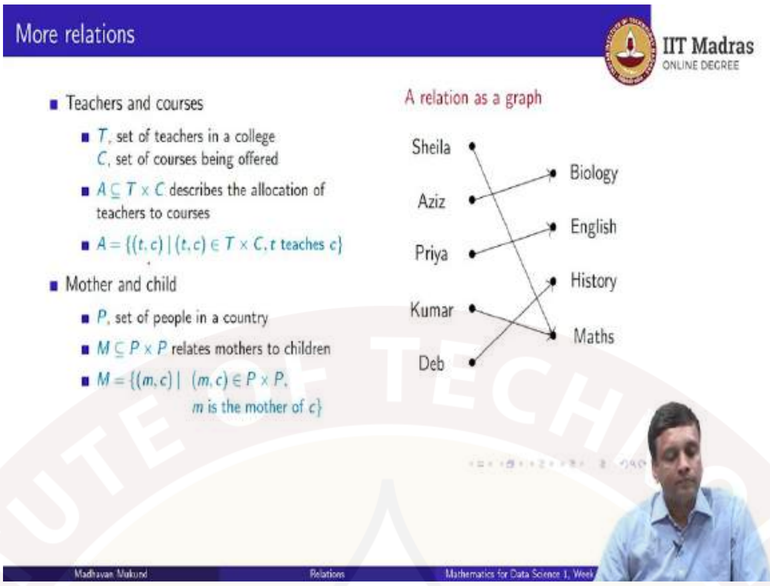
(Refer Slide Time: 03:59)



Another example would be pairs again of natural numbers (d,n), where d is a factor of n. Remember, d is a factor of n means that if I divide n by d, I get remainder 0. So, for instance 2 is a factor of 82, 14 is a factor of 56. So, these will be points in our relation. So, this is what is called a binary relation. So, formally it is a subset of the product. So, we take the Cartesian product all possible pairs and then we apply some kind of a condition which filters out the pairs of interest to us and it gives us therefore, a subset of pairs and this is what we call a relation.

Now, to denote the pairs that belonged to the relation either we can give the name of the relation as a set and say that (a,b) ∈R or sometimes to say that a is related to b, we use R as a kind of operator. We say a is related by R to b and so, we write a R b. So, these are two notations which you might see in different books and they mean exactly the same thing.

(Refer Slide Time: 04:57)



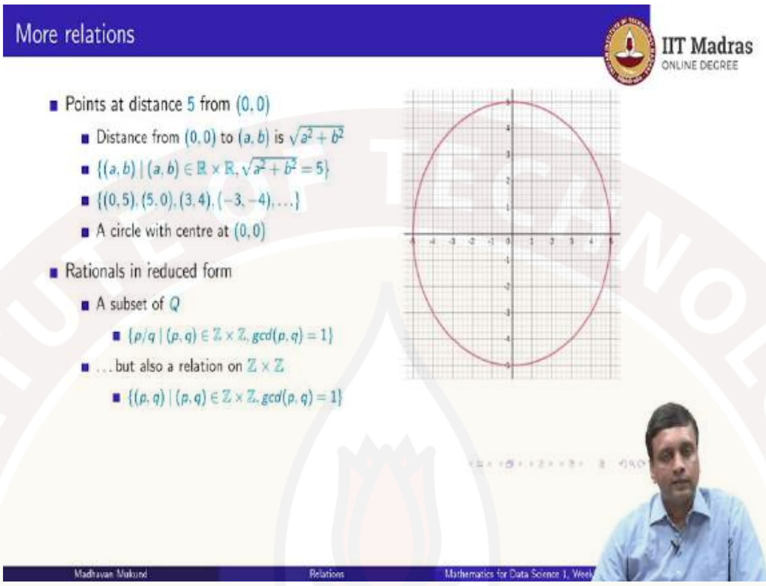
So, let us let us look at some other examples of relations outside the numbers. So, supposing you have a school in which there are some teachers and some courses to be taught. So, T is the set of teachers; C is the set of courses that are being offered in this term, then you need to describe which teachers are teaching which courses. So, we would have an allocation relation A which is a subset of all possible pairs T × C.

So, every teacher and principle could be teaching every course, but of course, this is not normally the case. We do not have all teachers teaching all courses, we have some teachers teaching some courses. So, we would specifically say take every pair of possible teacher course pairs, then we take out those were precisely the teacher T is actually teaching the course C and we collect those together to form this allocation relation.

So, here is a different graphical way of describing a relation not in terms of the grid and the graph that we have learned when we do graphs in school. So, this is also called a graph, but this is a graph in which we have some nodes representing the elements on the set. So, on the left hand side we have five teachers, on the right hand side we have four courses and the arrows from the left hand side to the right hand side connect the pairs which are in the relation. So, we see that Kumar teaches maths; Deb teaches history and so on.

So, this is a useful way of visualizing relations on finite sets and we will see this often as we go along. Another example of a similar type of a relation is that between a parent and a child specifically let us look at mothers and children. So, if we have a set of people in a country, then we can take the set of all pairs of people and then isolate from that pairs in which the first element of the pair is the mother of the second element. So, we want (m,c) which belongs to P × P such that m is the mother of c.

(Refer Slide Time: 06:49)



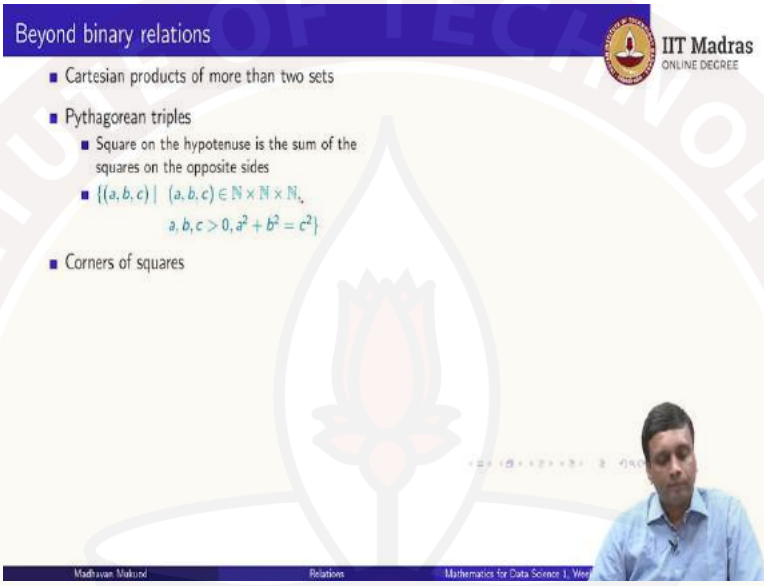
So, let us go back to numbers. So, supposing we want to plot all points which are in R × R which are at a distance 5 from (0, 0) which is normally called the origin. So, one thing you need to know for this we probably you should have learned this at some point is that if I take a point (a ,b) and calculate its difference from (0 , 0). So, this is calculated using the Pythagoras theorem and it comes out to be $\sqrt{a^2+b^2}$.

So, in other words, the relation we are looking for in this case is all (a,b) whose distance from (0 , 0) is 5. So, all (a, b) and R cross R such that $\sqrt{a^2+b^2}$is equal to 5. So, here are some of the points (0,5) for instance you can see (0, 5) is there because the sum is 0 plus 25 and the square root of that is 5. (3 ,4) is there because 3 squared is 9, 4 squared is 16, 9 plus 16 is 25, square root to 25 is again 5.

So, interestingly these points if we plot every such point in R × R which satisfies this actually defines a circle of radius 5 with center at (0 , 0). So, relations can define interesting geometric shapes and very often we do deal with geometric shapes in this relational form because it is easier to manipulate than looking at pictures. Now, depending on how we are going to view a relation, we can look at it in different ways.

So, remember that we looked at rationals in reduced form. So, we said that a rational in reduced form has p / q such that p and q are integers and the gcd is 1 right; that means, that they do not have a greatest common divisor other than 1. But, we can also think of this as a relation on integers itself. We want all pairs of integers. So, every rational is really a pair of integers, the numerator and the denominator and we want every pair of integers where the gcd is 1, that is, there is no common divisor.

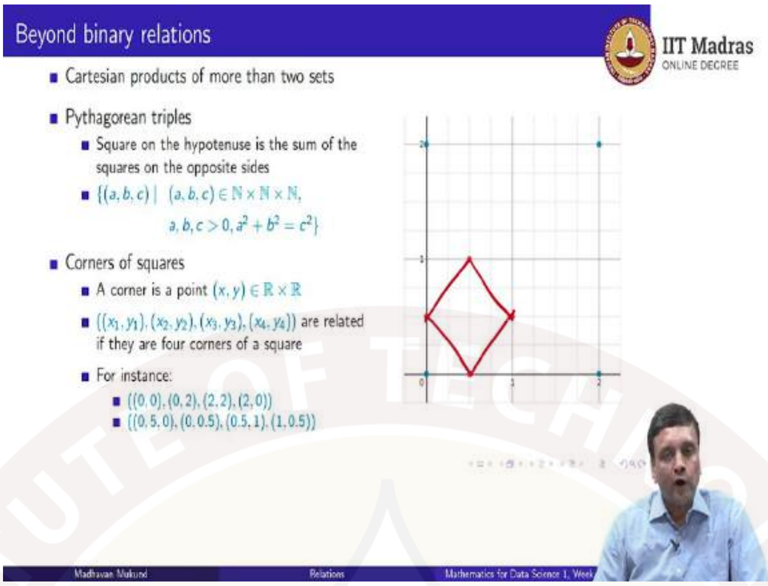
(Refer Slide Time: 08:51)



So, we do not have to restrict our self to binary relations. The Cartesian product notation extends to multiple sets. Let us look at three sets for instance. Remember, Pythagoras theorem which says that the square on the hypotenuse is the sum of the squares on the opposite sides. So, what values of a, b and c could be the sides of a right triangle are determined by Pythagoras's theorem.

So, we would say that a, b and c is a valid triple in the Pythagoras sense if (a, b, c) belongs to $N \times N \times N$. So, here we now have three copies of N and a, b, c must all be nonzero. They must all be positive length we do not want to have triangles in which one line one side is collapsed to a point and we want the constraint that $a^2+b^2=c^2$.

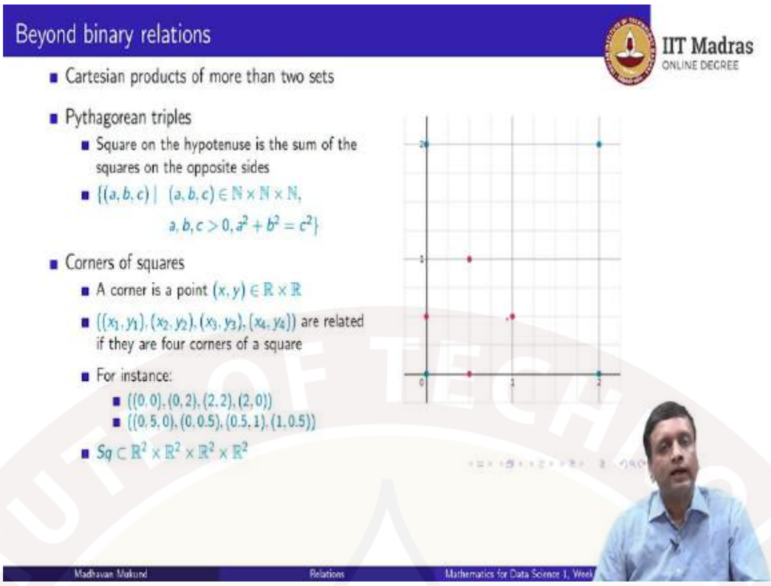
(Refer Slide Time: 09:41)



Here is another example. Suppose, we look at squares on the plane squares with real corners right. So, a corner is a point (x , y) which is in $R \times R$. So, we define the x coordinate and the y coordinate that defines the corner of a square and we want four corners which together form a square if we connect them by lines. So, for instance, if you look on the right the four blue dots correspond to a square which is cornered at (0, 0); (0, 2); (2, 0) and (2, 2).

The red square is also red points also define a square because this is a rotated square, but then if you rotate it vertically; you will turn out that this diamond is actually a square. So, there are many such four sets of points which form the corners of squares and we might be interested in all such four sets of points. So, now, we have a relation which involves four sets of points, but each point itself is a pair of real numbers; it is an x and a y.

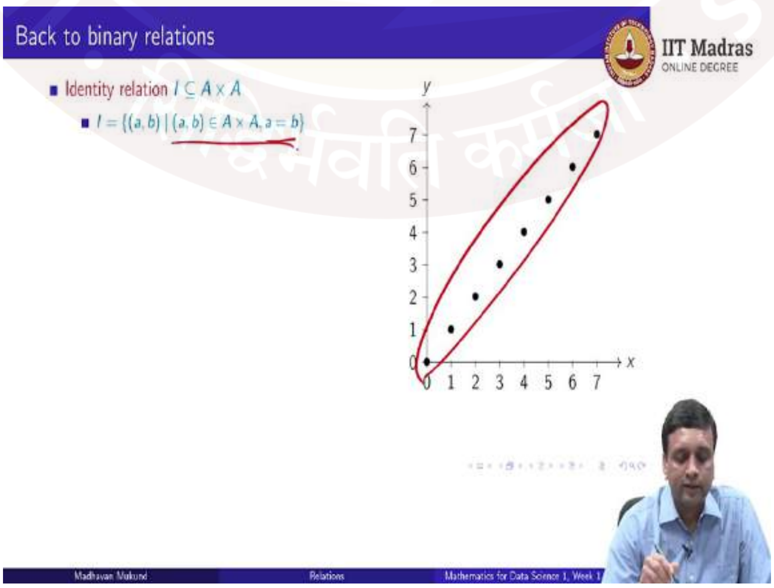
(Refer Slide Time: 10:35)



So, square if we think of it as a relation is actually a relation on R^2 that is the first corner times R^2 the second corner times R^2 the third corner and the fourth corner R^2 again. So, this is actually either a relation on eight copies of R or if you want to group it four copies of pairs of R.

So, this just says that we can take relations on arbitrary an arbitrary number of copies of a set and we get larger and larger from pairs, we move to triples we move to quadruples and in general if we have n copies we call this an n tuples.

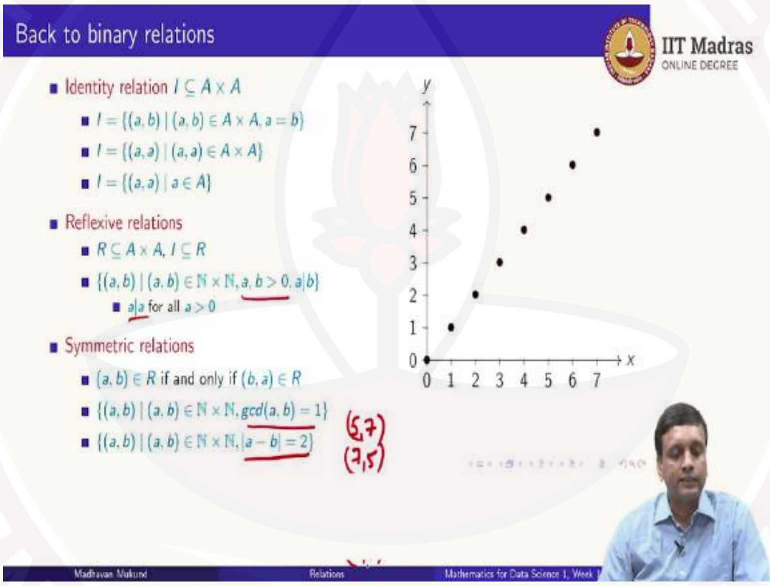
(Refer Slide Time: 11:11)



So, there are some special binary relations which pop up all over the place. So, it is useful to know their names. The first one is called the identity relation and as you would expect, the identity relation maps every element to itself. So, if I take A × A, so, first of all the identity relation is defined on two copies of the same set because identity means equality. So, I take A × A. So, this has all kinds of pairs (a , b), where both a and b belong to A and now, I want the condition that a = b.

So, in other words, I want things of the form (a , a). So, if I plot this for instance on the natural numbers and $N \times N$, then I get (0, 0); (1, 1) and so on, and these are the points which are drawn on the right in this grid.

(Refer Slide Time: 12:05)



Now, point of notation we sometimes it is tedious to write this notation as it is says us (a,b) time in n. So, we do not want to you know have to write out this long thing. So, sometimes we simplify this by saying I want all pairs such that a comma a belongs to A × A. So, what we are really saying is that the second day and the first day must be the same. So, we are collapsing the equality and this. Now, this is not technically correct, but this is often used in order to simplify the notation.

And, sometimes we might drop the product altogether. We might just say we want all pairs (a ,a)where a comes from the set A. So, in other words we are pulling out one copy of the element from the set and then we are constructing a pair by taking two copies of it. So, all of these are equivalent ways of writing this although only the first one technically follows the notation that we are using to introduce relations.

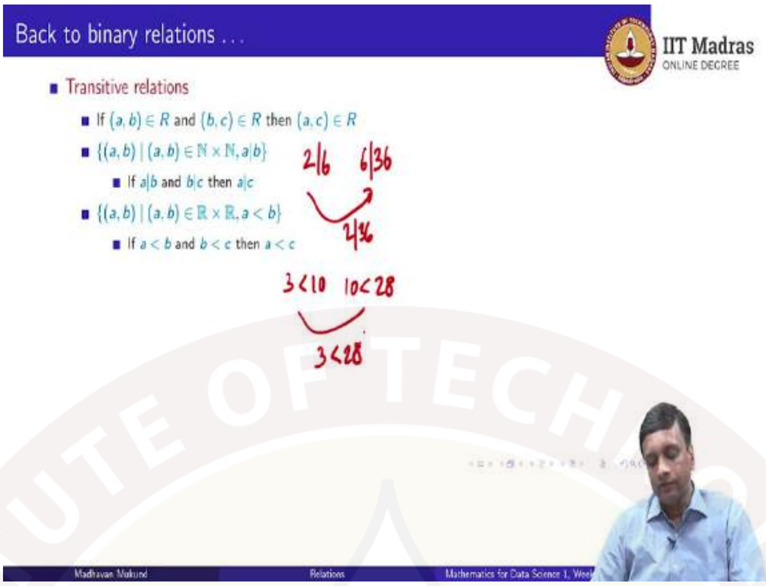
Now, there are some properties that relations may have. The first one is called reflexivity. So, reflexivity refers to the fact that an element is related to itself. So, a reflexive relation is one in which for every element a; (a,a) belongs to R. So, in other words based on what we just wrote above, it means that the identity relation is included in R. So, it does not mean that is the only thing. The identity relation has only the reflexive elements. A relation that is reflexive will have the identity pairs and it will have other pairs, but it must have all the identity pairs to be called reflexive.

A symmetric relation for instance is one where if (a, b) is there, then (b, a) must be there. So, for instance looking at reflexive relations, one example is the division relation. So, if we provided we make sure that the numbers are not 0, then we know it is reflexive because every number divides itself. So, if we take the reflect division relation as the relation that we introduced in the first part of this lecture that would be reflexive because a divides a for every a which is not 0.

Similarly, symmetric relations if we look at pairs where the greatest common divisor is 1, in other words they have no common divisors. This is what happens for example, in reduced fractions, then it does not matter whether we write it as (a , b) or (b, a). So, if (a, b) has greatest common divisor 1, so does (b , a). So, (a, b) and (b, a) must both either be there in the relation or neither will be there.

Similarly, if we look at this which is asking about the absolute value so, it is saying give me all numbers a and b such that a - b is either 2 or -2. So, the absolute value takes the difference and removes the negative sign. Now, we see that for instance if (5, 7) is there, then (7, 5) must be there because they both have the same difference depending on how we write it. Normally, in subtraction we have a sign difference, but because we are taking the absolute value there is no difference actually between these two. So, this absolute value relation also if we fix is a symmetric relation.

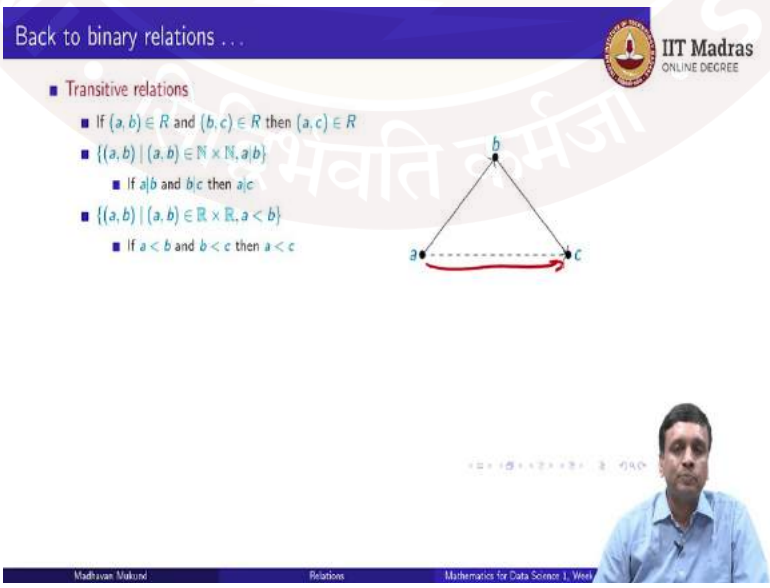
(Refer Slide Time: 14:51)



A third property that relations may have and which are useful is called transitivity. So, transitivity says that if we have two pairs which are related such that they share an elements. So, a is related to b and b is related to c, then a must be related to c. So, again our divisibility is a relation. So, supposing we say that 2 | 6 and we say that 6 | 36, then from this we can conclude that 2 | 36 as well, right.

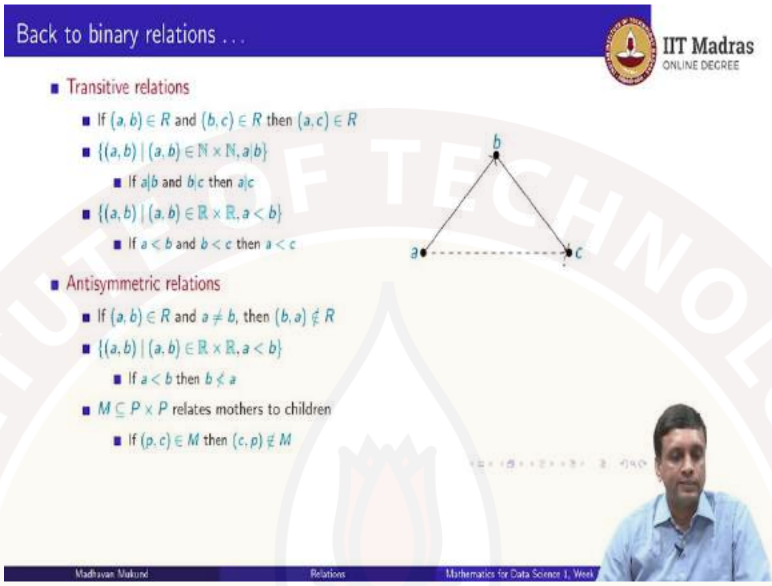
Similarly, if we take less than if we say that 3 < 10 and 10 < 28, then we know from this that 3 < 28. So, this is transitivity.

(Refer Slide Time: 15:34)



So, if we want to draw it pictorially if we have three elements a, b and c and this arrow remember we had this graph notation which says a is related to b and b is related to a, then this dashed line represents the requirement for transitivity a must be related to c.

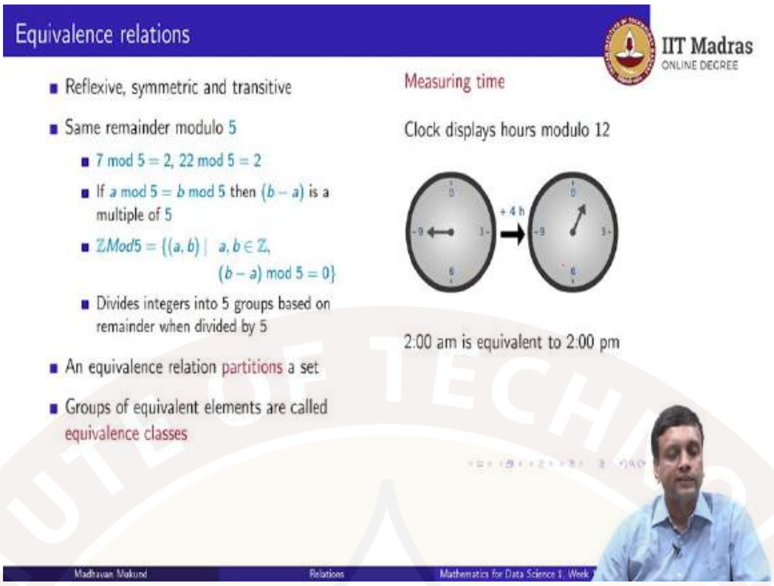
(Refer Slide Time: 15:49)



Now, we saw symmetry. So, symmetry says that if (a, b) is in R, then (b, a) must also be in R. Anti-symmetry says something different it says if (a, b) is in R, then (b, a) should not be in R. So, less than for example, which was transitive above is also anti-symmetric. If you take strictly less than, if a is strictly less than b; then it cannot be that be strictly less than a. So, this is an anti symmetric relation, but anti symmetry does not require that one of the two must be there. It only says that if one pair is there the opposite pair should not be there ok.

Similarly, if we look at our mother and children example; obviously, if p is the mother of c then c cannot be the mother of p ok. Now, there may be p and c such that neither p is the mother of c nor is c is the mother of p. So, that is allowed. We do not insist that every pair (p, c) must be related one way or another, but if it is related one way it should not be related the other way is what anti-symmetry says.

(Refer Slide Time: 16:46)



So, if we combine some of these conditions, we get an interesting class relations called equivalence relations. So, equivalence relation is something that is reflexive, symmetric and transitive. So, as an example supposing we connect together all numbers which have the same remainder modulo 5. So, for instance 7 has a remainder 2 with respect to 5 and so does 22. So, 7 and 5 would be related in this way if we define the relationship as having the same remainder modulo 5.

Now, notice that if two numbers have the same remainder modulo 5; that means, that going from one number to the other you are going in multiples of 5. So, for instance 22 -7 is 15 right. So, this is this modulo arithmetic. So, if you add the number that you are dividing by, then you get the same remainder and so, in set notation we can say that the integers modulo 5 are all pairs a, b such that b - a mod 5 is 0. In other words, we are not asking what is the actual remainder of b and a, we are just saying that b and a are separated by a multiple of 5 therefore, they must have the same remainder modulo 5.

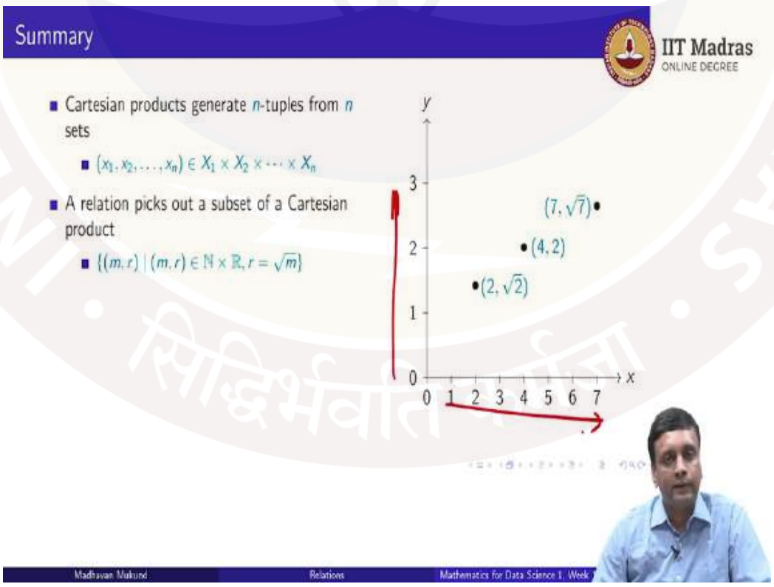
Now, this divides the integers into five groups if I based on the remainder. So, there are the group of numbers which are divisible by 5, they have remainder 0. Those like 6, 11 and all which have remainder 1; 7, 12 and all which one remainder 2 and so on. So, we have five possible remainders 0, 1, 2, 3, 4 and therefore, this divides the set of integers into five disjoint classes.

As an example of modulo arithmetic that we are all familiar with, consider what happens when we look at a normal clock. Now, a normal clock measures time from 0 to 12 and then cycles around again. So, though there are 24 hours in a day, the clock is actually partitioning these 24 into two sets where we have 0 and 12 as same, 1 and 13 as same and so on right. So, 2 am and 2 pm, there is no distinction on the clock.

So, the clock is actually showing us this equivalence class of hours regarding am and pm as being equal and we have to know from context whether the clock is showing am or pm. So, the main thing to note about an equivalence relation is that it partitions a set. It partitions a set into disjoint groups, all of the elements within a group are equivalent and all of the elements outside across groups are not equivalent to each other.

So, the groups of equivalent elements that we formed through an equivalence relation are called equivalence classes. So, this might look a little abstract now, but equivalence classes really represent a kind of equality and sometimes we are happy to work with this equality in terms of equivalence relations rather than actual equality and it has very much the same properties as equality does.

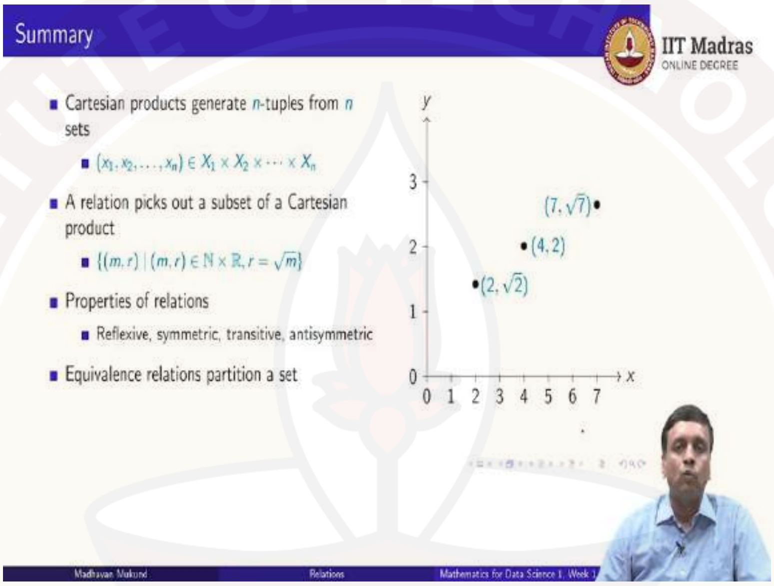
(Refer Slide Time: 19:28)



So, to summarize as we have seen a Cartesian product can generate n-tuples of elements from n sets. So, if we have X_1, X_2 , up to X_n, n sets these can be different or the same, then we can take one element from each set and form an n-tuple x_1, x_2, up to x_n. And, when we now pick out some particular subset of these n-tuples, we get a relation. So, for instance, if we take pairs from $N \times R$ and we want the second element of the pair the real number to be the square root of the first element, then we get $N \times R$ such that $r = \sqrt{m}$.

So, here on the right we have seen we show one picture of this. So, there are some elements like $(2, \sqrt{2})$, $(4, 2)$, $(7, \sqrt{7})$ and so on. Now, just notice that in this picture the y-axis is elongated compared to the x-axis. So, this is not in some sense to scale in both dimensions because the square root function behaves like this.

(Refer Slide Time: 20:27)



So, we have seen that there are some properties that we would like to record of binary relations – reflexivity, symmetry, transitivity and sometimes anti-symmetry. And, using reflexivity, symmetry and transitivity together we get what is called an equivalence relation, an equivalence relations partition sets into equivalence classes which behave like equality.

Thank you.

IIT Madras

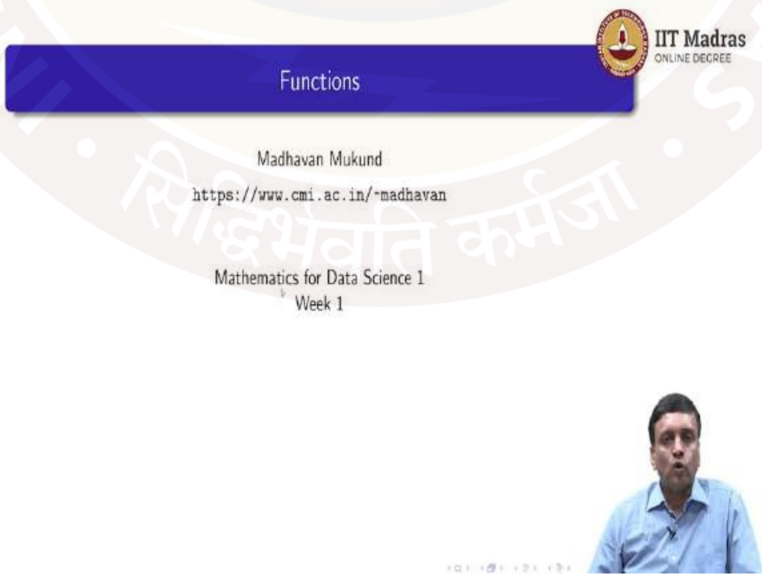
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

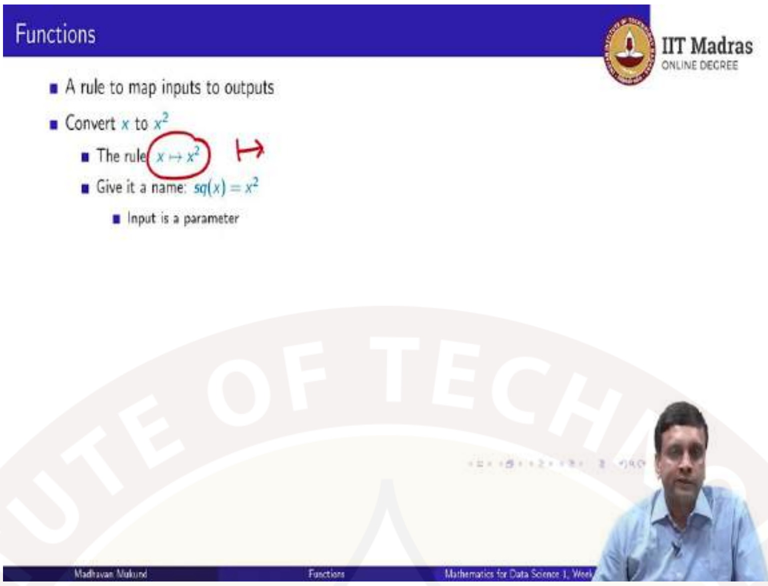
Week - 01
Lecture – 07
Functions

(Refer Slide Time: 00:05)

(Refer Slide Time: 00:14)



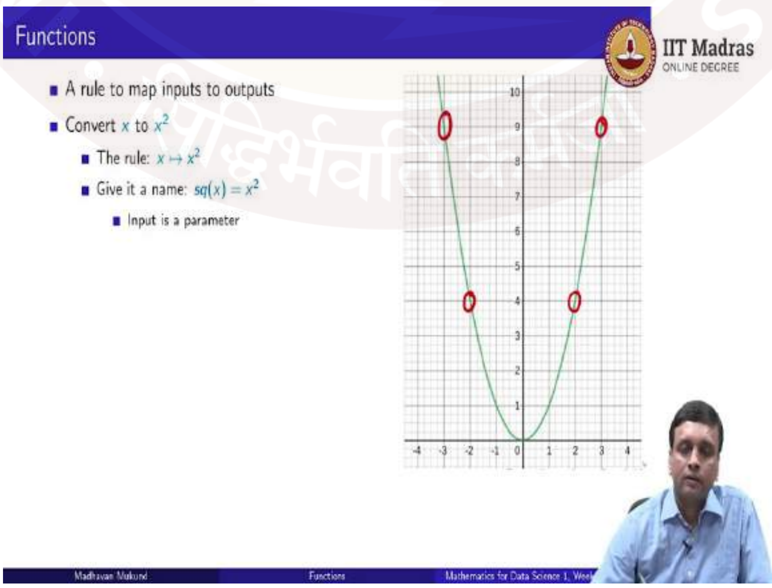
(Refer Slide Time: 00:17)



So, closely related to relations are functions. So, what is the function? A function is a rule that tells us how to convert an input into an output. So, for instance suppose we want a function that given an x returns as x^2, then this is one way to write the rule. We write this symbol which says x maps to x^2; given an x it is transformed to x^2, but more conventionally we also give a name to the function. So, in this case we can call it square(x).

So, square(x) takes a parameter x as input and it produces as output; some value which transforms this parameter, in this case x^2.

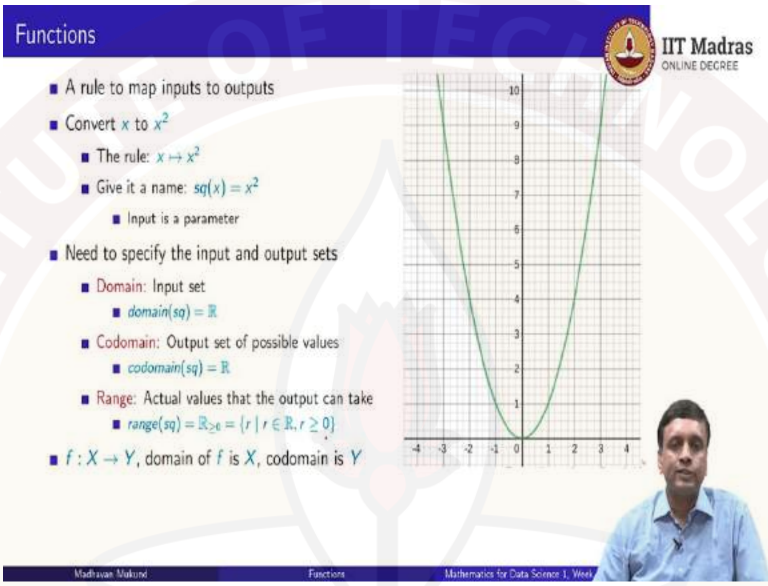
(Refer Slide Time: 00:56)



So, we can plot x versus x^2 by putting all the points where the second coordinate is the function value of the first coordinate. So, if we look at x^2 for instance, it forms this up you know inverted parabola shape which you should be familiar with. And notice that because for instance 2^2 is the same as $(-2)^2$, there is a symmetry about the y axis.

So, for instance 2^2 is the same as $(-2)^2$, and 3^2 would be the same as $(-3)^2$ and so on.

(Refer Slide Time: 01:31)



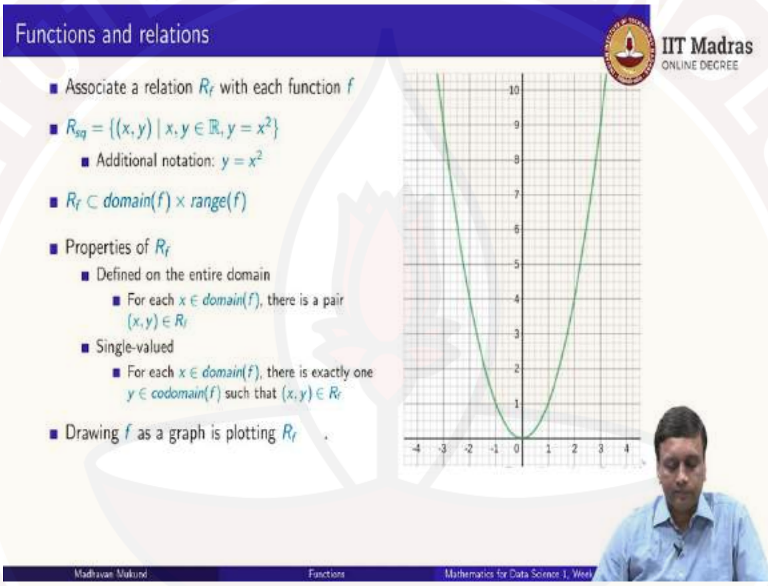
So, when we define a function, we have to be careful about specifying what set we take the input from and what sets the output produces. So, the input set is called the domain. So, for instance the domain of square as we have defined it above is a set of reals, so we can take the square of any real number.

Now the output when we apply square, we know that it is going to be a real number; so the codomain as it is called is the output set of possible values is called the codomain, in this case is the reals. But of course, we know that when we square a number; even if the input is negative, the output is going to be positive. So, even though the codomain is a set of all reals, we cannot get all reals as output of the square function. So, there is a separate name for that called the range.

So, the range of a function is a subset of the codomain; the range tells us what values the function can actually take. So, in this case the range of the square function is the non-negative reals. So, this is all real numbers greater than equal to 0 which is sometimes written like this and if you want to explicitly write it out; it is the set of all r in the set of reals such that r >= 0.

So, in order to specify a function abstractly and describe its domain and codomain, we usually write that f which is the name that we give to an arbitrary function is a function from X the domain to Y the codomain. So, this notation f : X to Y tells us without telling us what the function is actually doing; it tells us on what sets it operates, what is the input set and what is the output set.

(Refer Slide Time: 03:08)



So, the close connection between functions and relations is that we can associate with every function f a relation R_f; and R_f is merely all the pairs of inputs and outputs that the function allows. So, for example, with our square functions sq we have R_{sq} as all pairs (x,y), such that y is equal to x^2.

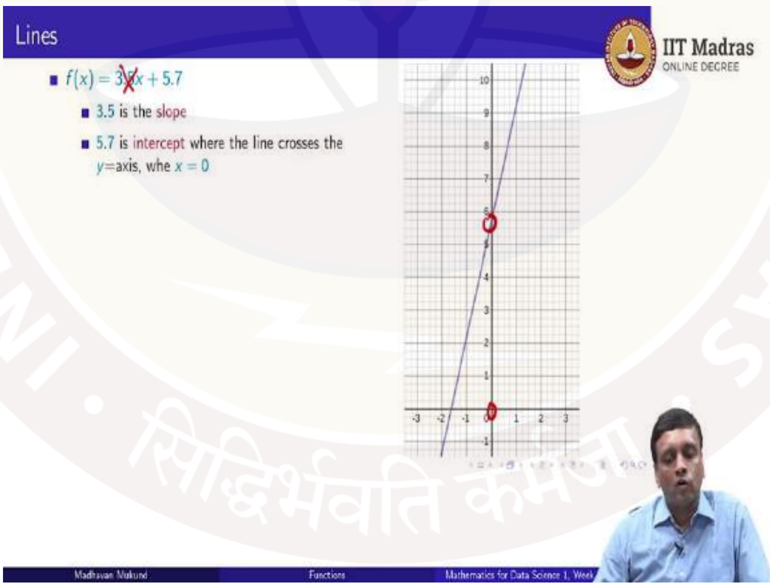
So, this is actually sometimes simplified by saying y is equal to x^2. So, we do not write out f(x) and then say f(x) is y; we just directly say y is equal to x^2 to denote that the output is the square of the input. So, this is an implicit notation, where we are implicitly naming the output for each x as y. So, notice that if we talk about it as a relation; remember that a relation is a subset of the Cartesian product of two sets. So, in this case, the Cartesian product is formed by the domain of the function and the range of the function, and then the relation is a subset of the domain X the range.

So, what are some properties of this relation? Well, first of all when we define the domain of a function, we really mean that the function is defined at every possible value in that domain. So, for every x and domain of the function f, there must be a valid value f(x); so there must be a y such that (x, y) belongs to the relation R_f. The other property is that this is a rule for producing an output from an input; so there can be no confusion about what the output is.

So, for each x that we feed in as a domain value to the function, there must be as exactly one output value f(x) that we get out. So, there is only one y in the codomain, such that (x,y) belongs to R_f. And in fact, we saw in the lecture on relations that, we would draw relations by plotting the points which form part of the relation. So, technically when we are drawing a graph of a function as we have done here for this parabola, we are actually drawing all the points which satisfy the relation R_f.

So, plotting a graph is the same for functions and relations; because implicitly we are plotting the relation that corresponds to a given function.

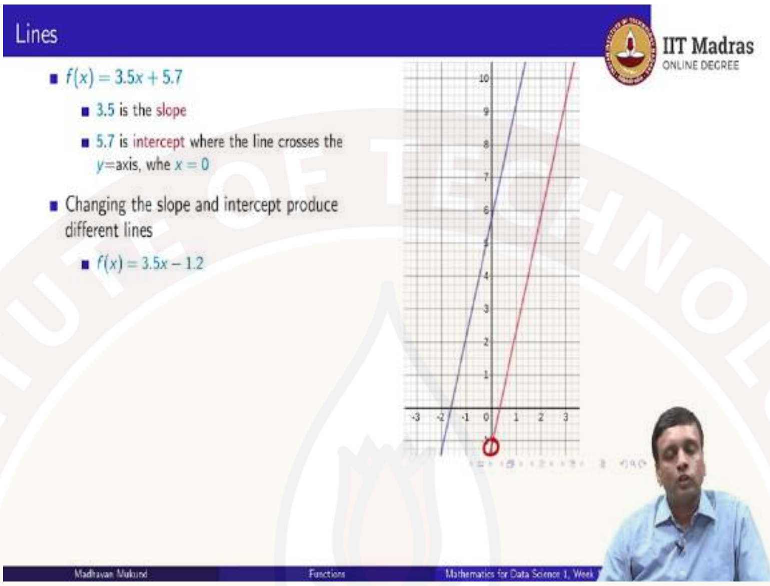
(Refer Slide Time: 05:23)



So, let us look at some other functions that we will encounter as we go along. So, if we have a function of the form something x + something. So, mx + c, then this defines a line. So, then the like we see a line 3.5x + 5.7. And what we will see as we go along in this course is that, the quantity which multiplies x is called the slope and it determines the angle at which the line goes; and the other quantity which is without x determines the intercept.

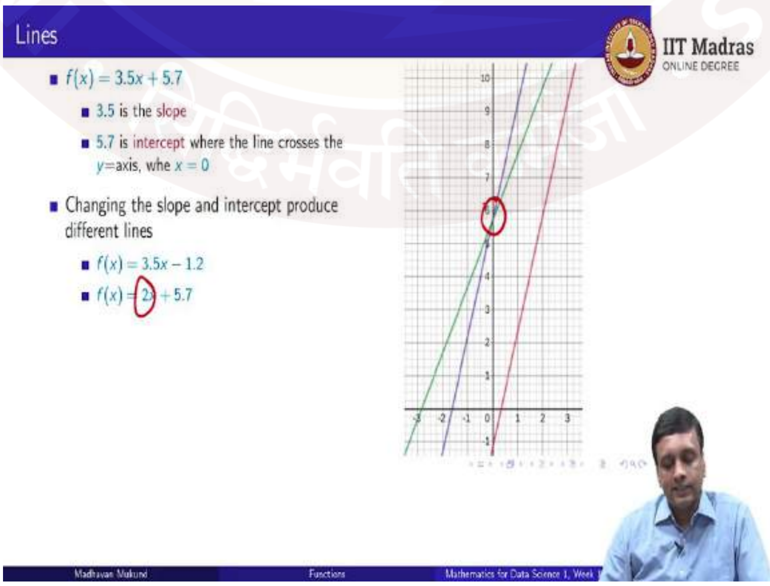
So, notice that if you set x = 0, then the first term goes to 0; this gets cancelled out, if x is 0. So, the answer will be 5.7. So, when x is 0, you get 5.7. So, what the second term tells us is where this line crosses the y axis.

(Refer Slide Time: 06:09)



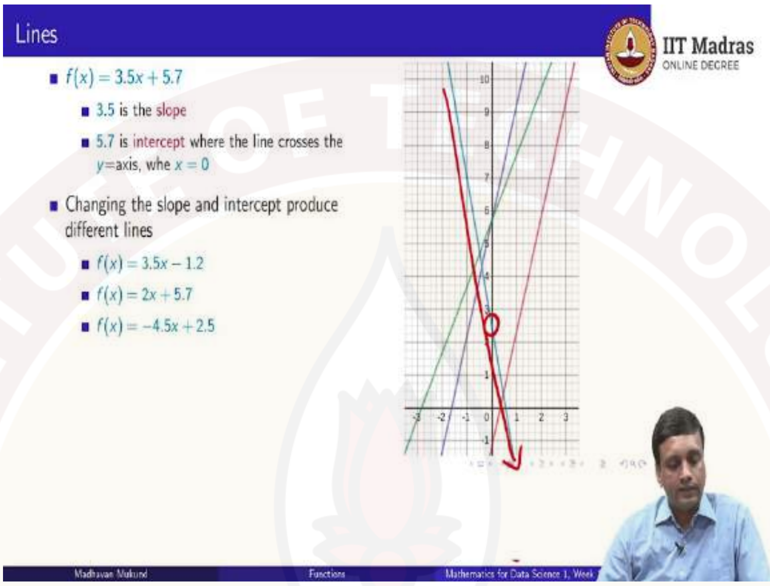
So, if we change these two values, we get different lines. So, for instance if we change the intercept and keep the slope the same; then we get a line which has the same slope it is parallel, it is at the same angle. But now the intercept is -1.2; so it crosses the y axis lower, so the whole line is shifted to the right.

(Refer Slide Time: 06:28)



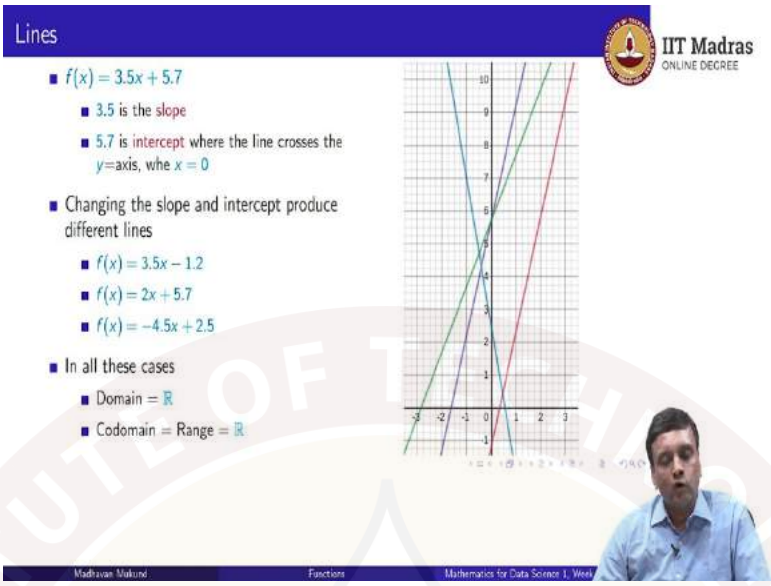
On the other hand if we keep the intercept the same; but we change the slope, we get a different slanted line. So, here we have reduced the slope from 3.5 to 2; so it is a shallower line and the green line passes through exactly the same point 5.7 as the previous one, but it has a shallower slope.

(Refer Slide Time: 06:49)



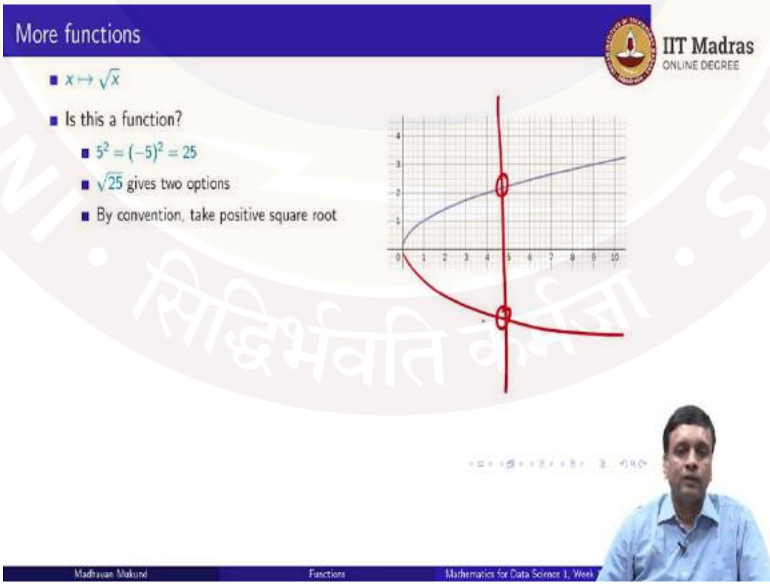
And we can change both and in fact, we can put a negative slope; so if you have a negative slope, it comes down rather than going up, so we have this line coming here. And notice that it crosses at 2.5, so that is the intercept. So, by changing the values of the slope and the intercept, we get many different lines and many different functions.

(Refer Slide Time: 07:07)



And for all of these functions that we have defined the domain is the set of reals, the codomain is the set of reals; but also because we can intuitively see that the line goes from way down -∞ to way up +∞ whether it is going up or down, it can take all values in the real. So, not only is the codomain equal to *R*, it is also the range.

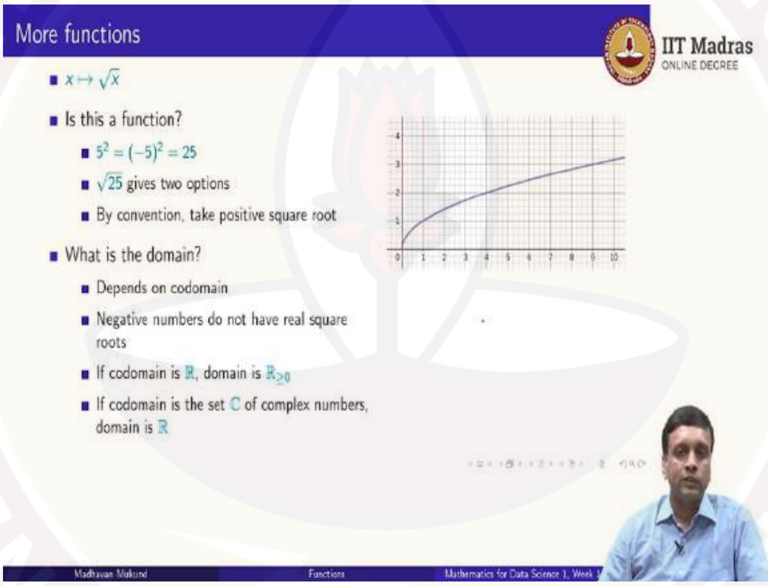
(Refer Slide Time: 07:28)



So, here is another function x maps to √x. The first question is, is this a function? So, remember that for a function, we need it to be defined on every input value and we also needed to have a unique output. So, remember that when we square a negative number, we get the same as when we square the positive version; so 5^2 and $(-5)^2$ are both 25.

So, technically if we take $\sqrt{25}$, we cannot determine whether we are talking about +5 or -5. So, when we write $\sqrt{x}$ as a function, our convention is that we are taking the positive square root. So, the function on the right plots the positive square root; if we were to take the negative square root, then it would be a symmetric curve going below. And now if we take both these together, then this is not a function; because if we take any x value, we have two possible outputs for this which is not allowed. So, we are taking by convention the positive square root.

(Refer Slide Time: 08:23)

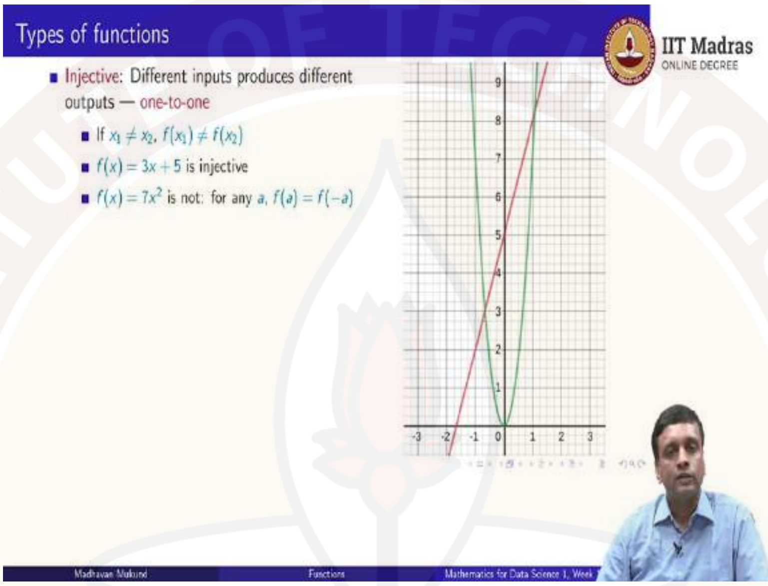


Now what is the domain of this function? Well it depends on what we allow the codomain to be. We have seen that negative numbers cannot have real square roots; no real number can multiply itself to produce a negative number, because of the law of signs for multiplication. So, if we insist that the output should be a real number, then the domain of this function, the function can only be defined when the input is not negative. So, we have this set which we defined before; the set of reals bigger than or equal to 0.

On the other hand, if we move to the set of complex numbers which we said we are not going to describe in detail; the set of complex numbers includes $\sqrt{-1}$ and implicitly through that the square root of all negative numbers. So, once we allow complex numbers as the output of our function, then we can define square root on all the real numbers.

So, the notion of domain and range is kind of flexible depending on how we are going to use the function. So, we have to be very clear when we are using a function what context we are using it in.

(Refer Slide Time: 09:24)

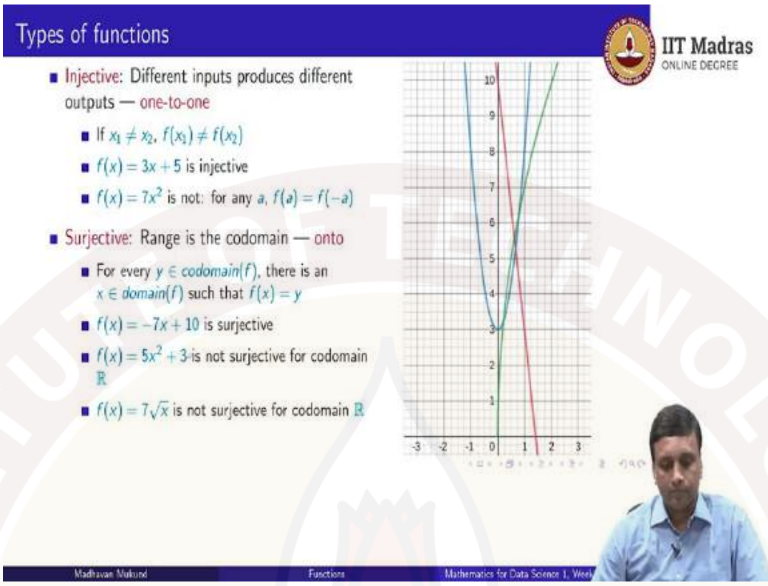


Now we saw when we looked at relations that there are some properties of relations which are interesting like reflexivity, symmetry and so on. Similarly there are properties of functions which are interesting; the first interesting property of function is whether it is one to one, whether it is injective.

What this means is; if I give you different inputs, does the function always produce different outputs? If $x_1 \neq x_2$, is it guaranteed that $f(x_1) \neq f(x_2)$? So, if we look at the linear function that we saw before the line, then we can see that it is injective; because if we change x, we move along the line to a new point. So, no two x points, point to the same y point; so therefore, this is an injective function.

If on the other hand, we take a parabola as function which of the other form something squared, so $7x^2$for instance. Then we already saw that f(a) is the same as f (-a), so there will be two points; the plus version and the minus version, both of which has the same output. So, it is not the case that distinct outputs produce distinct, inputs produce distinct outputs; so the square function is not injective.

(Refer Slide Time: 10:32)



On the other side we talked about the distinction between the codomain and the range; we said that the codomain is the set of values into which the function produces answers, but the range is the actual set of values of the functions can take.

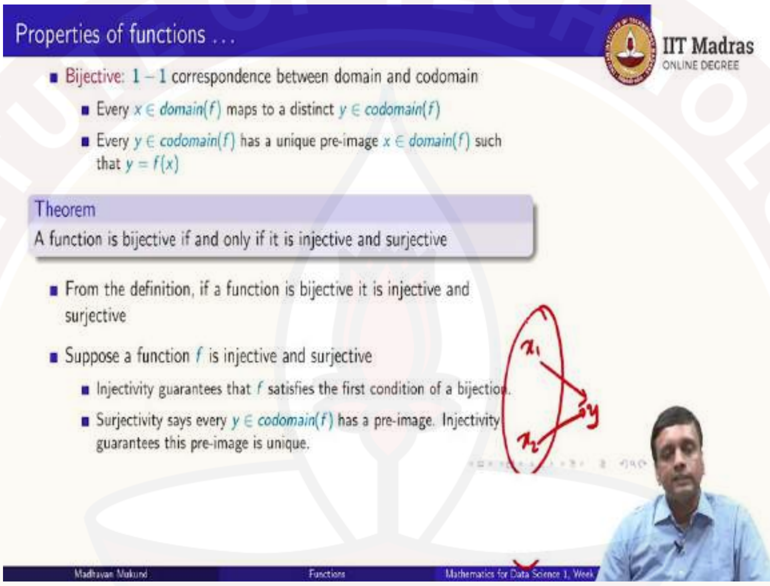
So, the question is, whether or not all values in the codomain are actually touched by the function and this is called surjectivity or onto. So, the range of a surjective function is in fact equal to the codomain, which says that for every y which is in the possible codomain of f; there is actually an x in the domain of f, such that f (x) = y.

Now, once again if we take a line, then this is surjective; because if I pick any point y, I can find a point x, I can solve for x for example, which gives me that y. On the other hand if I take a parabola, in this case we have shifted the parabola up, so it is $5x^2+3$. Then we can see that, first of all a parabola with no shifted, if I did not have this +3 term; then we know that it can only take positive values, because x^2will always be a non-negative number.

Now if I further add +3, it can only take values 3 and above; so this definitely is not surjective, the domain codomain is a set of all reals, but the actual range is only if the reals which are bigger than or equal to 3. Similarly if I take this $7\sqrt{\square}$function, then we know that even if we take the codomain to be R; so we only take square roots of positive numbers. We know that we will never get a negative answer, because by convention we have taken positive square roots.

So, this is again not a surjective function. So, these are two important properties of functions, are they injective is it one is to one; if I give you different inputs, do I get different outputs and is it surjective, is it onto, does every possible output have a corresponding input that maps to it.

(Refer Slide Time: 12:25)



So, if you combine these two, you get something called a bijective function. So, a bijective function is something with where there is a one to one correspondence between the domain and the codomain.

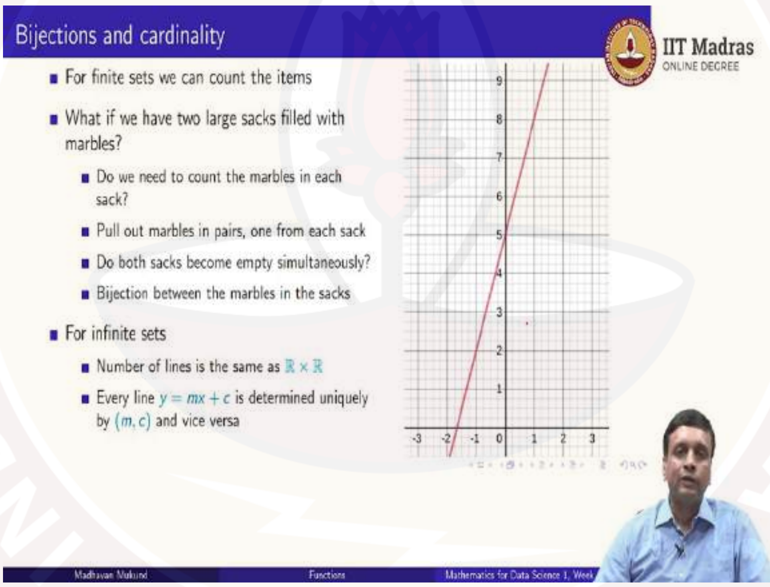
So, every x in the domain maps to a distinct y in the codomain and every y in the codomain has a unique x that maps to it. So, from the statement it looks clear that this corresponds to injectivity and surjectivity. So, actually this is the theorem that a function is bijectuve if and only if it is both injective and surjective.

Now this may look obvious, but actually only one direction is obvious, from the definition, we can see that if a function is bijective; it must be injective, because it says every x maps to a distinct y, so no two x will map to the same y.

It also says it is surjective, because it says every y in the codomain has a unique pre image. So, the fact that a bijection implies injectivity and surjectivity is part of the definition; the other way requires a small argument. So, supposing a function is injective and surjective, we have to show that it is bijective. So, for this, we have to guarantee first that every x maps to unique y; but this is guaranteed because the function is injective, injectivity says if I have two inputs x_1 and x_2 which are not the same, $f(x_1) \neq f(x_2)$. So, this is fine.

What about surjectivity? So, surjectivity says that everything in the output comes from some input not necessarily unique; but if two things map to the same output right, if two things map to the same output, if I have a y such that I have x_1 and x_2 mapping to the same y. So, if it has even, if a surjective function if the output has two pre images; then these two pre images do not satisfy injectivity. So, if I combine surjectivity in the presence of injectivity, I know that the pre image is unique; and therefore these two conditions guarantee that I have a bijection.

(Refer Slide Time: 14:15)



So, an important use of bijection is to count the items in a set. So, remember we said that the cardinality of a set is the number of items and if you have a finite set, we can count them. Now supposing somebody gives you two large sacks filled with marbles or balls and ask you to check whether the two sacks have the same number of balls each. So, think of these sacks as sets and these balls are a large number of elements.

Now, you could of course, count the marbles in each sack, but this is a bit tedious; because we know that as we are keeping track of these small objects, we often lose count or miss count or add one or plus one. So, at the end, we have to be doubly sure that we have counted correctly, so we will count it a number of times. So, counting the marbles in each sack and then checking if the two counts are equal is a tedious process and it is error prone, if we do it manually.

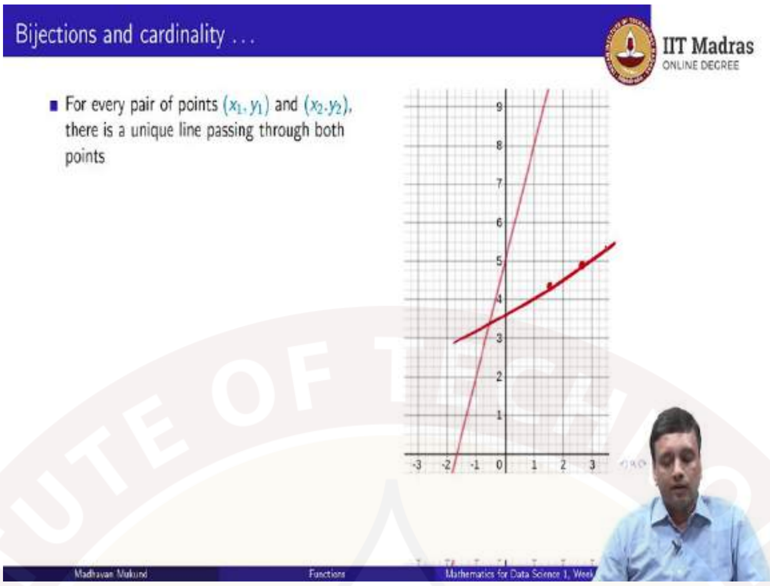
Now, here is a manual process which is less error prone. Supposing we put our hand into each sack and pull out a marble from each sack and put it away somewhere; then we put our hands again in and take out one marble each again and put it away somewhere. So, with each move, we are taking out one marble from each sack. So, what can we say; well if the two marbles sacks get empty together, then we pulled out one from each. So, we have actually established that there is a one to one correspondence between the marbles in the first sack and the marble in the second sack.

If on the other hand when we find one sack is empty and the other sack is not empty; this means that up to this point, we pulled out an equal number of marbles from both sacks and now one sack has extra marble, so they were not equal. So, in this way establishing a bijection is equivalent to saying that two sets have the same cardinality. So, for finite sets this is a convenience; but for infinite sets this is the only way in order to establish that the cardinality is the same.

So, for instance supposing we want to know whether the number of lines that we can draw is the same as the number of points on this plane R X R . So, R X R is a set of all points that you can draw on this plane and the number of lines we can draw is a number of such straight lines that we can draw; are these the same? Now it may not seem obvious how to argue this one way or another; but remember that we said that every line can be represented by a function of the form mx + c. And we also said that if you change m, you get a new line and if you change c, you get a new line. So, m and c together uniquely define a line.

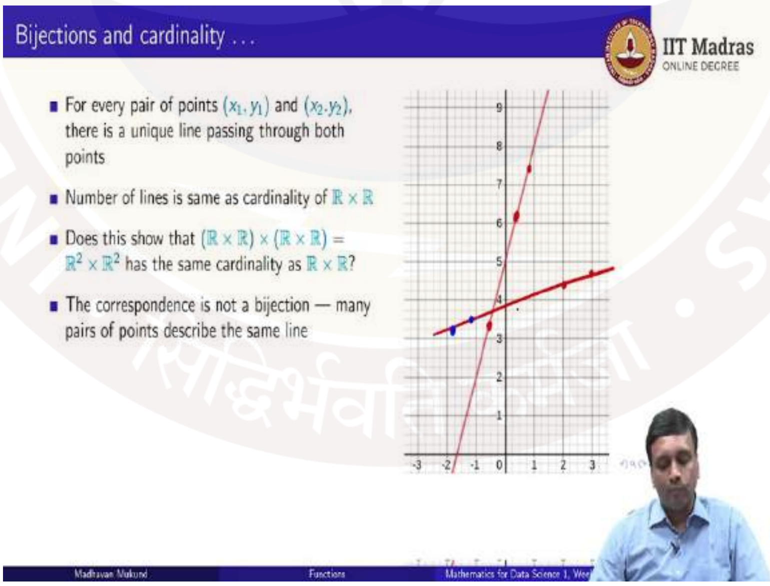
So, since m and c together uniquely define line; every pair (m, c) defines a line and every line defines a pair (m, c), so there is a one to one bijection between the lines and the pairs of points on this plane. So, actually the number of lines is the same as R X R. So, think about it, because this may not be obvious at first sight; but by establishing a bijection in this way, we can say that the number of lines that we can draw on a plane are equal to the number of points on a plane.

(Refer Slide Time: 17:09)



Now, suppose we extend this argument; if we take any two points right, if we take two points say $x_1$andx_2, we can draw a unique line passing through these points. So, this is a well known fact from geometry.

(Refer Slide Time: 17:23)

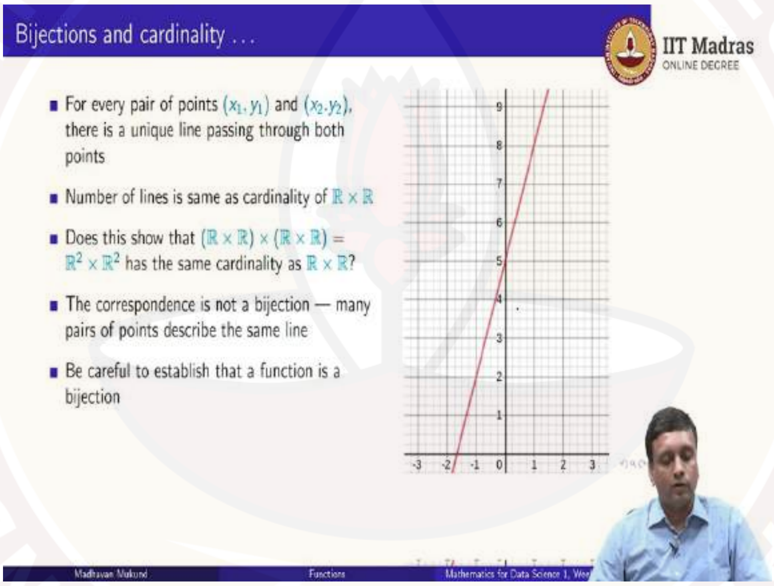


So, we know that the number of lines has the same cardinality as R X R that is what we claimed in the previous argument. Now we say that every pair of points defines a line. So, can we say that every pair of points therefore, has the same cardinality? So, remember this is a pair of points.

So, we have one point here and one point here. So, do we say that every pair of points has the same cardinality as the set of all points? So, it is R^2X R^2 the same as R XR, is this an argument for that? So, important thing is to ensure that we have a bijection; the problem is that this is not a bijection, because along any line we have many points, right.

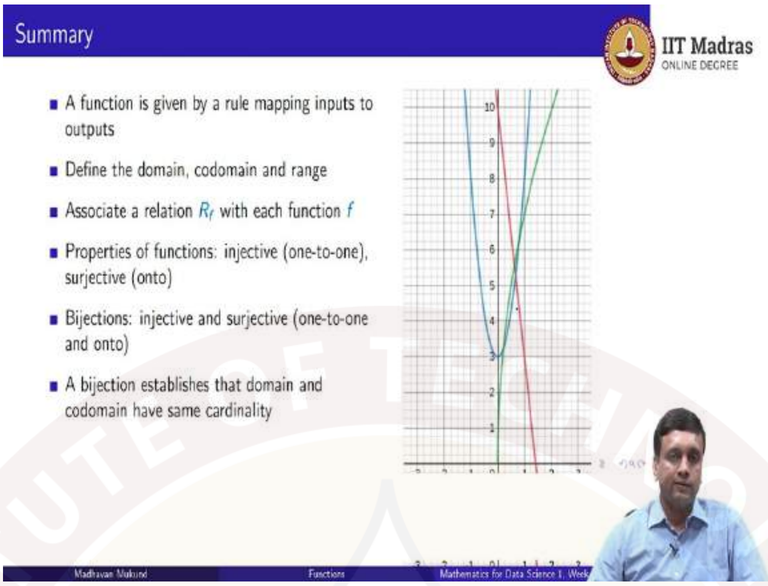
So, if I take these two points, indeed it forms a unique line; but I get the same line if I take these two points for instance. So, it is not the case that every pair of points that I pick generates a different line. So, unless I can show you that pairs of points, different pairs of points generate different lines; I do not get a one is to one correspondence between pairs of points and lines, and therefore this bijection breaks down.

(Refer Slide Time: 18:24)



So, whenever we are trying to use a bijection to describe some kind of a correspondence and count points especially in an infinite set, count elements of an infinite set, compare infinite sets against each other; you must make sure that the function you are defining is really a bijection.

(Refer Slide Time: 18:39)



So, to summarize a function gives us a rule to map inputs to outputs. And with each function we have to specify three sets; we have to specify the domain, so the function must be defined on every set in the element of the domain set, the codomain what are the output elements supposed to look like and the range which was actually the output assumed by the function once we applied.

So, not all elements in a codomain may actually be attainable by the function; the range is those elements which you can reach through the function. With each function we can associate a binary relation consisting of all pairs (x , y), such that y = f (x). Then we saw some interesting properties that we would like to prove for functions in order to make use of them; one is injectivity that is every pair of distinct inputs produces distinct outputs, so this is one is to one. And surjectivity which says actually that the codomain and the range match; everything that I could possibly generate, can in fact be generated by applying the function.

Then we saw that a bijection combines these two. So, a bijection gives us something which is an injection and a surjection; something that is one to one and onto. And once we have a bijection between two sets, we can actually argue that the two sets have the same cardinality and this is often the only way to prove that two infinite sets have the same cardinality.

Thank you.

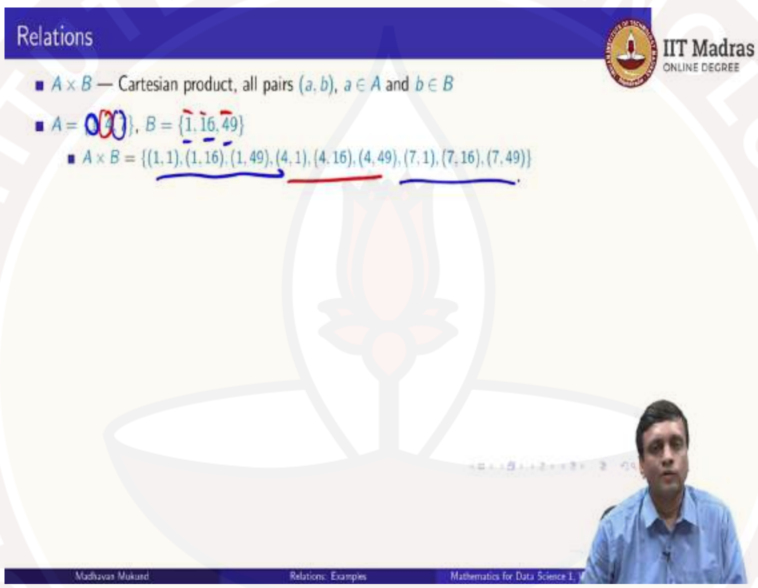
IIT Madras

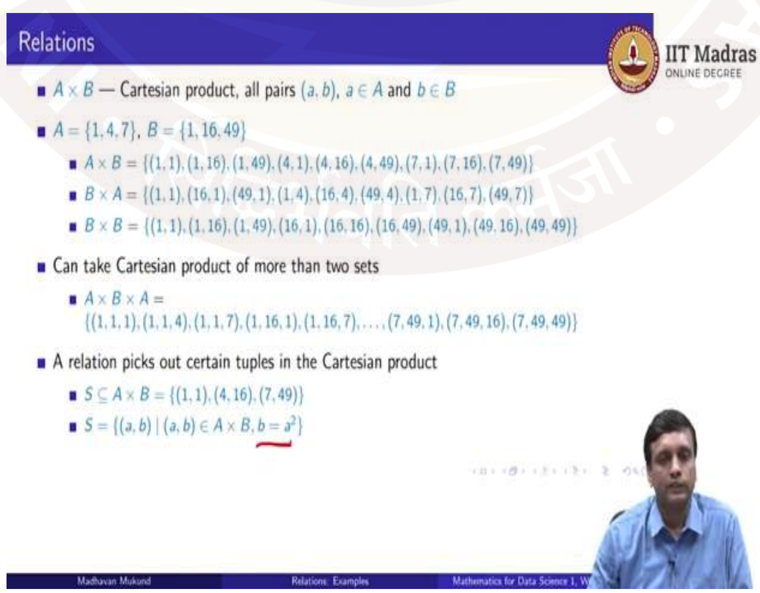
ONLINE DEGREE

Mathematics for Data Science 1.
Professor Madhavan Mukund.
Department of Computer Science
Mathematical Institute, Chennai.
Lecture-7A.
Relations: Examples.

So, earlier we defined relations as subsets of elements of a Cartesian product which have special properties. So, let us take a look at relations again and understand why we are so interested in relations.

(Refer Slide Time: 00:25)





So, remember that a Cartesian product takes all pairs of elements from a collection of sets. In particular, if you say A cross B, you are taking 2 sets A and B, and you are taking every pair of elements of the form small a small b such that the first small a comes from capital A and small b comes from capital B. The order is important, the first element in the pair comes from the first set, the second comes from the second set. So concretely, let us look at these 2 sets.

So, suppose A = {1, 4, 7}, so it has 3 elements, and B = {1, 16, 49}. So, if you now look at A × B, it looks at every pair. So, if you can take this one, and combine it with 1, 16 and 49 to get this. Then you can take this 4 and combine it again with 1, 16 and 49 to get these for 3 pairs, and finally you take 7 and then you combine it again with 1, 16 and 49 to get these pairs.

So, it is easy to see that if you have m elements in the first and n elements in the second, every one of those m elements is paired with every one of the n elements, so you get m × n pairs. Now, the first thing to remember is that the Cartesian product is ordered. So, there is a first and there is a second. So, if you reverse this and say B × A, you do not get the same set of pairs, every pair is reverse. So, (16,1) replaces (1,16), (49,1) replaces (1,49). So, this is the first thing to remember about Cartesian products.

The other thing to remember is that there is no relation, there is no constraint on what you can take the Cartesian product of. You can very easily take the Cartesian product of a set with itself. So, the set to itself is not just pairs of identical elements, but also pairs of non identical elements. So, if you take B × B, you get Of course, (1,1), (16,16), (49,49). But you also get the dissimilar pairs like (1,16), (16,49), (49,16), and so on.

So, this is an example with 2 sets, but there is nothing to restrict us to 2 sets. So, in general, a Cartesian product can take a large number of sets and gives us tuples. So, for instance, if we take 3 sets, we get these triples, each element has 3, each element in the Cartesian product has 3 elements in order.

So, here for instance, if I do A × B × A, I take every element in A, combine it with every element in B and then with A again. So, I have 1 from A, 1 from B and 1 from A. Then I have 1 from A, 1 from B and 4 from A, the second copy of A and the first copy of A are different.

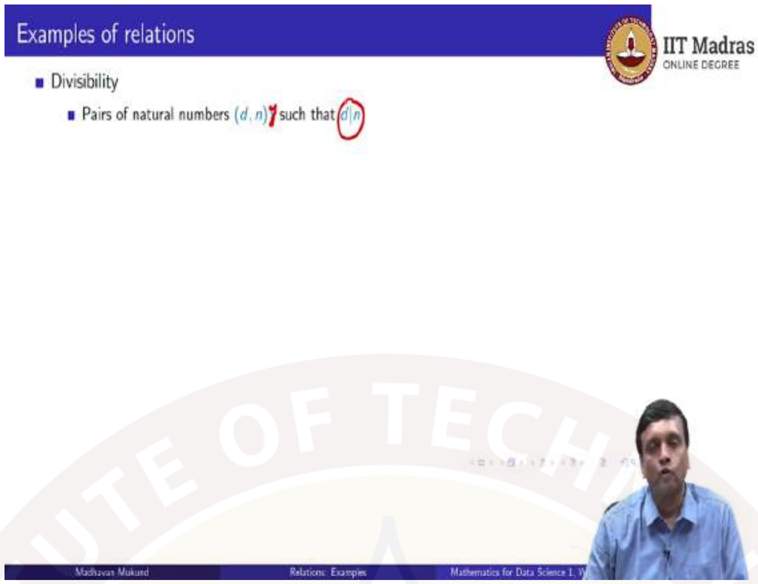
So, I have (1,1,1), (1,1,4), (1,1,7), then I move to the second element of B, I have (1,16,1), (1,16,7). Now, ultimately the Cartesian product is a set, so it does not matter in what order I write these triples. But to order to write them down systematically, it is convenient to write them down in this particular way, where we go through each set one by one, otherwise, we may miss out on something. So, the reason we need Cartesian products is because they are the building blocks of relations.

So finally, what we want is not all these pairs or triples, but some of them which are of interest to us. So, for example, from the first Cartesian product A × B, we may be interested in the pairs where each element from A is paired with a corresponding position B. So, the first element in A is paired only with the first position in B, second with the second and so on. So, we might want to say that we want S, a set which is a subset of A × B, which from those 9 different pairs picks out only 3 of them of interest, (1,1), (4,16) and (7,49).

Now, if as in this case, there is some way of describing this, which is more abstract, you can also use a set comprehension. So, we can talk in terms of positions or observe that in this particular case, the second element is always a square of the first element. So, we could also write this as the set of pairs (a,b), where (a,b) comes from A × B, so we are generating every possible pair in the Cartesian product.

But then we are filtering, remember that we had these filters, so we are filtering it so that we only retain those pairs for which the second component B is the square of the first component. So, this is how relations are defined. They are typically defined as subsets of the Cartesian product. And we can either write out the subset explicitly or try to express it implicitly using the set comprehension notation.

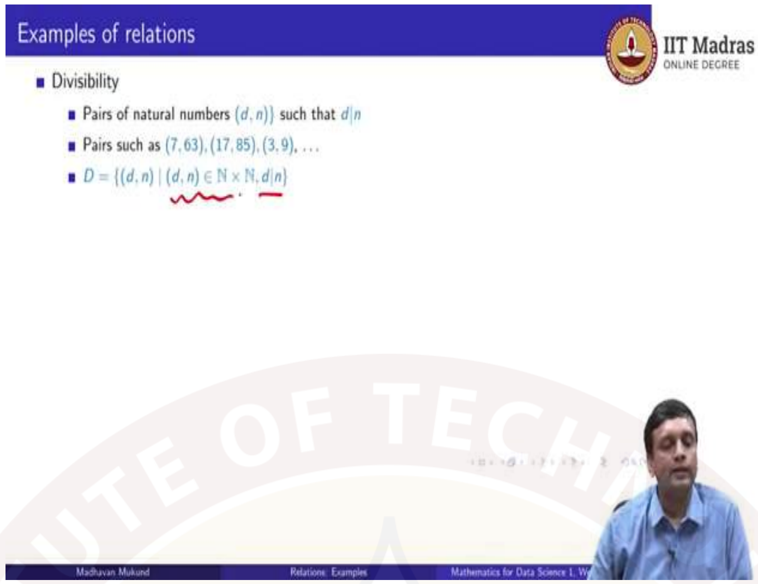
(Refer Slide Time: 4:57)



So, we saw some examples. So, let us look at these examples again more carefully, some examples from numbers. So, divisibility is an important relation when we are talking about natural numbers or integers. So, divisibility talks about pairs of natural numbers, such that the first one divides the second one. So, we want (d,n) such that d divides n, remember this notation, this perpendicular bar for numbers denotes, this is not the same as the one that we use in set comprehension.

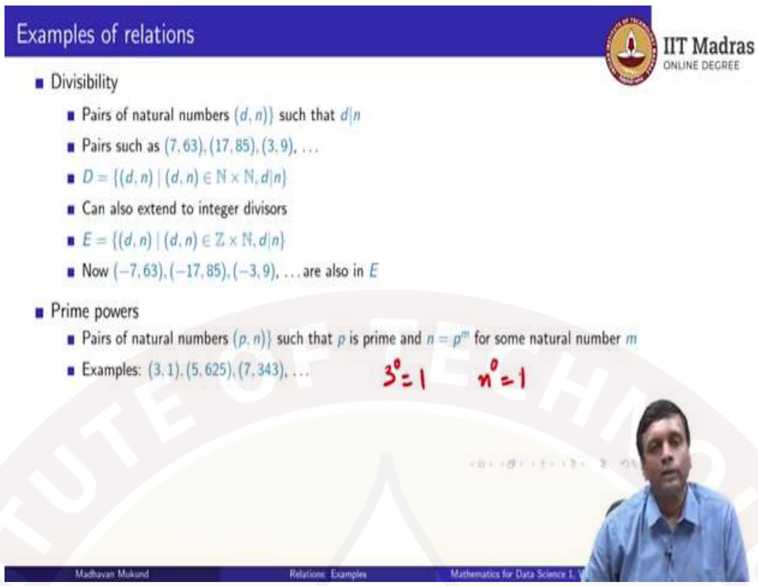
So, here it is an operation, arithmetic operation which says d divides n, so if I divide n by d, there is no remainder, it is a 0, d perfectly divides n. So, this would have this divisibility relation would have pairs like (7,63) because $7 \times 9 = 63$, or (17,85), because $17 \times 5 = 85$, and so on. So, we have a large number of pairs of divisors and numbers which the divisors divide equally, evenly. So, this we can write in our set comprehension notation because this is an infinite set, so we have no other way of listing everything.

(Refer Slide Time: 06:09)



So, we take all pairs N × N, (d,n), such that d | n. So, this is our filter. So, we want to generate everything of this form, but filter out under the condition that d must be a divisor of n and keep all such pairs. And this we can call d, the divisibility relation.

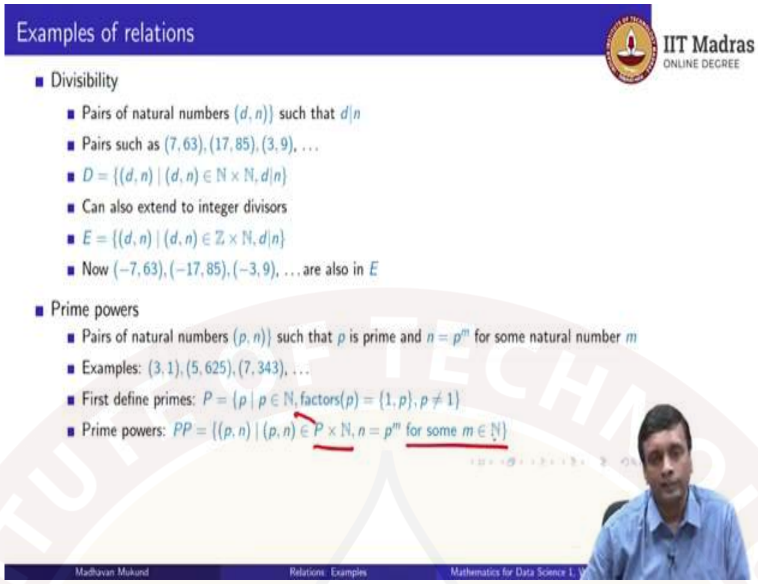
(Refer Slide Time: 6:2 1)



Now, this is the relation on pairs of natural numbers, so we only get positive divisors. If we extend it to integers, then we will get even negative divisors. We know that $(-7) \times (-9)$ is also 63, because the 2 negative signs will cancel out. So, if you extend the generating set from N to Z, from the natural numbers to the integers, then we get a larger set of divisor pairs. So, we get minus and plus elements for the same pairs that we had in the original relation.

Here is another example. Let us look at what we call prime powers. So, a prime power is something that is a prime multiplied by itself for a certain number of times. So, for instance, we can say that $5^5 = 3125$. So, $5^2 = 25$, 5^4 rather, 5^2, $5^3 = 125$, and $5^4 = 625$. So, 625 is a prime power, similarly $343 = 7^4$, so it is a prime power and so on. Why is (3,1) in this relation because anything to the power 0 is 1 by definition. So, $3^0 = 1$, in fact, anything to the power, so any number to the power 0 is 1. This is by definition. So, for every number comma 1 will be a prime power.

(Refer Slide Time: 7:41)

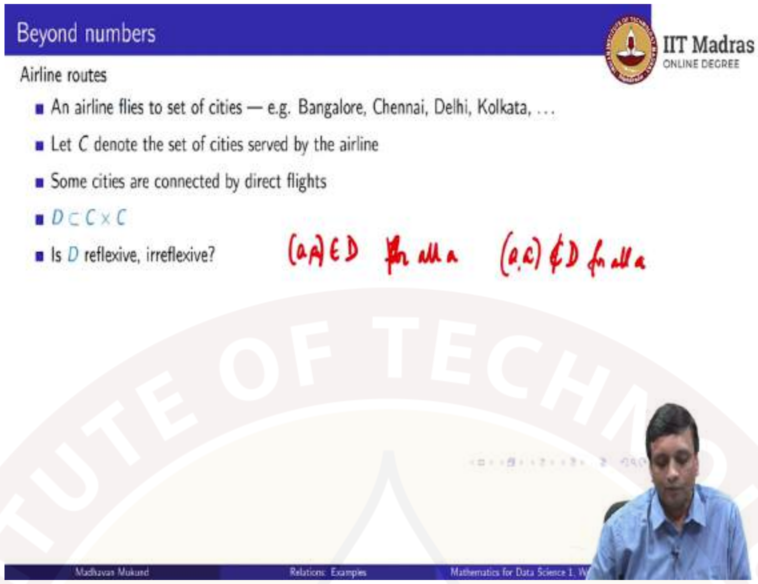


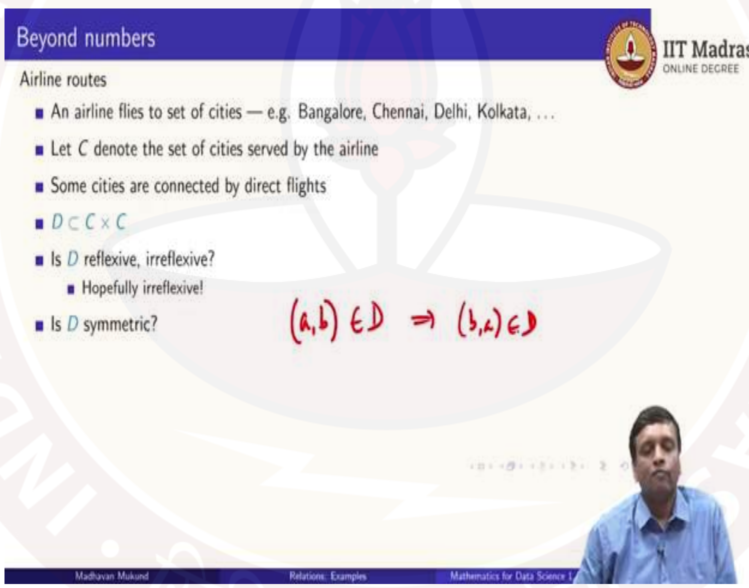
So, if you want to define prime powers, it is useful to first define primes. So, one way we can define primes is to say, give me a natural number, such as the factors of the natural number consists of exactly 2 elements, 1 and the number itself. And because in sets, we do not distinguish duplicates, in this definition, if I just say factors(p) = {1, p}, it includes a case where p is 1, because factors(1) = {1,1}, which is just 1. But I do not want to count 1 as a prime number. So, we also specify that P is not 1. So, this is the set of primes.

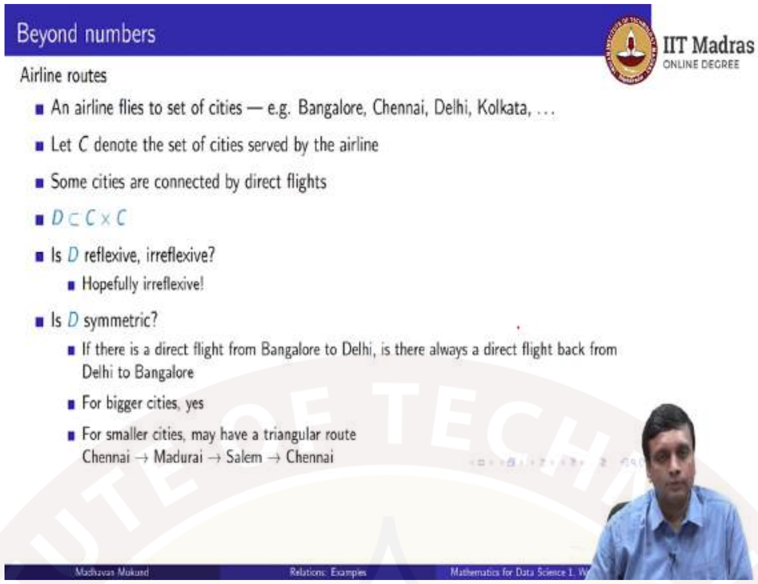
And now, we can say the set of prime powers is the set of all pairs in P × N, where P is defined above, P × N, such that n is the power of p. So, n = p^m for some m, which is a natural number, which could be 0. That is why we get (3,1). So, this is an example that we also talked about. It is saying that when you are writing the set comprehension, you can write these kinds of statements.

So, you do not have to be very precise about what you are writing mathematically in terms of notation, as long as the understanding is clear, there is no ambiguity about what you mean. So, you can write words like for some, you can also write it in a mathematical notation using symbols for there exists and for all and so on, but it is not necessary. As long as you are precise, you can use set comprehension notation in a flexible way.

(Refer Slide Time: 9:00)







So, these are relations in a formal sense. But why are we so interested in relations especially in the context of computing and data. So, let us look at relations which go beyond numbers. So, here is an example. Supposing we are talking about an airline, which serves a set of cities and we are interested in the routes that this airline serves. So, let us C be the set of cities where the airline operates. So clearly, the airline operates between some pairs of cities, but not all of them.

So, some of these cities are connected by direct flights and for other situations, you have to take a hopping flight which goes from city A to city B and then from city B to C. So, let us look at that subset D of direct flights between cities in C. So, this is an example of a relation. Not every pair of cities is connected by a direct flight. So, if you take all possible pairs of cities, some of them are connected by direct flights, and some are not. So, this way, information about an airline's route is really a relation in the sense that we mean.

Now, we have defined certain properties of relations, we said that the relation is reflexive. Now, this is useful to ask this question because we are talking about a relation between a set and itself. So, we can ask whether every element in the set is related to itself or is not related to itself. So, reflexive means that always we have (a,a) in D, for all, for every a. And irreflexive means, exactly the opposite of this is never in D and for all A.

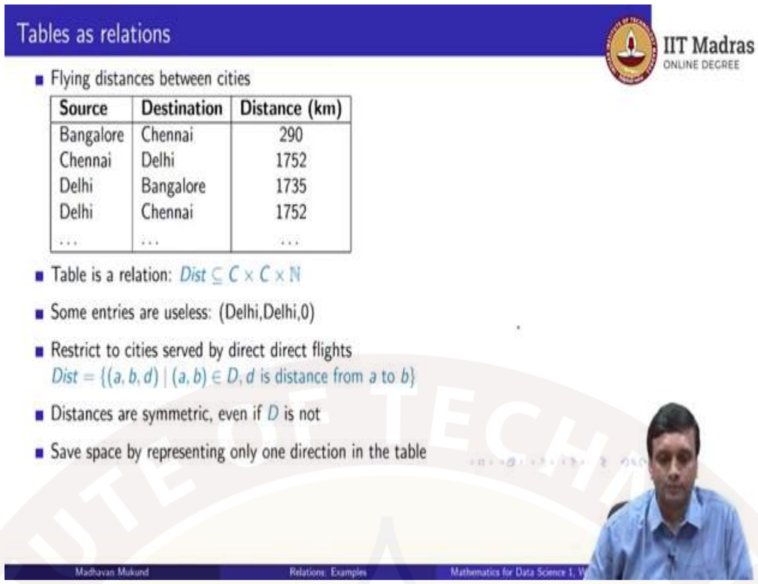
So, the question is, in terms of direct flights, is this going to be a reflexive relation and irreflexive relation or neither. Well, it is easy to see that this should not be reflexive. Because we do not expect an airline to actually operate a flight which takes off from an airport and then lands immediately in the airport. And in fact, we would precisely like it to be irreflexive, that is, this should never happens.

So, this should not be reflexive because we do not want every airport to serve itself and we want it to be irreflexive because we want no airport to serve itself. So, this is an example of an irreflexive relation. Now, is it a symmetric relation? So, symmetric relation says that whenever I have a pair of cities in the relation, then I will also have the reverse pair in the relation. So, if I can fly from one city to another directly, then I can also fly back.

So, concretely for instance, if I take any 2 cities a direct supposing there is a direct route from Bangalore to Delhi, then is there always a direct flight back from Delhi to Bangalore. Now, if you think about airlines, this is usually the case. But actually, if you look at domestic flights in particular, this is typically true only for the bigger cities, it will certainly be true for all the metro cities and the largest state capitals and so on. But if you look at smaller cities, this is not necessarily in the case.

For instance, it is quite common for airlines to serve 3 cities in a triangular route. So, you might have a flight that takes you from Chennai to Madurai, but if you want to come back from Madurai to Chennai, you cannot fly back directly, but you may have to fly to Salem and then come. So, between these 3 cities you can get from one to another, either directly or indirectly depending in which direction you are going. So, this relation is going to be irreflexive but not necessarily symmetric, it depends on the context.

(Refer Slide Time: 12:29)



Now, one thing you can do is to extend this to a table. So, here is a useful table that we might want to keep, which might be used to derive other things such as how long it takes to fly or how expensive a ticket is like to be. So, here we are just recording a fact which is what is the flying distance between a pair of cities. So, this table says that if the source is Bangalore and the destination is Chennai, it is 290 kilometers, whereas if the source is Chennai and the destination is Delhi, it is 1752 kilometers.

So, for every direct flight which our airline operates, you can record this distance and put it in a table. So, what is important to recognize and this is why relations are so useful in computing and data is a table is just a relation. So, every column represents a potential set of values. Here, the first column represents a possible city, so it is taken from the set C, the second column is also taken from the set C, the third column is a natural number.

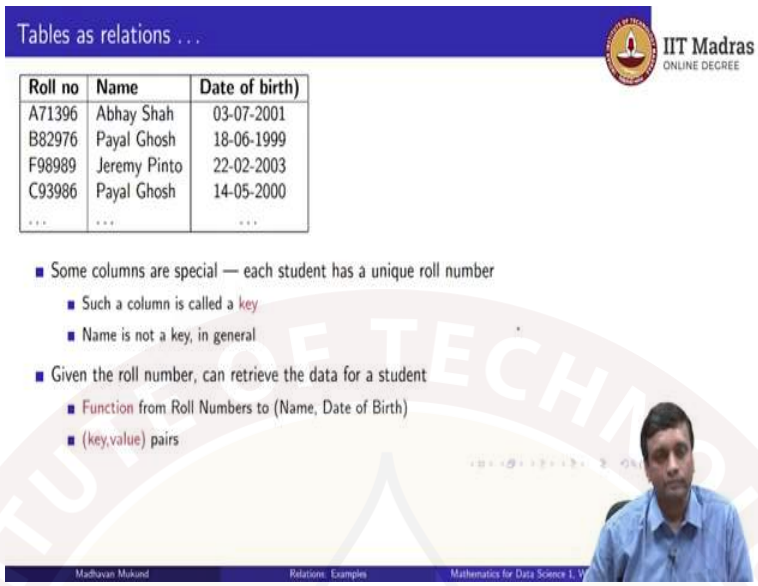
If you take pairs of cities which are the same, you could put 0, so it could be from Delhi to Delhi it is 0. So, in general, you have all possible pairs of cities and all possible numbers, but only some of them are interesting. Namely, when I have 2 cities which are actually connected by a flight and the distance the number is actually the real distance. So, it is a relation on $C \times C \times N$.

As we said, some relations are useless so we would not record them even though we know them. We know that for every city, the flying distance from the city to itself is 0, so there is no reason to record it in the table. The other thing is that unlike our direct flight's relation, this is actually a symmetric relation. So, first of all, we will only keep direct flights because we do not want indirect flights. But distances are definitely symmetric.

So, it doesn't really matter whether there is a direct flight from Chennai to Delhi and back or whether there is a direct flight from Chennai to Madurai and not back. It is enough to record the distance from Chennai to Delhi and Chennai to Madurai once each. I do not have to keep the distance from Delhi to Chennai separately as you can see above, in this example, Chennai to Delhi and Delhi to Chennai are both exactly the same distance 1752 because that is how distances work, distances are symmetric.

So, if we have symmetric entries, in a practical sense, when we represent a relation as a table, we can save on space by not recording the symmetric entries and making a note separately that this relation is symmetric. So, that is why it is important to know the property of the relation. It is not just an abstract question, is this reflexive, is this irreflexive, it is actually a practical consideration, a symmetric relation can be represented by only half the entries in the relation, the other half followed by symmetry.

(Refer Slide Time: 15:12)



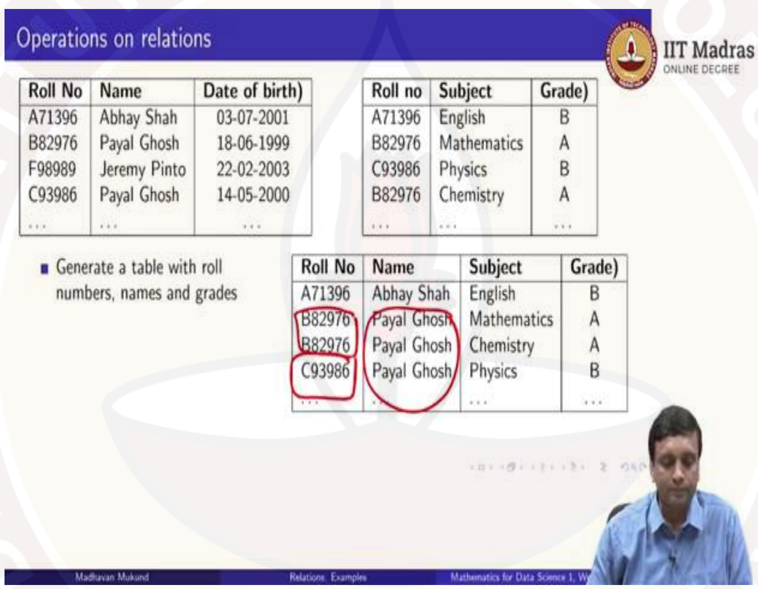
So, let us go further with this. So, another place where we often encounter tables are, for instance, when looking at data about people. Let us look at students. So, typically a college would record or a school would record information about students in this form. So, they would assign a roll number, then they would record maybe the name, the date of birth, and there would typically be other personal information like maybe their home address, phone number, and so on.

So here, what is important is that some columns are not natural in the sense. So, we know that everybody has a name and they are born on a particular date, but this roll number is actually assigned to them by the school or college. And this is something which is designed to be unique, so no 2 students get the same roll number. So, this kind of column is called a key. And this is because we want to identify, define each student directly and individually without getting confused about which student we are talking about.

And unfortunately, the other columns are not keys, 2 students could have the same name. And it is even possible for 2 students to have the same name and the same date of birth. So, we cannot rely on the fact that the other columns will uniquely distinguish. So now, if we have a unique roll number for every student, then each row is identified by the roll number. So, we can actually think about the row as being something where if I give you the roll number, you can tell me which row it is and give me the other values in that thing.

So, this is more like a function. A function says given an input give me a unique output. So, given a roll number, tell me all the values associated with the roll number, the name, the date of birth, and so on. So, this kind of a stored table is also called sometimes a set of key value pairs, given the key there is a unique value. I can change the value for a given key by updating it. But if I add a new entry, I have to add a new key so there is no confusion.

(Refer Slide Time: 17:03)

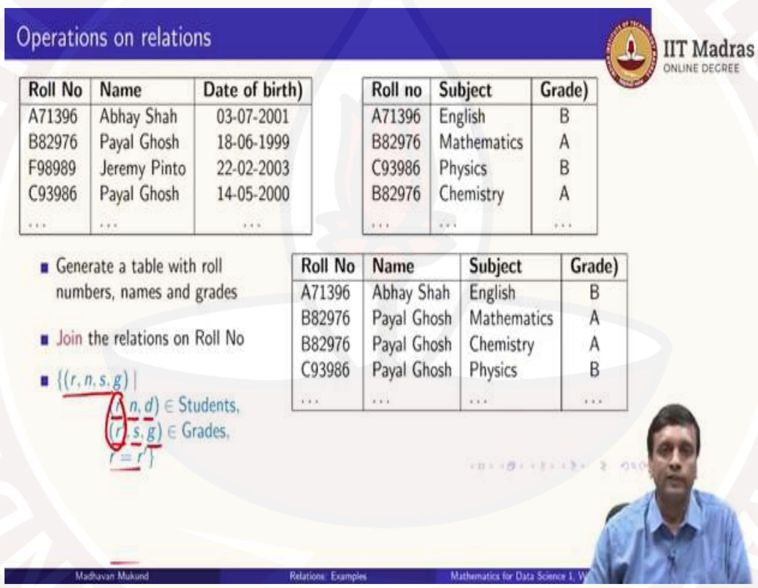


So, usually a school or college will maintain more than one table of this kind. For instance, there might be a separate table, where we maintain the marks of the student or the grades of a student in the courses that they do. And here for conciseness, we might keep only the roll numbers and the subject names and not the names of the students. So, for instance, in the second table, we have the roll number, subject and the grade. Here is a typical requirement when we have to generate a report card.

The grade card has, the grade table has the roll number and the subject and the grade but it does not tell us who the student is. And that is, for example, it may be difficult for an outsider who except for the student themselves to know whose roll number belongs to whom, because nobody would recognize these strange character sequences. So, we want a table that looks like this which has the roll number and extra column with the name which is not there in the grade table which is taken from the first table and then we want the subject and the grade.

And here, we see why it is important to have keys because we have this name Payal Ghosh, which is ambiguous, there are 2 Payal Ghosh'es. And in fact, they have 2 different entries in this table because they have 2 different roll numbers. So, the Payal Ghosh who got an A in mathematics is not the same as the Payal Ghosh who got a B in physics. So, this is an operation which combines these 2 tables. And remember that a table is a relation.

(Refer Slide Time: 18:23)

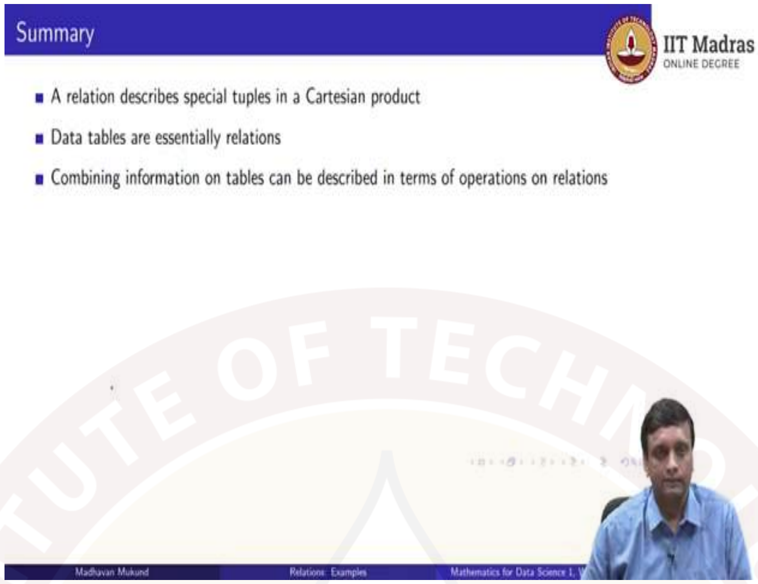


So, this operation, which combines 2 tables is also an operation which combines 2 relations, and it is an important operation in computing and in data science called a Join. So formally, a Join takes tuples from 2 relations and combines them on common values. So here, for instance, you take any arbitrary roll number, name and date of birth from students, you take any arbitrary roll numbers subject and grade from grades, but you want that the roll number in the roll number of the 2 sides belongs the same.

So, the r comes from students and the r′ comes from grades and you want r = r′. And if this is the case, then you put out a new tuple, which combines the n from the left hand side throws away the date of birth, we are not interested in preserving the date of birth, keeps the n and keeps the subject and the grade s and g and of course keeps the roll number which is the same on both sides.

So, this will ensure that we do not get rows merged, where they correspond to 2 different students. So, the marks for Abhay, or the grade for Abhay will not be merged with the name and date of birth for Jeremy Pinto, because they have 2 different roll numbers. So, this is called the Join and this is a very important operation on relations, and therefore on tables. And this is something that we use implicitly all the time.

(Refer Slide Time: 19:39)



So, to summarize, a relation describes special tuples in a Cartesian product. And what is really important for us from a computing and data science point of view is that we work with tables all the time and tables are really relations. So, that is why relations play such a central role in many of the things that we are going to look at. So, it is important to get the terminology of relations right.

And when we combine information on tables, these are actually operations on relations such as the Join operation that we described, this is only one kind of Join we may have different types of operations, which we will see in other courses later on. But please, keep in mind that tables are relations. Thank you.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

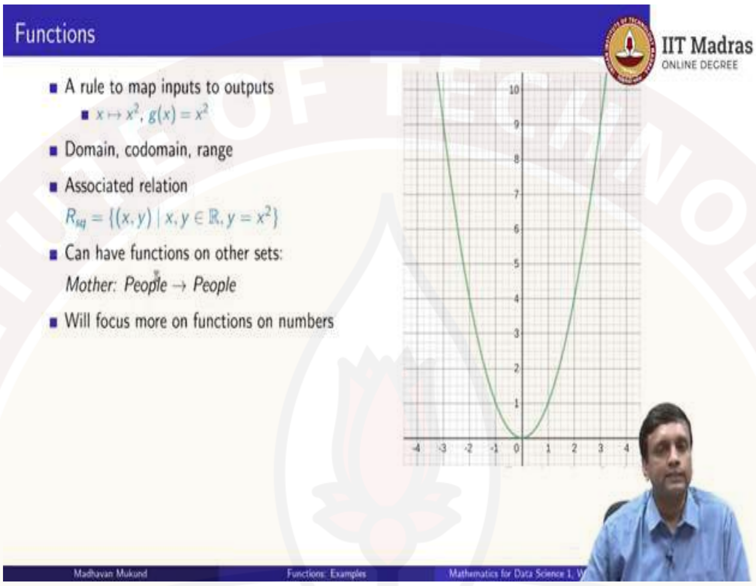
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Madhavan Mukund
Department of Computer Science
Mathematical Institute, Chennai
Lecture-1.7B
Function: Examples

So, let us take a closer look at functions now.

(Refer Slide Time: 0:19)



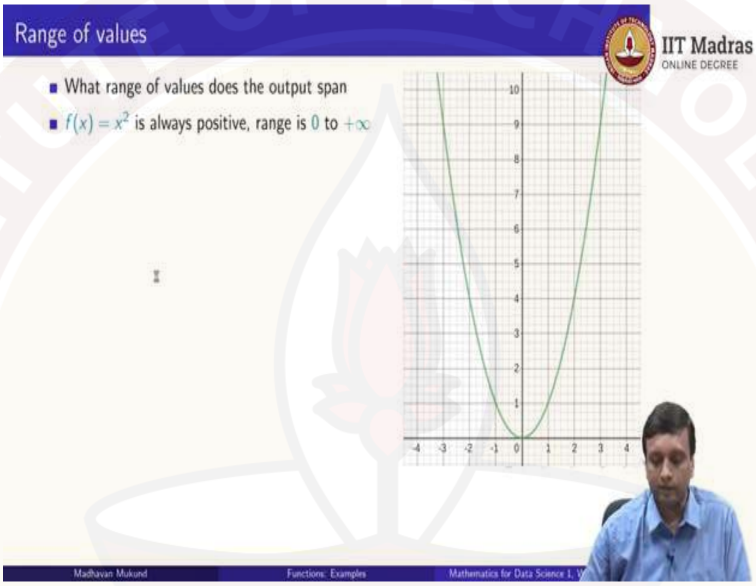
So, remember that a function is a rule that map's inputs to outputs. So for instance, if we are looking at numbers, a function could take an input x and map it to x^2, which can also write given a name saying $g(x)$ is equal to x^2, which says g is the name of a function, which when it takes an input of the form x produces an output of the form x^2.

And with such functions, we have a notion of a domain that is what are the inputs that are allowed, the set from which we take inputs. Codomain, what is the set to which the outputs belong and range which is the actual outputs that this input set generates for this given rule. So, for instance, we have for this function this relation associated with it, all pairs x comma y such that x and y are reals. So, the domain and the codomain are both reals, the rule is y equals x^2, so that is the filter that we put, we only want such pairs.

And if we plot all the points which belong to the relation, we get this graph on the right. And this actually tells us that the range of the function even though the codomain is all reals, the range of the function actually keeps this function above 0, so we only get non negative reals as outputs. Now, we are not restricted to looking at functions on numbers, we can also look at functions on other sets.

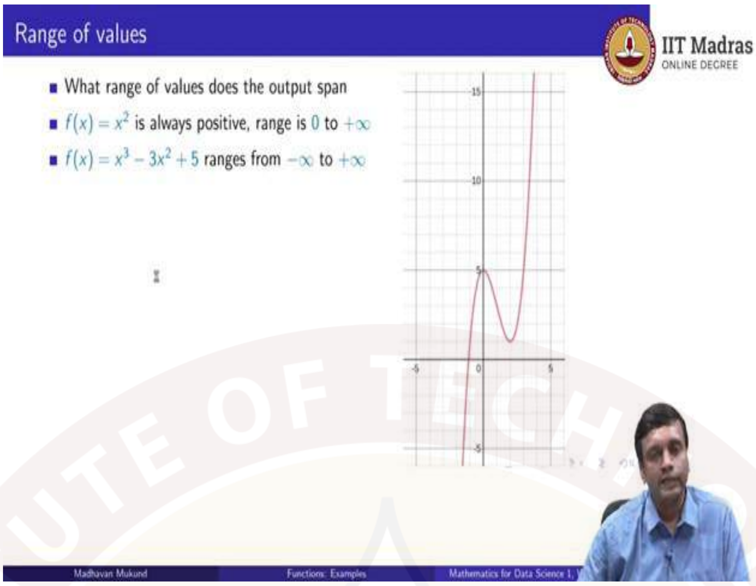
So, for instance, if we look at the set of all people in the universe, in the world, in the country, in any range of geographical regions, we can look for the function mother which says, given a person, this will map the person uniquely to the mother of that person. So, this is a function because every person has 1 mother. So, in this lecture, and in general, when we are talking about functions in this course, we will look more at functions on numbers. So, let us look at these a little more closely. What are the questions that we really want to ask about functions on numbers?

(Refer Slide Time: 2:09)



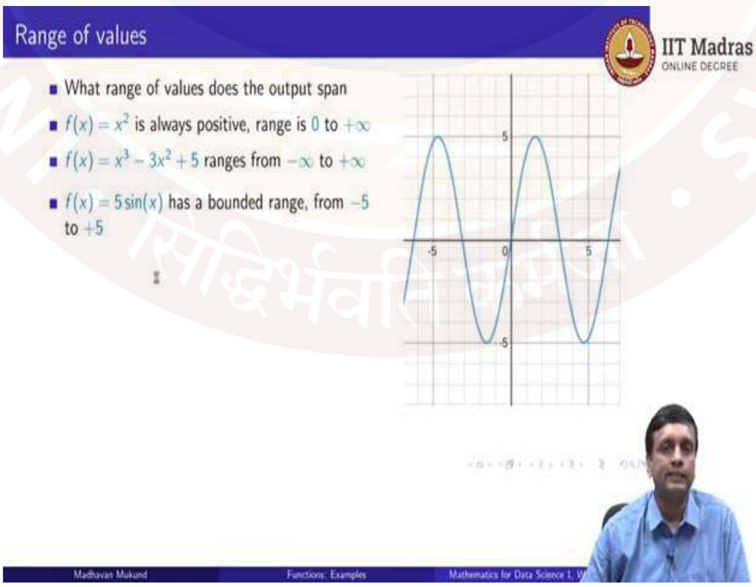
So, one of the basic questions is, what are the ranges of the values that we can get. So, in other words, we have a core domain. But what is the range of values that we can actually achieve through the function. So as we saw, this square function, $f(x)=x^2$ is always positive, so we always get something between 0 and $-\infty$, there is no upper bound, but we never get something which is negative.

(Refer Slide Time: 2:35)



On the other hand, if we take a cubic function of this form $f(x)=x^3-3x^2+5$, then when x becomes very small, the x^3 becomes very small because the cube of a negative number is a negative number. So, cube have a large negative number, I mean magnitude, the $(-1000)*(-1000)*(-1000)=-10^{-9}$. So, as we go into negative, large negative values, we can at large negative outputs, same for large positive values. So, this has a range from minus $-\infty$ to $+\infty$.

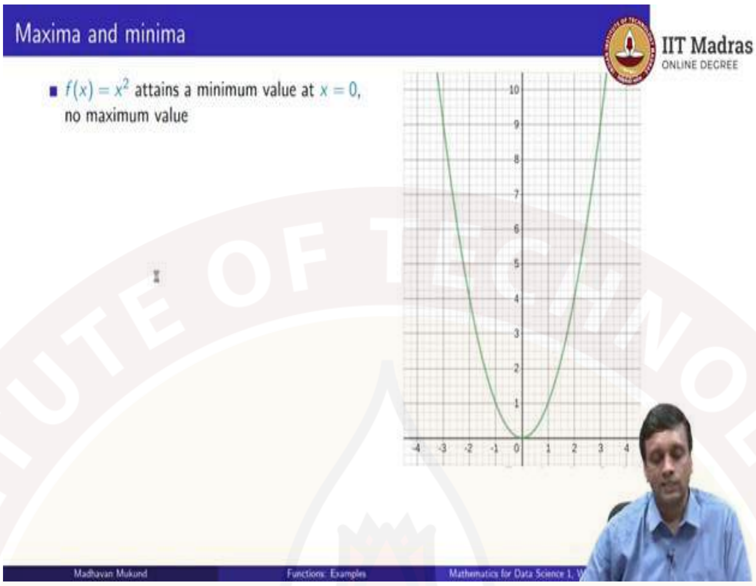
(Refer Slide Time: 3:05)



And then there are some functions like the trigonometric function $\sin x$, which oscillate between an upper bound and lower bound. So, if you take $\sin x$, usually it is between +1 and -

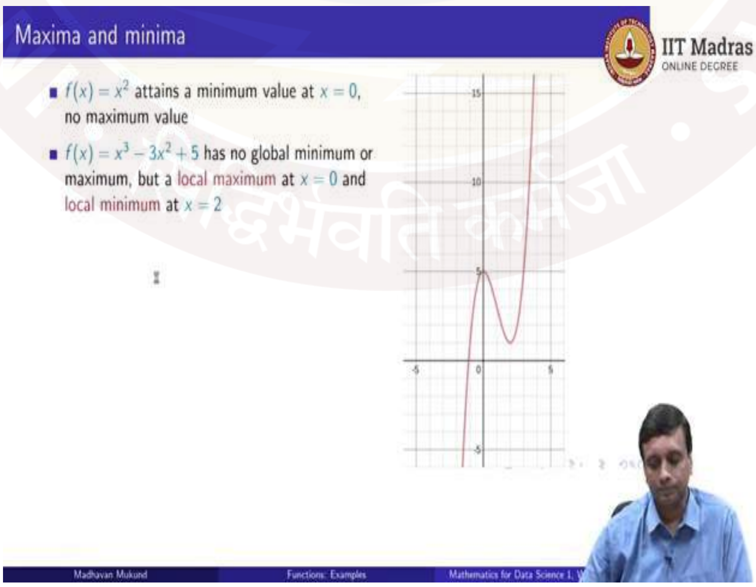
1. If we take $5\sin x$, then it will be between -5 and $+5$. So, this has a bounded range. Even though we consider all possible inputs, we never go outside this range from -5 and $+5$.

(Refer Slide Time: 3:28)



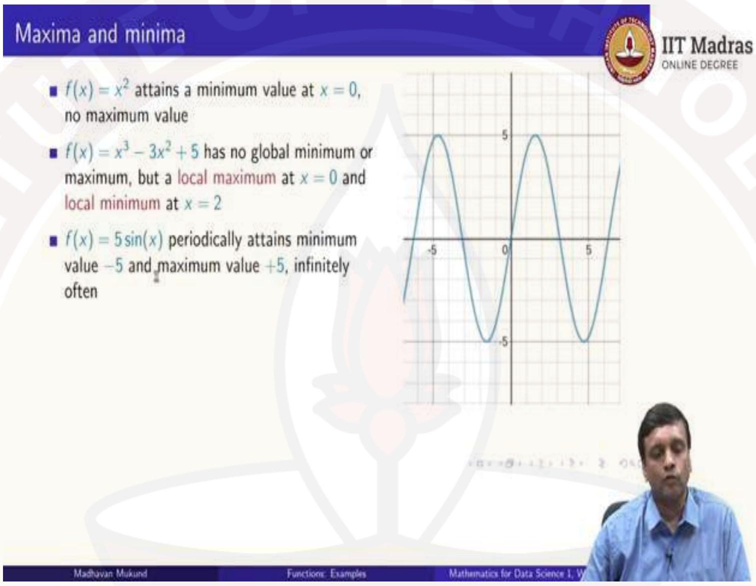
Now, within the range of values that it can take, we are often interested in specific points, in particular, where the value are a minimum and where they are maximum. So, for instance, this function that we have seen before $f(x)=x^2$, it is clear from the graph on the right that at 0 the output is 0 and at all other points is bigger than 0, so it attains its minimum value at 0. And because it keeps growing indefinitely in both sides, there is no maximum value.

(Refer Slide Time: 3:57)



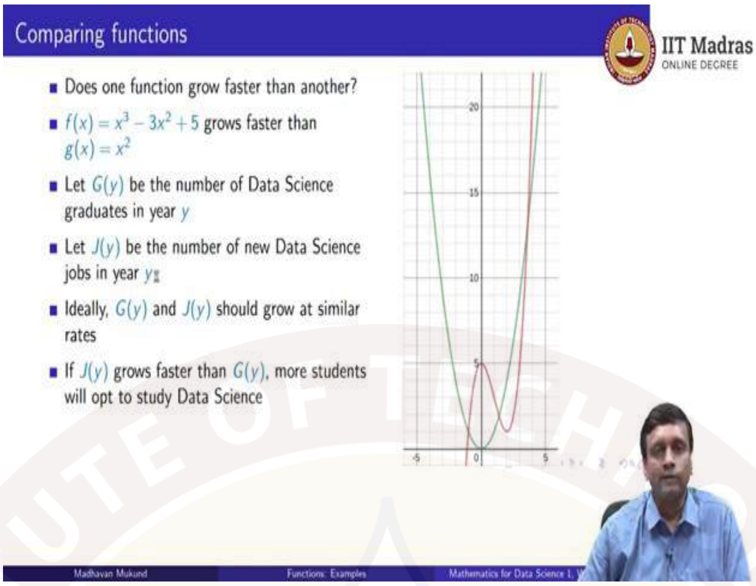
Now, the cubic function we said grows arbitrarily small as we go to the negative inputs and arbitrary large. So, there is actually no maximum and minimum, but it has an interesting behavior in between because it zigzags it goes up, comes down and goes up again. So, there is something called a local maximum and a local minimum. So, at $x=0$, it turns around, so it achieves a maximum value and starts falling briefly and then at $x=2$ it turns around again. So, it achieves a local minimum and goes up again. So, we are interested in finding out where these local maxima and minima are for various reasons.

(Refer Slide Time: 4:32)



And similarly, if we look at something like $\sin x$, then it has, of course, local minima and maxima, $-¿5$ is a local minimum and $+¿5$ is a local maximum, it is also a global minimum and maximum because these are the maximum and minimum values that the function can ever attain. And now, these values are actually attained infinitely often periodically as we go from left to right.

(Refer Slide Time: 4:53)



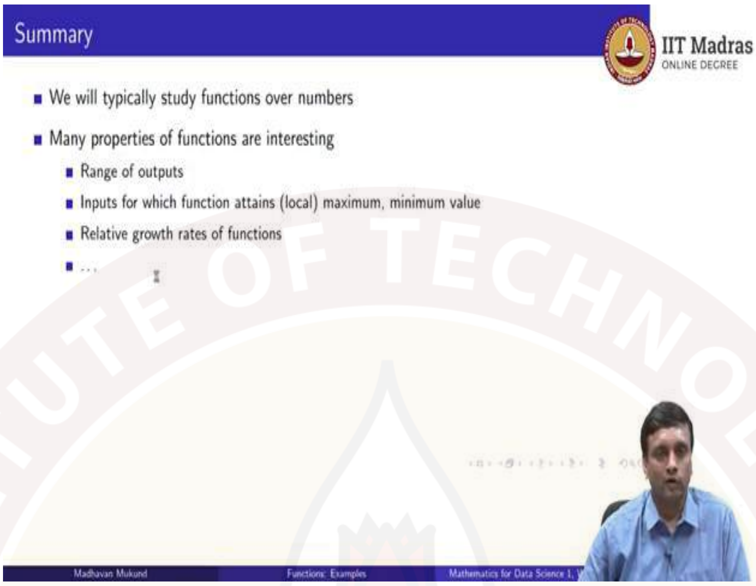
Another thing which we are interested in about functions is how fast they grow. Thus one function grow faster than another. So, if you look at our 2 functions, $f(x)=x^2$, and $f(x)=x^3-3x^2+5$, and we look at their 2 graphs, then it is very clear that the red line, although initially on the right, it is below the green line, it overtakes it, and after that, it is never going to be below the green line. So, in this way, the cubic function grows faster than the square function.

Now, why is this interesting? Well, we often see this informally stated in various contexts. So, let us look at a context which is relevant for you. So, let $G(y)$ be the number of data science students graduating in a year y. So, as the year increases, so we go from 2020 to 2021, and so on, the value G takes a certain number and hopefully because courses are growing, this number is increasing.

At the same time, there are jobs being created in data science. So, let $J(y)$ be the number of new data science jobs in a year . Now, ideally, you would like that these 2 are comparable, that the jobs are growing because the number graduates is growing and vice versa. If the number of jobs increases more than the number of graduates then there is a demand for graduates and of course, more graduates will opt to study data science. So, you would expect a demand for this kind of course.

Of course, the unfortunate case might happen the other way around, if suddenly there is a slump in demand, then people who graduate with a degree in data science will not be employable and then there will be a reverse trend. So, these are some of the reasons why when we look at data, we are interested in comparing the growth rate of functions and we will look at this in the context of the functions that we study mathematically.

(Refer Slide Time: 6:33)



So, to summarize, we will typically study functions over numbers. And we are looking at many properties of these functions which are interesting to us, for instance, the range of outputs, where these functions attain local minima and local maxima and what are their relative growth rates and many other things which we will come across as we go along. Thank you.

IIT Madras

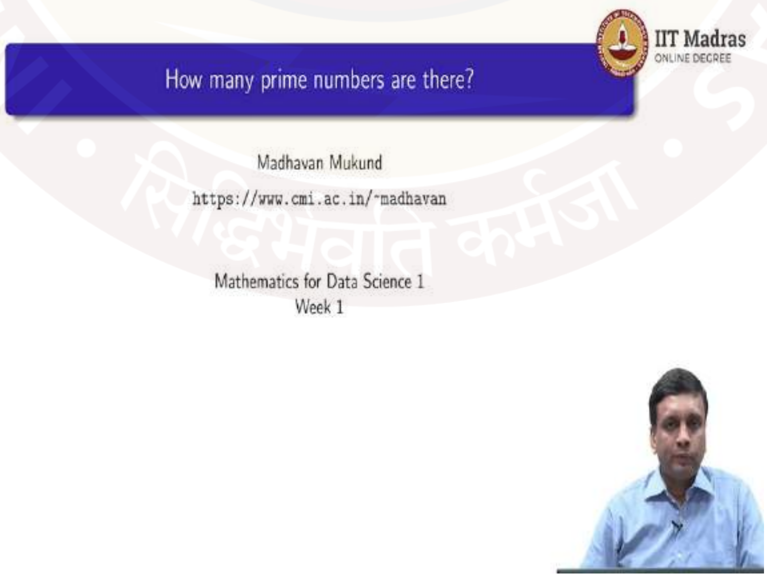
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture – 08
Prime Numbers

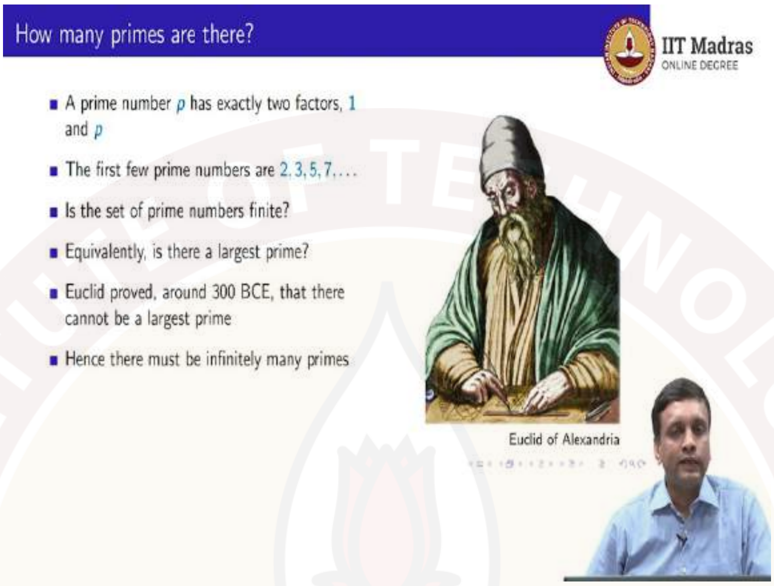
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, when we looked at the natural numbers, we talked about divisibility and we talked about the prime numbers. So, we know that the prime numbers start with 2, 3, 5, 7 and so on. So, how many prime numbers are there?

(Refer Slide Time: 00:25)

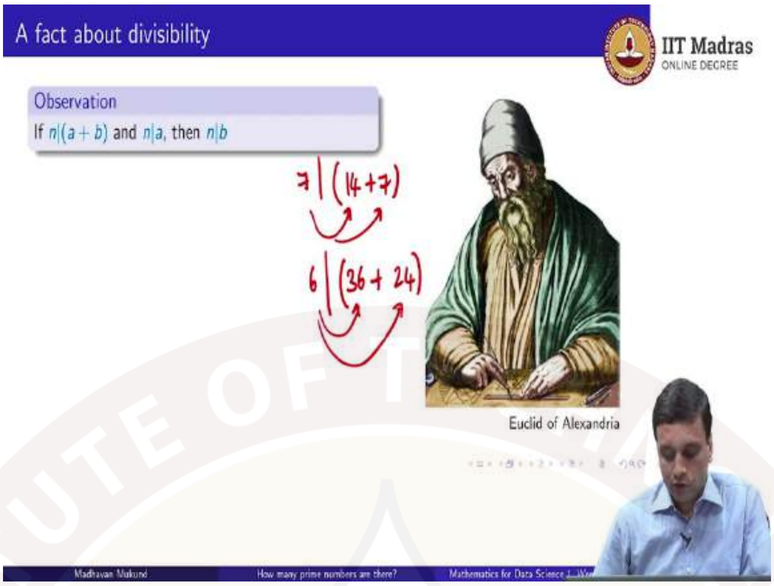


So, remember that a prime number is something that has only two factors 1 and itself. Now, it must have exactly two factors. So, 1 is not a prime. So, the first few prime numbers are 2, 3, then 4 is not a prime – because 4 is divisible by 2, then 5, again 6 is not a prime and so on. So, the question is, is this set of numbers these prime numbers is it a finite set or are there infinitely many prime numbers?

Now, if there is a finite set of prime numbers, there will be a largest prime number. So, the same question can be asked by asking is there a largest prime? So, if it is a finite set, in that finite set, there will be a largest one. And if there is a largest one, then below that largest one there are only finitely many numbers, so there can only be finitely many primes. So, asking whether the set of primes is finite is the same as asking whether there is a largest prime.

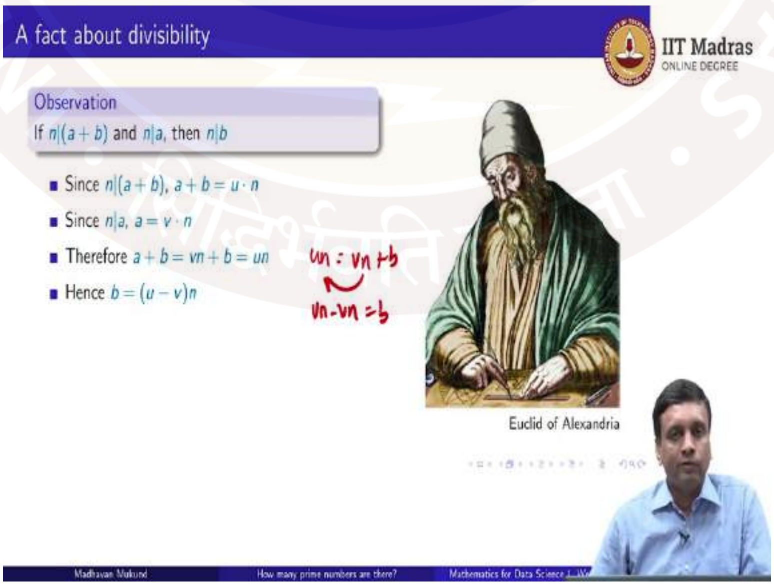
So, what we are going to see is a version of a proof that goes back to Euclid from about 300 BCE, which shows that there cannot be a largest prime. And as we argued if there is no largest prime, then it must be that the set of primes is actually an infinite set.

(Refer Slide Time: 01:35)



So, to go ahead with this we need a basic fact about divisibility. So, this says that if a number divides a+b and it also divides a, then it must divide b. So, let us look at an example. So, supposing you say that 7 divides 21, and I write 21 was 14+7 then 7 also divides 14, and therefore, it also divides 7. Similarly, if I say 6 divides 36+24 which is 60; then since 6 divides 36, it must also divide 24 right. So, this is not very difficult to prove. So, let us prove it just to get a feel of how such proofs go.

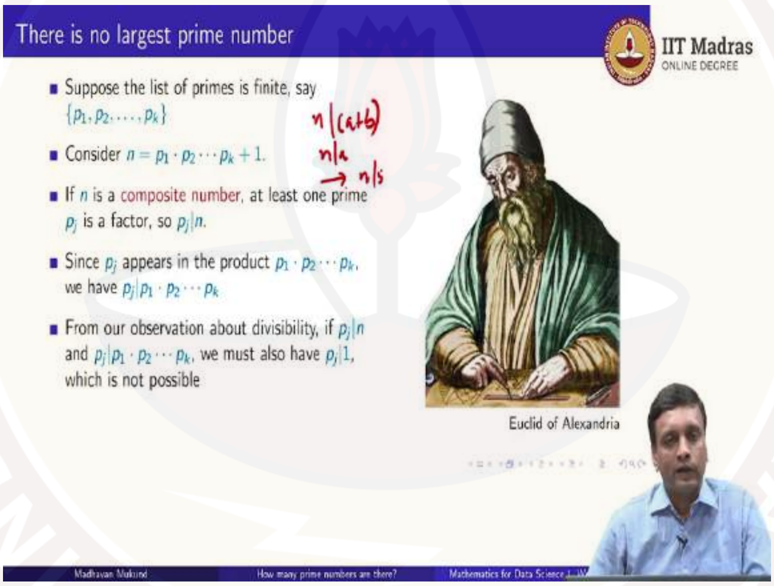
(Refer Slide Time: 02:17)



So, since n divides the sum a+b, a+b can be written as a multiple of n. So, let us call it u times n. Similarly, since we have assumed that n divides a, a can also be written as a multiple of n; let us call it v X n. So, what we are told is that any a + b is u X n for some u, a itself is v X n for some v. And the question is b also some multiple of n does n divide b?

Well, because of what we have just discussed a + b can be written as v n + b, because a is v n, and the sum v n + b which is the same as a + b is in fact u n. So, now, we can do some simple rearrangement. So, we can take u n = v n + b, and just take the v n to the other side and we get u n - v n = b and so b is (u – v) times n. So, this simply proves to us that if a number divides a sum and it divides one part of that sum, it also must divide the other part of the sum. And we will use this in order to show Euclid's result.

(Refer Slide Time: 03:21)



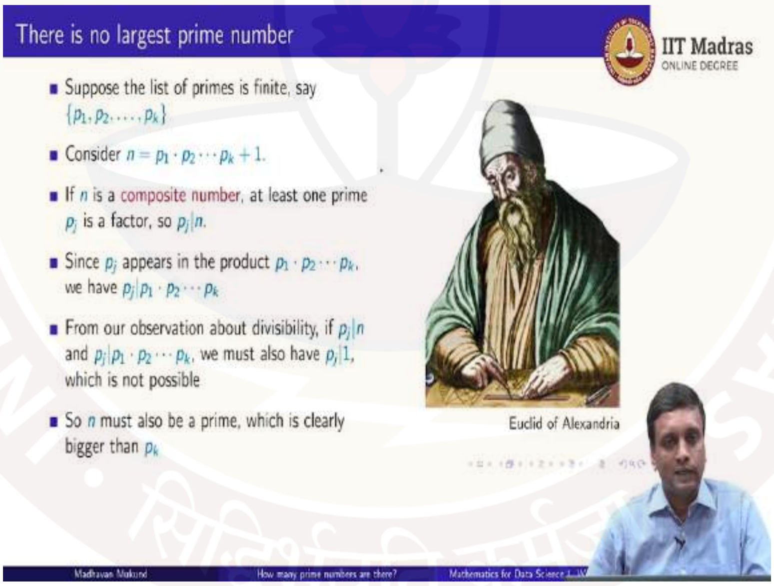
So, what Euclid said is that suppose the list of primes is finite. So, if it is finite, then we can list them out and it is a finite set, so it is some $p_1$to p_k. We do not have to be in any particular order. we can assume that p_1 is the smallest one; it is 2, $p_2$is 3 and so on. But it does not really matter as long as this exhaustively completes all the primes.

Now, we construct a new number which is the product of all these primes, you multiply all these primes by themselves to each other and then we add 1 right. So, n is p_1X p_2X ...X p_k+1. So, now, the question is what is the status of n? So, since we have assumed that the list of primes is finite, n must be a composite number, because this is not one of the primes that we had before right, it is bigger than all of them because it is the product of all of them plus 1.

Now, since it is a composite number it must have a factor other than 1 and itself. And because we have listed out all the primes one of the primes among them must be a factor. So, let us assume that p_j is a factor. So, p_j divides n right. So, there is one in this p_1 to p_k, there is a p_j which divides n. But on the other hand, let us look at this part right. The first part the first part is the product of all the prime So, p_j appears in that product.

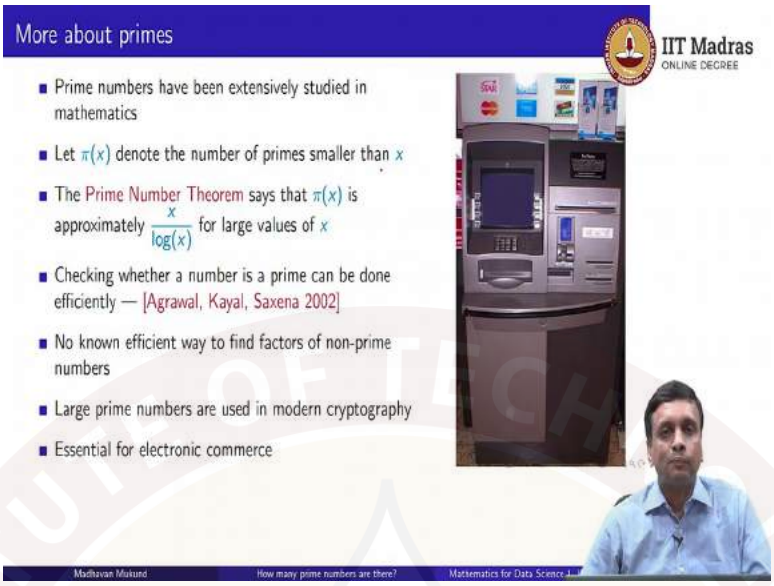
So, if it is one of the factors of the product, it must divide the product right. So, p_j divides n, and p_j also divides one part of the sum. So, remember what we said that if some number n divides a + b and if some number n divides a also, then n must divide b. So, in this case a+b is the product of the primes plus 1, and a itself is a product of the primes and we have argued that there is one prime p_j which divides both of these. So, therefore, by that divisibility result that we showed in the previous slide p_j must divide 1. But of course, we know that p_j is a number bigger than 1, it cannot divide 1. And so we have a contradiction right.

(Refer Slide Time: 05:21)



So, what is the contradiction? Well we assume that n was a new number was a composite number because we have exhausted all the primes, but in fact, it cannot be composite because then we cannot find a proper divisor for it among the primes. Therefore, n itself must be a prime. And notice by construction n is actually bigger than all these. So, it also shows that there is no largest prime, because for any set of primes we can always construct a larger prime. So, this is essentially what Euclid did.

(Refer Slide Time: 05:47)



So, we know more about prime numbers. So, prime numbers are very mysterious because their distribution is kind of unclear, but they also have important properties as we will see. So, prime numbers have been extensively studied in mathematics in an area called number theory. So, one of the things that is studied about prime numbers is how they are distributed. So, as we go a larger and larger in the set of natural numbers, how frequently do we find primes?

So, $\pi(x)$ is supposed to denote the number of primes that is smaller than any given number x. So, for instance, π (4) would be 2, because 2 and 3 are the only 2 primes below 4; π (10) would include 2, 3, 5 and 7. So, π (10) would be 4 and so on.

Now, as you go larger and larger, the gaps between the primes become larger. And in fact, you can prove amazing things like the prime number theorem which says that $\pi(x)$ is approximately x / log x for large values of x. Now, it does not matter if you do not understand what this means, but it is important to understand that this is a very significant type of argument that you can give about the distribution of a set of numbers which is quite in a way randomly distributed.

Now, in terms of modern applications of primes, it might seem that primes are very strange things, and we would only need to study them a number theory. In fact, the famous mathematician G. H. Hardy once said that he was very proud of the fact that he did number theory and nothing that he studied had any application. Well, it is not quite true because primes as we will see are actually quite useful.

So, one of the questions that you might want to ask is given a number check whether it is a prime. Now, of course, there is a brute force way of doing it which is to try and enumerate all the factors by looking at all the numbers below n and dividing n by them, but that is not considered to be an efficient way to do it. And in fact, this was proved by three Indian computer scientists from IIT Kanpur, Manindra Agrawal, Neeraj Kayal, and Nitin Saxena in 2002, and it is one of the breakthrough results in theoretical computer science in the history of the subject.

So, checking whether a number is prime can be done efficiently. But what about the other question, if I know a number is not a prime, can I factorize it? So, we know number is not a prime, but how do I find two non-prime, two non-trivial factors that is not 1 or itself. Now, it turns out that there is no efficient way to do this. So, this is quite paradoxical. We can check whether a number is prime or not, but if it is not a prime we can factorize it fast. And this in fact is the reason why we are so concerned about prime numbers, because we would like to find numbers which are not prime, but which are actually products of large primes. So, their factors are only large prime numbers, and this is a very important in cryptography.

And cryptography in this sense is something which affects not just you know military secrets, but it affects us in day-to-day life because whenever we do electronic commerce our transactions are protected by cryptography to prevent unauthorized transactions from being executed on our behalf or to prevent them from being tampered with they are all encrypted. And a lot of this encryption is based on the existence of large prime numbers, and the fact that factorizing the product of two large primes is difficult. So, prime numbers though they are very exotic in number theory are actually a very, very important part of our day-to-day life.

IIT Madras

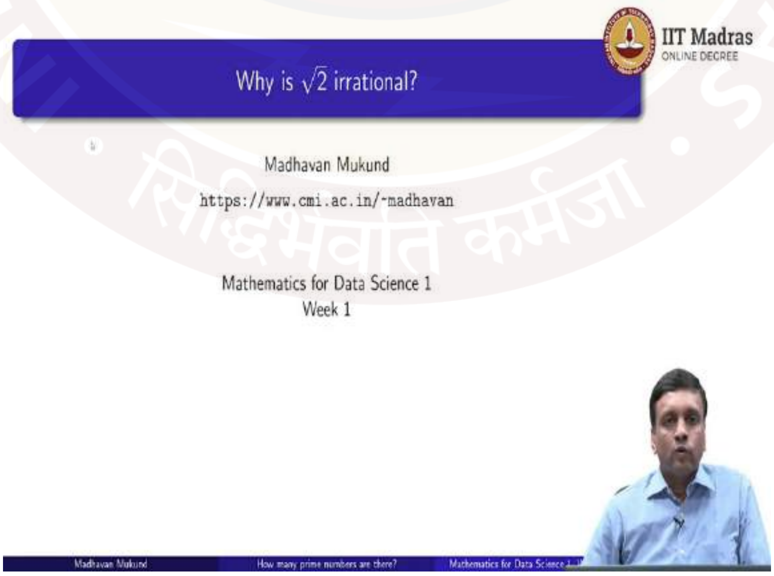
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

Week - 01
Lecture – 09
Why is a number irrational?

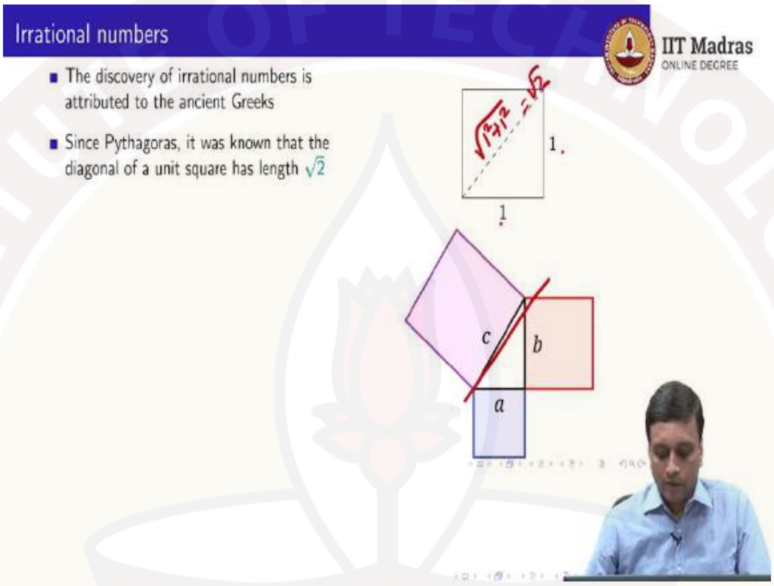
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



When we looked at the different types of numbers, we started with the natural numbers, move to the integers, then to the rationals which are expressed as $\frac{p}{q}$. And then we argued that the rationals do not exhaust all the numbers that we need; and in particular, we claim that the √❑ cannot be expressed as a rational numbers, so it is what is called an irrational number. So, let us try and ask why √❑is an irrational number.

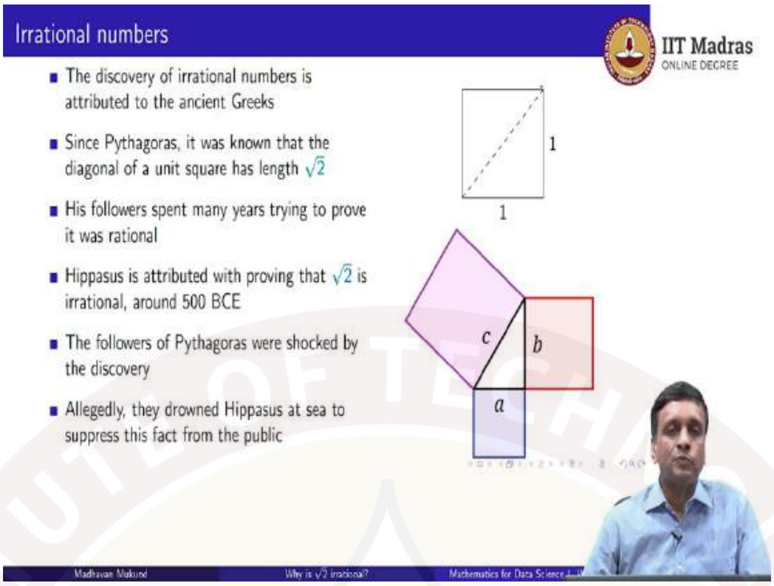
(Refer Slide Time: 00:37)



So, the discovery of irrational numbers actually is attributed to the ancient Greeks; and in particular, it comes from Pythagoras. So, remember that in Pythagoras's theorem which you must have studied in school. If you have a right angled triangle, then the square on the hypotenuse that is the square on the long diagonal side – this one, has an area which is the sum of the squares on the other side. So, in other words, if you have a right angled triangle and you measure the three sides, you get $a^2+b^2=c^2$. So, from this, knowing a and b, you can compute c.

So, in particular, if you draw a square which has one and one as its two sides, then this must be the √❑which is the √❑. So, you can actually physically draw if you assume that you can measure out a unit length using some kind of a measure, then by drawing a square, you can actually construct a length √❑.

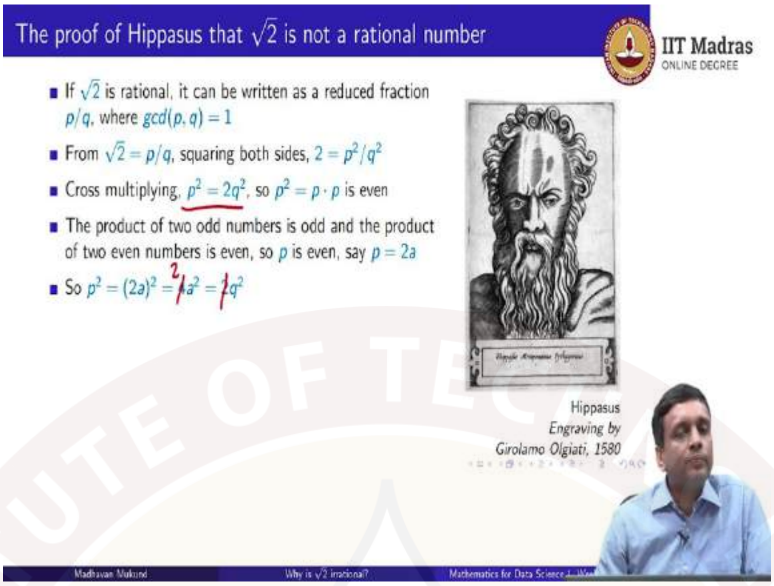
(Refer Slide Time: 01:35)



So, for Pythagoras it was very important to understand how to describe the √❑as a rational number, and he and his followers and many times many years trying to prove that in fact it could be expressed as a rational number. Much after Pythagoras, about 50-60 years after Pythagoras, one of his followers Hippasus is claimed to have proved that √❑is irrational this was around 500 BCE.

Now, the followers of Pythagoras had a very mystical idea about numbers, and they felt that numbers could solve everything. And in particular they were very keen that rational rational numbers should form the basis of all of what we could call it modern day time science and philosophy. So, the followers of Pythagoras were really shocked by this discovery of Hippasus, they found it to be a, I mean they could not argue with it; at the same time they felt that this discovery could not be revealed to the public because they felt it was very dangerous. So, in fact, it is said that they allegedly drowned him in the sea to prevent this from being made public. So, the √❑being irrational has a rather colorful history. And let us see now how Hippasus proved that this was actually the case.

(Refer Slide Time: 02:41)



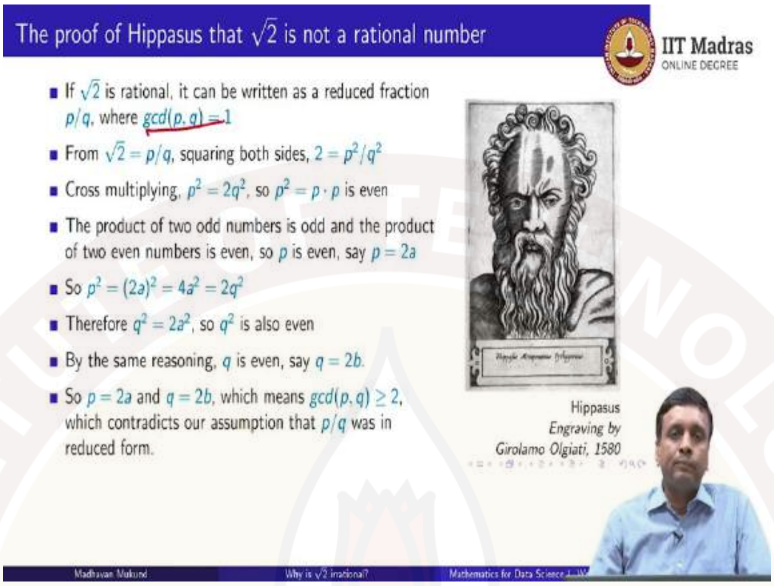
So, let us assume as in many of our arguments. Let us assume that $\sqrt{2}$ was rational. So, if it is rational, then we know that it can be written as a ratio or fraction of two integers p and q; and in particular we can assume that it is in reduced form. So, p and q have no common divisor, their gcd is 1. So, if we take $\sqrt{2}$ is equal to $\frac{p}{q}$, and we square both sides, then $\sqrt{2}$ times $\sqrt{2}$ is 2 on the left hand side, and $\frac{p}{q}$ times $\frac{p}{q}$ is $\frac{p^2}{q^2}$. So, we get 2 is equal to $\frac{p^2}{q^2}$. So, we can cross multiply as usual, take the q^2 from the denominator on the right hand side to the left hand side numerator, and we get $2q^2$ is equal to p^2.

So, what is p^2? p^2 is p X p. And if it is of the form 2 times something, then it is an even number, because an even number is something which has 2 as a factor. So, p^2 has 2 as a factor. So, p^2 is an even number. Now, it is a basic fact about natural numbers that if you multiply two odd numbers, you get an odd number; and if you multiply two even numbers, you get an even number. So, if p^2 is even, and p^2 is p X p, then both p and p – the two copies must both be even; so p must be an even number in other words.

So, if p is an even number, then we can write p as 2 times something because p is even p must be of the form two times something say 2a right. So, from this initial assumption, we have concluded that the numerator of this fraction which represents $\sqrt{2}$ is actually an even number of the form 2a.

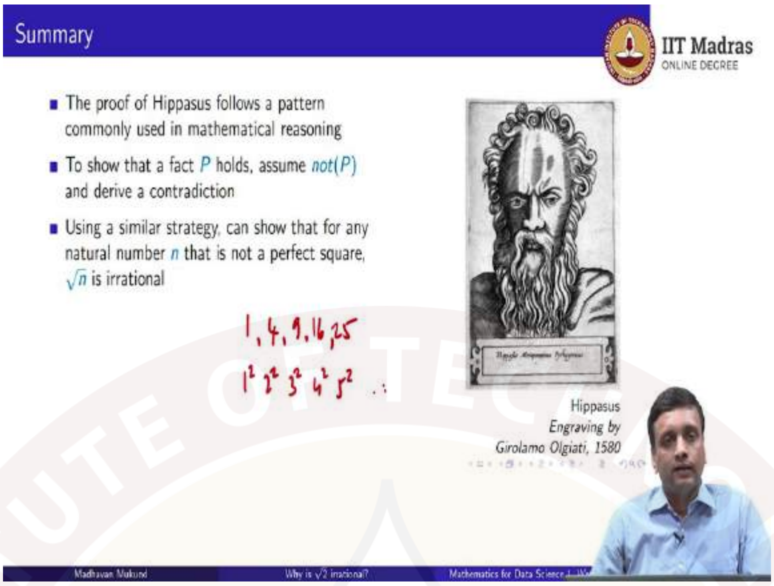
So, now, let us substitute in this equation for p^2right. So, $p^2$is $(2a)^2$ is 4 times a^2. So, now $4a^2$ is equal to $2q^2$. So, now, we can cancel right. So, we can take this 2, and this 2, and cancel it.

(Refer Slide Time: 04:39)



So, we have in other words that $q^2$is $2a^2$. And if $q^2$is $2a^2$, then by the same argument as before $q^2$is also even, and so q must be even. And therefore, q can be written as the form of 2 times some other number b. So, we have that p is of the form 2 times a and q was of the form 2 times b. But what this means is that the gcd of p and q must be at least 2, because both of them are even numbers. So, they are both multiples of 2. So, we claimed initially that the gcd of p and q is 1. We said that they were actually both in reduced form. So, there was no common factor other than 1. And now we have shown that if we assume that we in fact generate 2 as a common factor. So, this cannot be the case. So, the only contradiction that we can resolve with this is by assuming that $\frac{p}{q}$could not have been there. So, therefore, √❑ cannot be represented by any reduced fraction $\frac{p}{q}$.

(Refer Slide Time: 05:32)



So, this argument of Hippasus is a common way of arguing things in mathematics right. To show that some fact capital P holds you first assume that not P holds, it is negation holds. So, we wanted to show that there is no way that √❑cannot be expressed as rational. So, we said let us assume the negation. Let us assume that √❑can in fact be express as a rational, and then you take that assumption and derive a contradiction. And since you cannot accept a contradiction, your assumption must be wrong and therefore, what you tried to prove originally was correct.

So, in fact, it is not just √❑that is irrational, √❑is also irrational. Now, 4 is a perfect square. So, we know that √❑ is 2. What about√❑; that is also irrational. So, among the integers among the natural numbers we have the perfect squares 1, 4, 9, 16, 25 and so on which consists of 1^2, 2^2, 3^2, 4^2, $5^2$and so on. So, a perfect square is one whose square root is also a natural number.

Now, it turns out that anything which is not a perfect square has an irrational square root, and the proof is not exactly the same because we have used a property of 2, and evenness in this proof, but with a very similar argument you can show this is the case. So, therefore, there are a lot of irrational numbers that you can generate just by taking square roots of non-perfect squares.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute

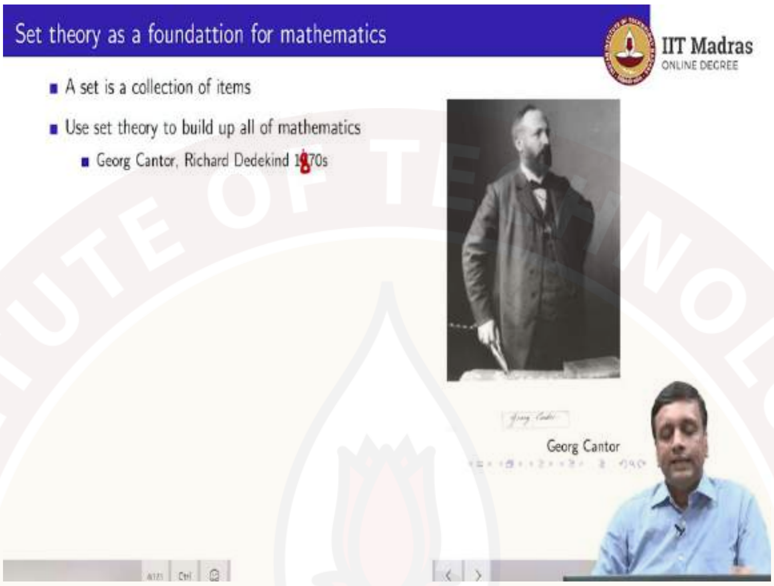
Week - 01
Lecture – 10
Set versus Collections

(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)

So, we have looked at sets, and we said that a set loosely speaking is a collection of items. And then we made some remarks in that lecture that not everything can be thought of with a set. So, let us ask whether every collection is in fact a set, and if not, why not?

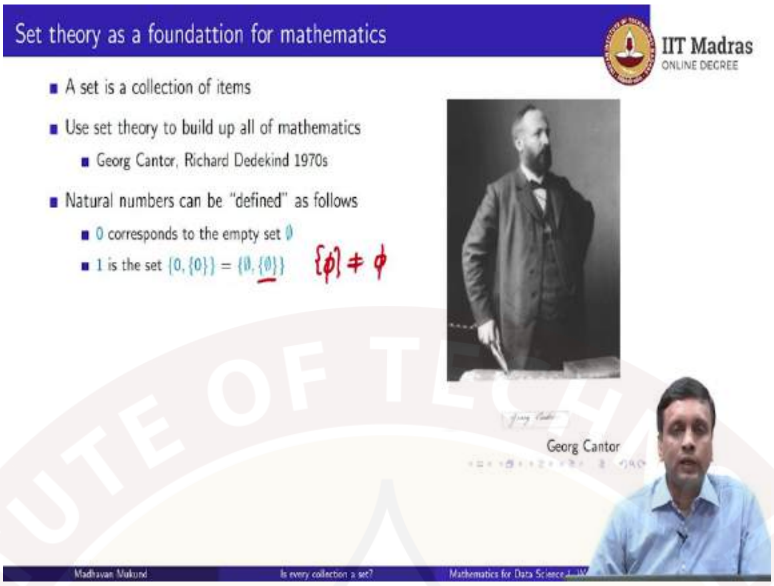
(Refer Slide Time: 00:29)



So, as we said a set is a collection of items. And when set theory was investigated formally starting from the late 1800s, the idea was to make set theory a foundation of mathematics. So, let us try to briefly understand what that means. So, we wanted to the mathematicians of the time wanted to start off with very basic things and build up all of mathematics from that, and they felt that set theory was a good place to start.

So, some of the mathematicians who are involved in this was Georg Cantor and Richard Dedekind from the 1870. So, this is a mistake, this is not the 1970s of course, but the 1870s.

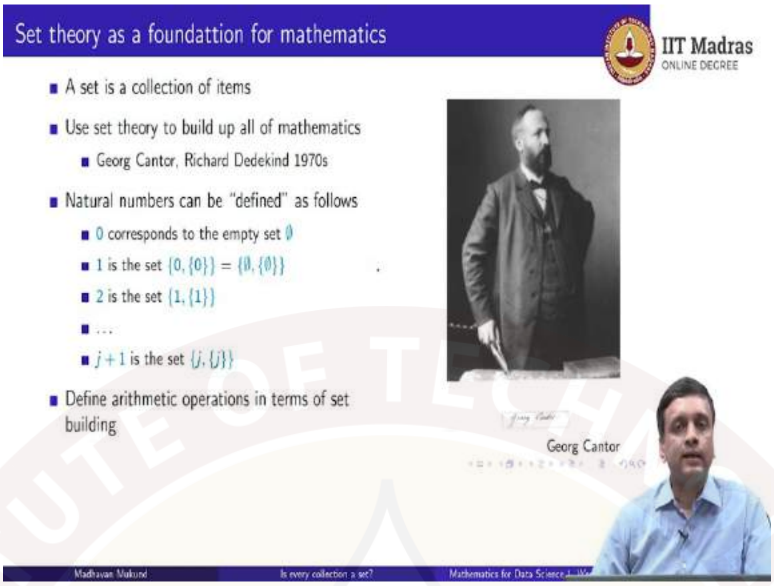
(Refer Slide Time: 01:05)



So, one is aspect of this foundational nature of set theory is for insists how do you generate numbers if you have only sets. So, one of the things that you need if you start with set theory is the empty set. So, you have it for free. So, what they said is that 0 can be thought of as the empty set.

So, we are going to use sets to represent numbers, and we are going to use the empty set to stand for 0. So, what is 1? Well, 1 is a set that consists of 0 and the set containing 0; in other words it is a set containing the empty set, and the set containing empty set. So, remember that the set containing empty set this is not the same as this right. The empty set has no elements; the set containing the empty set has one element.

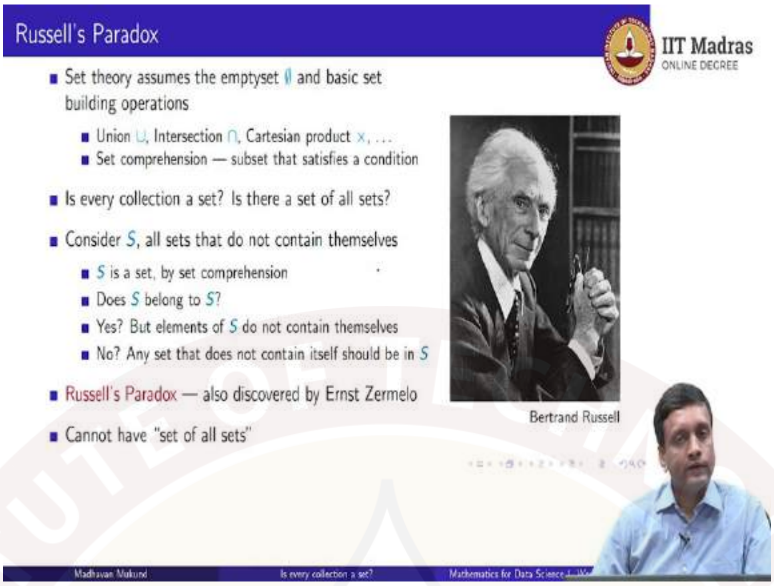
(Refer Slide Time: 01:52)



Similarly, 2 would be the set which contains 1 in the representation above, and the set containing 1. So, it is a bit tedious to write out. So, I have not expanded it. But you just take the expression for one in terms of the empty set replace it twice, and you get the number 2. And in this way for any number j plus 1, you can get it from the number j by taking the representation of j adding the set containing the representative j putting it into a new set.

So, these are the natural numbers as expressed using sets starting from the empty set. And then you can actually define set theoretic ways of combining these two define, the addition of two numbers and this format to get a new number which is the sum and the product and so on. So, this is what it means to use something like set theory as a foundation of mathematics.

(Refer Slide Time: 02:37)



So, basically set theory assumes that you have the empty set, and then you have basic set building operations. For instance, you can take the union of sets, you can take the intersection of sets, you can take the Cartesian product which we saw when we were looking at relations. And you can of course to set comprehension which is that you can take some elements from a set which satisfy a condition and build a subset.

So, now into this picture came Bertrand Russell and he asked whether this would make sense or not. So, here we come back to our fundamental question is every collection a set? In particular he asked can there be a set of all sets? So, remember that sets are objects just like anything else. So, we can collect them together. So, is this collection of all sets in fact a set?

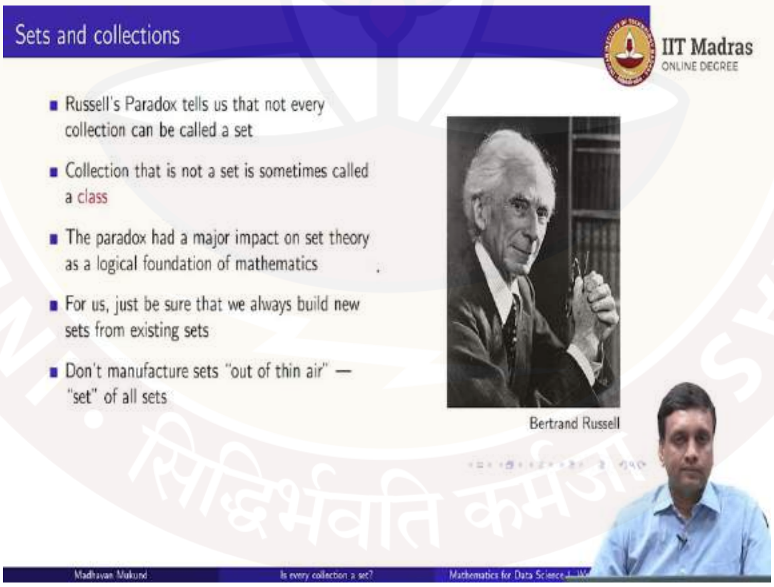
Well, supposing it is a set, then we can do the following. We can apply set comprehension right, and we can pick out some sets from this collection of all sets. So, we will call capital S, the subset of all sets that do not contain themselves. So, this is a subset of this hypothetical set of all sets. So, this capital S is a set because we have applied set comprehension to the set of all sets. So, we have the set of all sets. And among all sets we have pulled out those sets which do not contain themselves. So, this is the condition we have applied, and this is allowed by set comprehension.

Now, the question is does the set that we have constructed belong to itself, does S belong to S? Well, if it does belong to itself, then it does not satisfy its own definition because elements of S should not contain themselves. So, S cannot belong to itself, because if it did it would contradict to way we have pulled out S from the set of all sets. But if it does not belong to itself, then that is also a contradiction, because then S does not belong to S and by the condition that we have applied to pull out sets S must be included in that condition.

So, either way we have a paradox; we have a contradiction. So, S can neither belong to itself nor can it not belong to itself. And this is called Russell's Paradox. He was the first person who published this and made it publicly known, but this was also independently discovered by another well known set theorist of the time called Ernst Zermelo. So, what this really tells us? If you remember our argument is that we made some assumption, and then from that assumption we realized that we have a contradiction or an observed situation.

So, something must be wrong in one of our assumptions. And here it turns out that the assumption that goes wrong in all these is the assumption that there is a set of all sets. If we did not have a set of all sets, we could not have done the set comprehension, and therefore, we would not have reached this observed conclusion.

(Refer Slide Time: 05:23)



So, what Russell's Paradox really tells us is that, not every collection can be called as set in particular the set of all sets does not exist. So, he went through an exercise of trying to formulate a different version of set theory which he called type theory and so on, but in modern mathematics typically if you are not sure that what you are dealing with is a set then it is safer to just called such a collection a class. So, a class is just a collection of objects which does not have any of the implied properties that you expect from the sets.

So, this paradox as we said came in the context of set theory being used as a foundation of mathematics. And, this seem to casts doubts on whether it could be used at all. So, it had a major impact on this whole mathematical exercise of deriving mathematics from logical foundations which went on into the 20th century which we will not be able to discuss here unfortunately, but it is a fascinating subject in its own right.

For us what we have to be clear about is that whenever we use sets we must make sure that we always start with sets that we have and build new sets from existing sets. So, we can assume that the numbers are sets. So, we have the set of natural numbers, the set of integers, the set of rationals, the set of reals and so on. And, whenever we construct a new set we just have to verify that the set that we started with to construct the new set was already a set.

So, we take a Cartesian product or a union or set comprehension, we always start with old sets and make new sets. So, those old sets must be well-defined. So, in other words, we should not manufacture sets out of thin air such as the set of all sets.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Madhavan Mukund
Department of computer science
Chennai Mathematical Institute

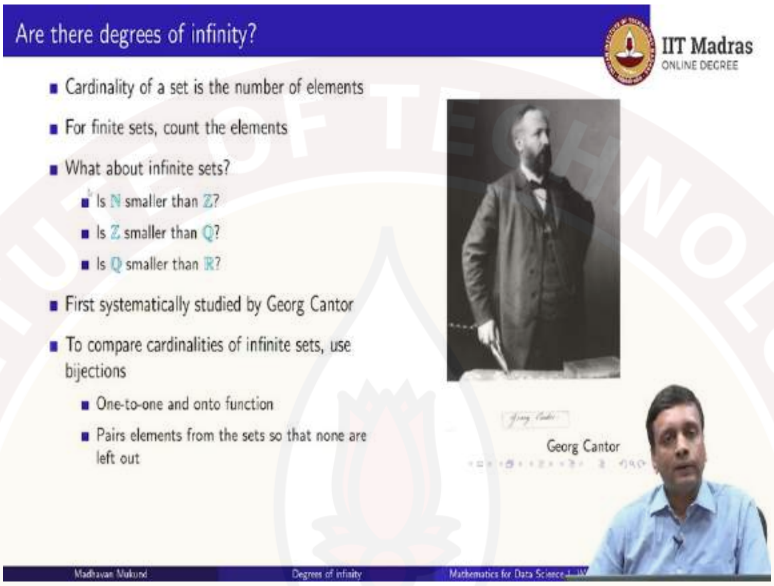
Week - 01
Lecture – 11
Degrees of infinity

(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)

So, when we looked at the sets of numbers, we said that we have various kinds of infinite sets – the natural numbers, integers, reals, the rationals, some of them are discrete, some of them are dense. And the question that we asked was whether they all have the same size, or there are more of one than another?

(Refer Slide Time: 00:29)



So, the question that we want to ask is, are there degrees of infinity? So, we know that for a set the cardinality denotes a number of elements, and if it is a finite set we just have to count these elements. So, for a finite set, there is no problem about cardinality which is the count the number of elements and we are done.

We get a natural number which is the cardinality of the set. Now, the question is what do we do for infinite sets. So, let us look at the natural numbers for instance. So, in which we move from the natural number to the integers, we added negative numbers. So, clearly we have added an infinite set of numbers we roughly doubled the set. So, is the set of natural numbers are same as the integer number in size or not?

Similarly, when we move from the integers to the rationals, we move from a discrete set where we had a next and previous element to a dense set where between any two element there is an another element. So, this suggests that there should be more rational than reals rationals than integers, but is that true or not?

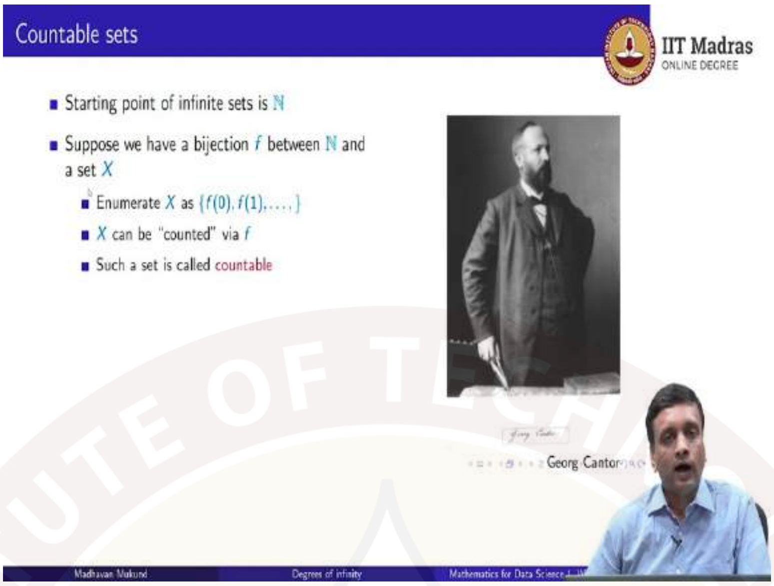
And finally, when we move from rationals to real numbers we added a whole bunch of irrational numbers which cannot be expressed in the form $\frac{p}{q}$. So, clearly the real numbers have a large number of new things which are not in the rationals. So, again is the set of reals larger than the set of rationals or not? So, this study of the cardinality of infinite sets was actually undertaken by Georg Cantor in the 1870s. And as we have seen when we studied functions the correct way to compare the cardinality of infinite sets is to use a bijection.

So, what is the bijection? The bijection is one-to-one and an onto function. In other words, it allows us to map one set to another set in such a way that two elements are always map to two different elements and everything on the other side is map 2 from something here that is the onto part. So, it is one-to-one no onto elements map to the same one, and it is onto no element on the right hand side is missed out.

So, intuitively what this allows us do through this function this bijection is to pair up the elements from the one side with the elements from the another side. So, I take an element on the left hand side, through the bijection I pair up it with an element of right hand side. And because it is one-to-one and onto, this pair this paring actually exhaustibly covers all the elements in both sides or nothing is left out.

So, we have paired up everything and therefore, the two sides have the same cardinality. So, this is the technique that we will investigate in order to resolve these questions about the cardinalities of the infinite sets of numbers that we have discussed above.

(Refer Slide Time: 02:55)

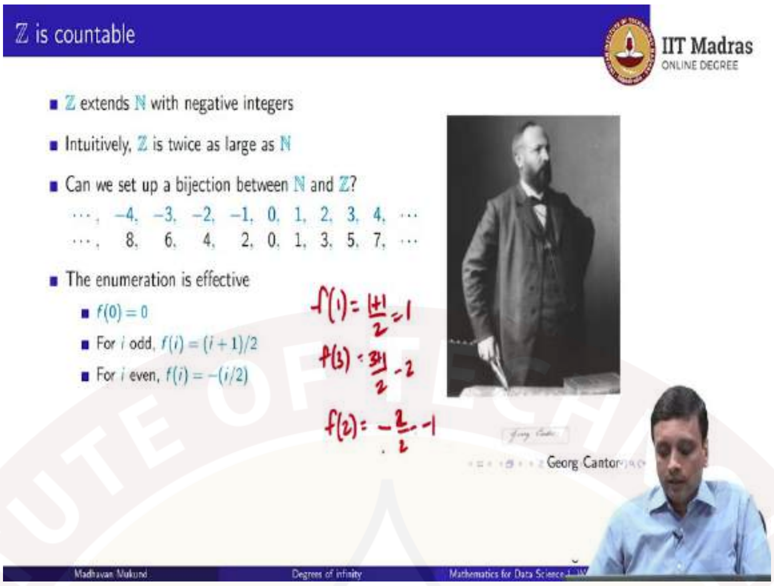


So, our starting point is the set of natural number, because this is the first infinite set that we have to begin with. When we start counting we realized that there is no largest number because we can always add 1. And so if we take all the finite numbers that we can used to count, we get an infinite set called the natural numbers.

Now, supposing we find a bijection between the set of natural numbers and some other set X, does not matter what this set is, but supposing there is a bijection. We can pair of the natural numbers with the elements of X. This means that we can actually effectively enumerate the elements of X, we can take the number paired with 0, f (0) and call that the beginning of X, then f (1) is an X element, f (2) and so on.

And because we are doing this kind of enumerating X, we can count X in a way via f and so we call any such set countable. So, countable set is one which can be bijectively paired up with the set of natural numbers. So, when we are looking at other sets, we will first check whether they are countable or if not we have to argue that they cannot be counted.

(Refer Slide Time: 04:02)



So, let us begin with set of integers and show that it is countable. So, why should be, why should it not be countable, or why should it be a surprise if it is countable? Well, because Z extends N with negative integers right. So, for every, if you do not count 0 in the calculation, for every positive natural number there is a corresponding negative integer in Z.

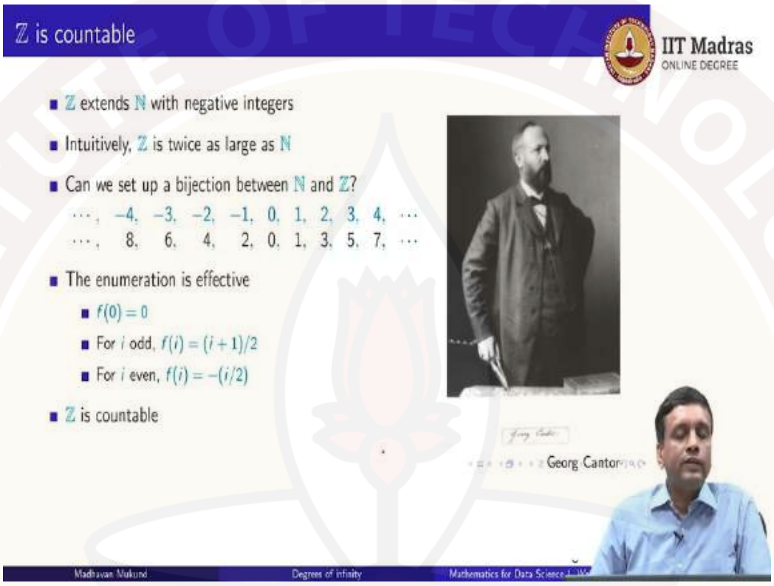
So, Z is referring twice as big as N; for +1 you have -1; +2 you have -2 and so on right. So, it seems contradictory that you can double the set, and still have the set of the same size that you started with. So, the question now is for Z to be countable, can we set up a bijection between the natural numbers and Z?

So, let us look at Z as we do on the number line. So, it starts from some -∞ and then it comes to -4, -3, -2, -1, 0, 1, 2, 3, 4 and continues. So, we start our enumeration at 0. So, we enumerate 0, the 0 of Z as the 0th element, then we map 1 to +1, map 2 to -1. What do we do next? Well, we map 3 to +2, and 4 to -2.

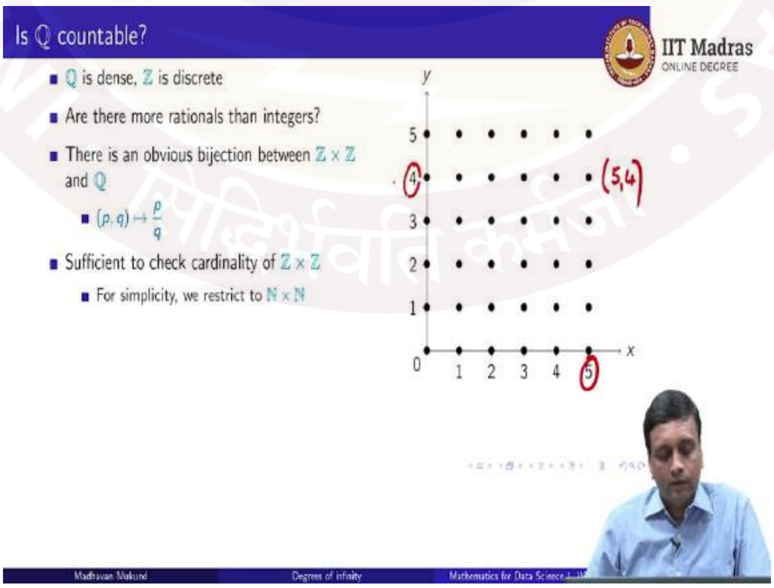
So, we keep zigzagging to the right hand to the left, we count Z by starting with the center moving right one, moving left one, moving right one, moving left one. So, in this way we could now enumerate the number +3 as 5, -3 as 6, +4 as 7, -4 as 8. So, in this way we can actually enumerate Z effectively. So, f (0) is 0 as we saw. If i is odd for example, 1 then f (i) is $\frac{i+1}{2}$.

So, f(1) for instance is (1 + 1)/ 2 = 1; f(3) = (3 + 1) / 2 = 2 and so on. So, if f is odd, I have $\frac{i+1}{2}$. And if it is even like 2, then I take - $\frac{i}{2}$. So, I take -2/2 which is -1. If it is 4, I get -4/2 which is -2 right. So, we have actually given an effective way of assigning a position in some sense or count to every number in Z, and this shows that the set of integers is actually countable.

(Refer Slide Time: 06:17)



(Refer Slide Time: 06:20)



Now, what about the rationals? One reason why we might suspect that the rationals are not countable is because the rationals we saw a dense between any two rational numbers there is an another rational number.

Whether the integers and the rational numbers are discrete, you can always find a next number; and in the case of integers you can always find a previous number. For natural numbers 0 has no previous number, every other number has a previous and a next. So, given that rationals are dense and the integers are discrete, the question is are there more rationals than there are integers?

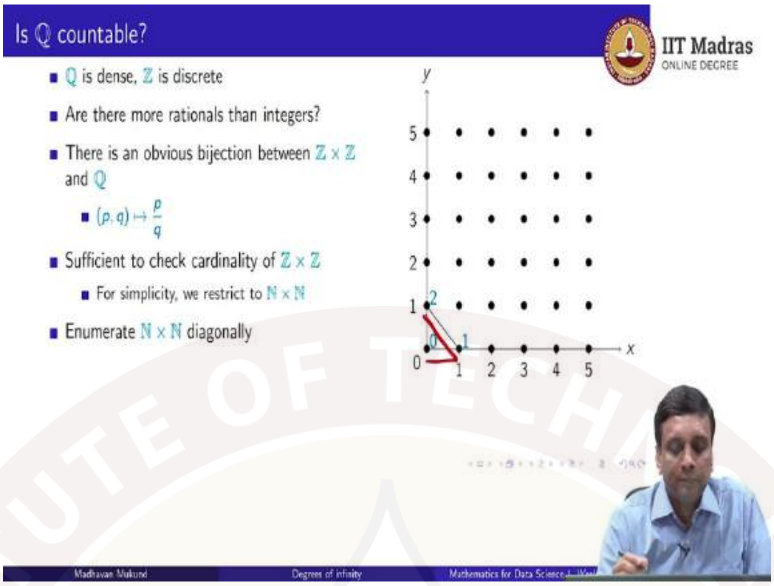
Now, there is an obvious bijection between pairs of integers and rationals because that is what a rational is, rational is a pair of integers p upon q. So, I can take a pair (p , q) and Z cross Z and directly connect it in an bijective way to the fraction $\frac{p}{q}$. So, every pair gives a unique rational number, every rational number gives me a unique pair.

There is no surprise here, there are no we are not talking about reduce forms of for example, we have different numbers like $\frac{1}{10}$, then we have $\frac{2}{20}$, and $\frac{3}{30}$, these are all different rational numbers they may represent the same value, but they represent different pairs. So, this is a clear bijection between Z cross Z and Q. So, Z cross Z has the same size of Q.

So, if we are looking at the cardinality of Q, we can also look at the cardinality of Z cross Z. Because if we can measure the cardinality the size of Z cross Z, then through this bijection, Q must have the same size, there is no need to separately measure the size of Q.

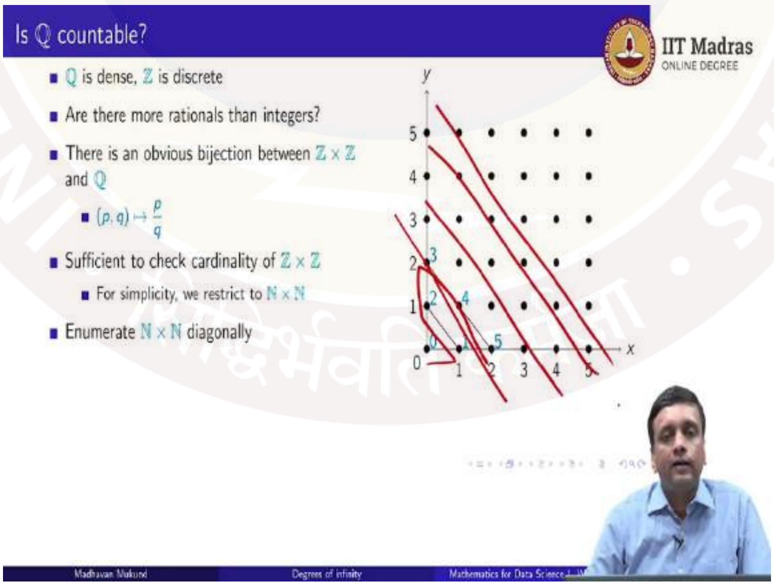
So, instead of Z cross Z just to make the picture easier to see, we will actually do N cross N, and then I will show you how to extend it to Z cross Z. So, here is a picture of N cross N. So, remember that we think of N cross N in a two-dimensional grid and at each point (i, j) I have a dot representing the pair (i, j). So, for instance this pair, this pair is (5 , 4), because it comes from the 5 and the 4 over here right. So, every dot in this pair in this grid is a pair in N cross N.

(Refer Slide Time: 08:22)



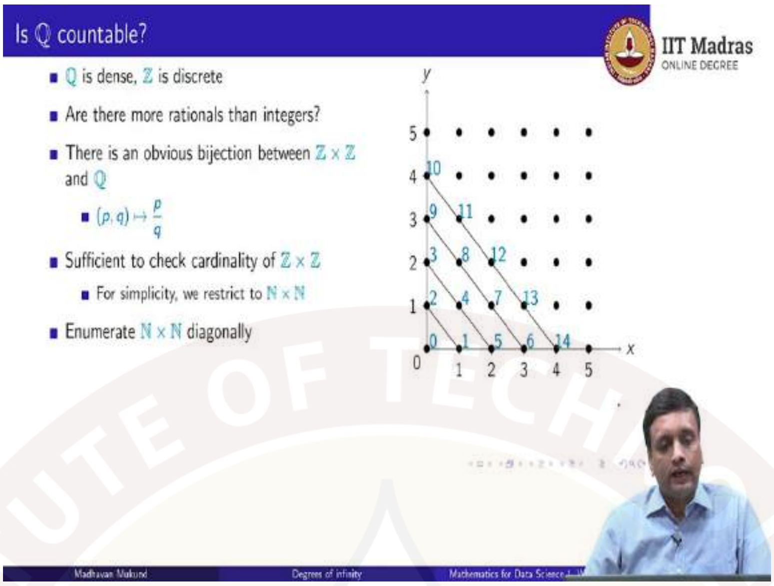
Now, I am going to enumerate this in a particular way. So, here is a one enumeration. So, you start with the 0th element as the element at the bottom left corner what is normally called the origin. Then you enumerate the first diagonal right, so you go from here and then you go right and then you go up. So, you enumerate in this way then you continue.

(Refer Slide Time: 08:47)



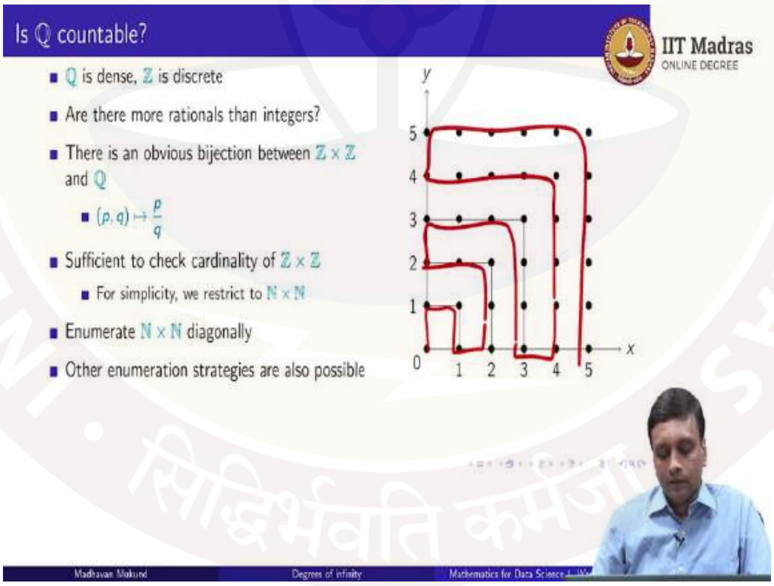
So, you started from here went up, then you up there, and come back down again right. So, you can slice this thing like this right. So, you can slice this gird like this, and enumerate it diagonal by diagonal.

(Refer Slide Time: 08:58)



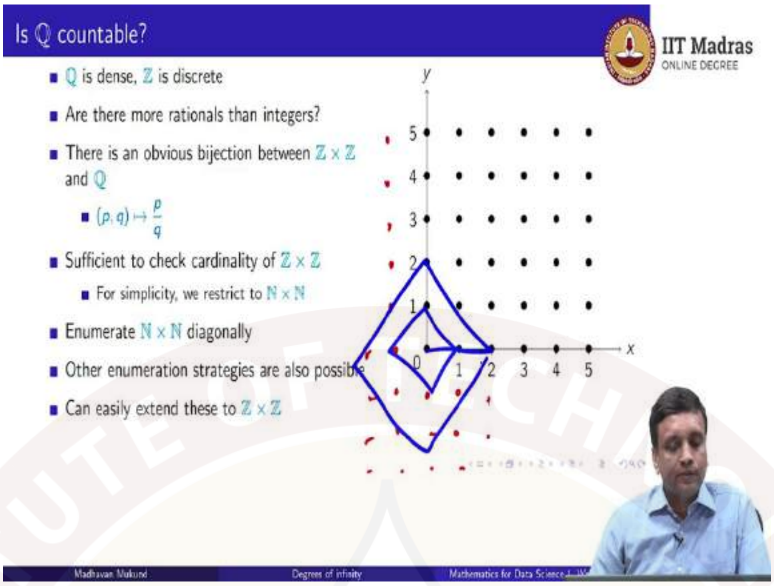
So, this gives us an effective enumeration of N cross N.

(Refer Slide Time: 09:04)



But we can also enumerate in different ways. For instance, we can enumerate in these larger and larger squares. So, we can start here, then finish this, then do this, then do this, then do this and so on right. So, long if we do not miss out any point in the grid we are done.

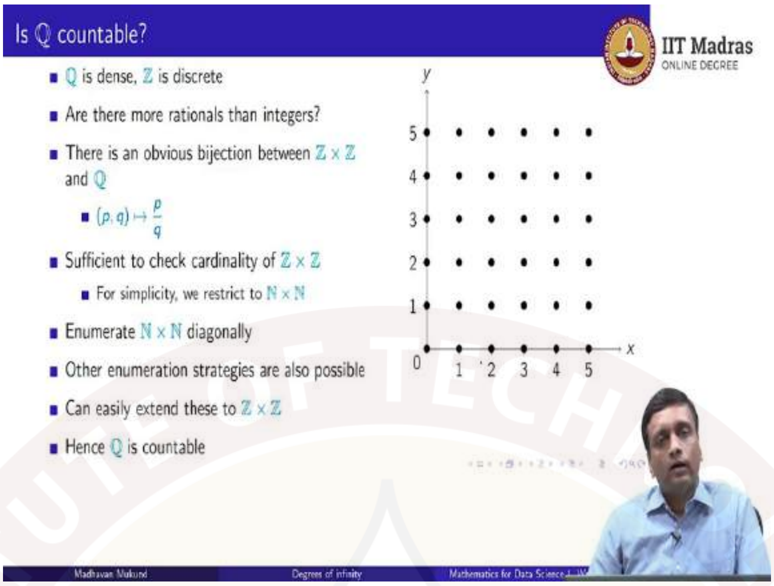
(Refer Slide Time: 09:21)



So, this shows us that N cross N is something that we can enumerate. Now, how would we do it for Z cross Z? Well, it is very simple. If I had Z cross Z, I would also have points on this side, and I would also have points below right. So, I would have points to the left and below 0 because I would had negative numbers.

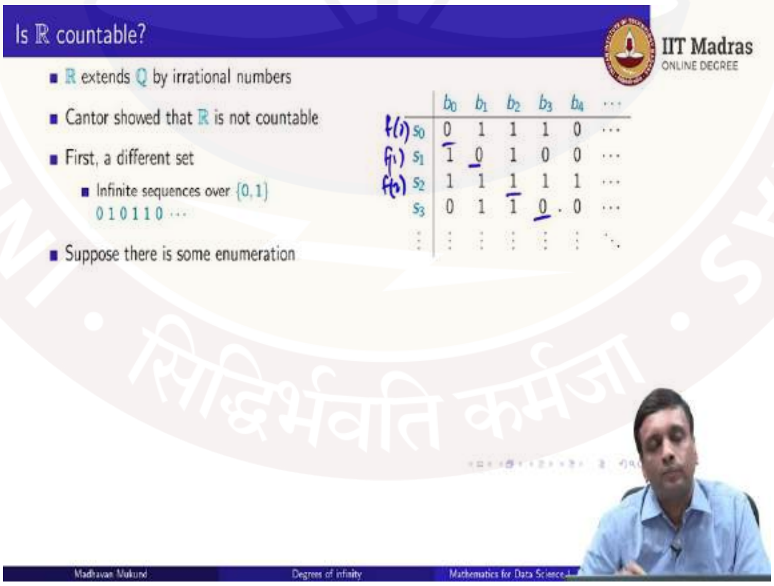
So, now, if I wanted to enumerate Z X Z, I would start here, then I would do this, and I would complete this diamond, then I would go here, and then go here, and then complete this diamond and so on right. So, instead of doing just the diagonal, I would extend the diagonal around to form a diamond, and in this way I would start from the center and spiral out so that I enumerate all the numbers in Z X Z. So, N X N can be enumerated as we saw, and this can be easily extended to Z X Z.

(Refer Slide Time: 10:10)



So, therefore, the set of rational numbers though it is dense and then it looks superficially to be much larger than the set of integers, actually both the integers and the rational numbers have the same number of elements which is quite surprising, but it is true.

(Refer Slide Time: 10:26)



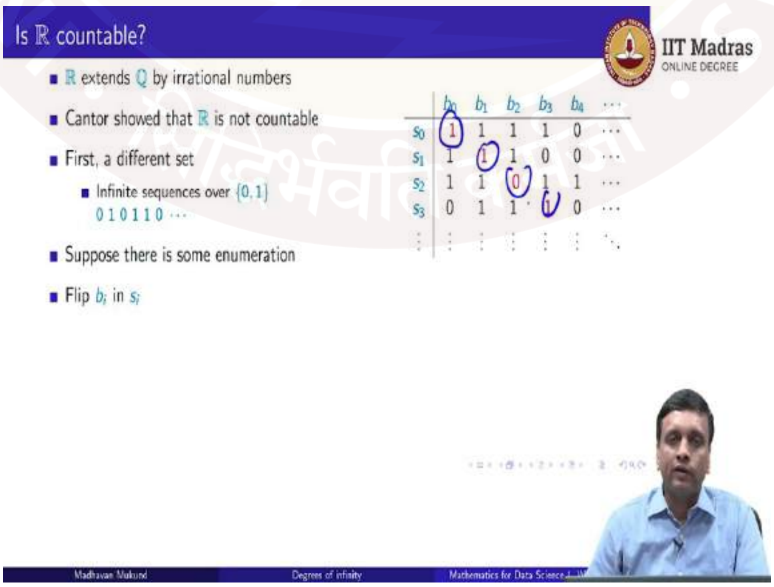
So, for all the infinite sets we have seen are countable right. Of course, the natural numbers are countable by definition, and then we saw integer are also countable, and the rational are also countable. So, what about the real numbers? So, how did we get to the real numbers? We took the rationals and then we added all these irrational numbers like √2, π, e and so on. So, Cantor showed that R actually is not countable. So, let us see how this proof works.

So, actually he did not, he did have a separate proof that R is not countable, but later on he made another proof which is easier to present which starts with the different set. So, instead of looking at *R*, we will look at something which looks quite different. We will look at infinite sequences over 0, 1. So, an infinite sequence of a 0, 1 is just something like you just keep writing down 0 or 1 infinitely many times without stopping right.

So, what Cantor argued is that this set is not something that you can count. So, supposing you can enumerate the infinite sequences over 0, 1, then on the right to see some enumeration; we are not looking at a particular enumeration in some particular order. We are just saying is there any enumeration at all, so that I can write down the 0-th sequence. So, this is the 0th sequences, this is f (0) in some sense, this is f (1), this f (2) and so on.

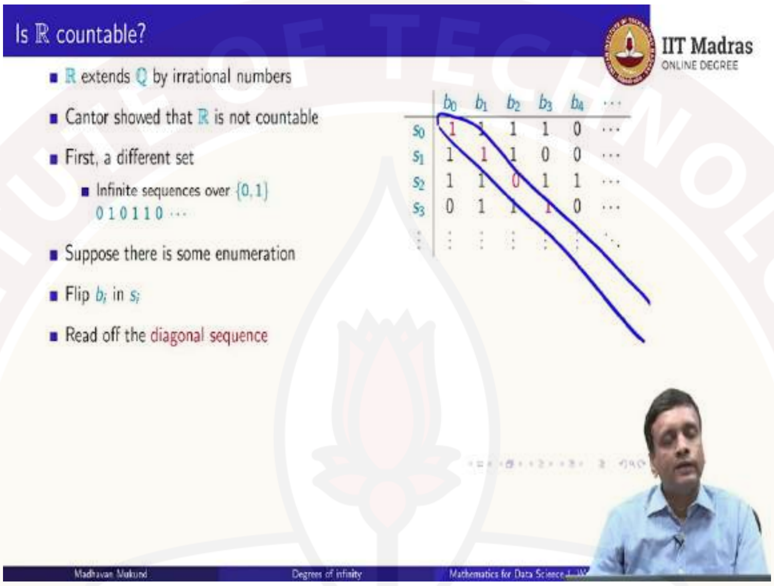
So, I have just written f (0) as s_0, and f (1) as s_1, and so on. And each sequence has positions which I have written b for bits because these are binary digits 0 or 1. So, each sequence has an infinite sequence of bits which characterize what it is, and no 2 rows are the same they are all different infinite sequences of 0s and 1s right. So, hypothetically this table is an enumeration of such sequences. So, if this is a enumeration of all such sequences, can be derive a contradiction? So, this is how Cantor derived a contradiction.

(Refer Slide Time: 12:26)



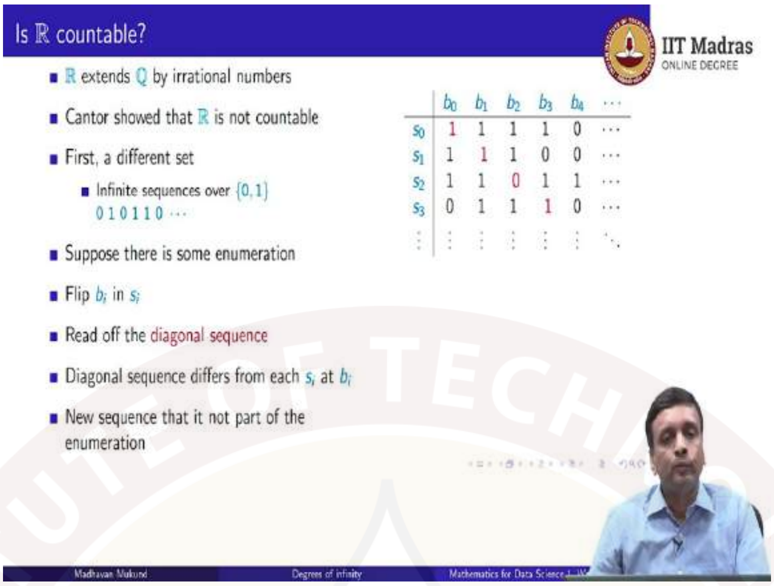
So, he said let us take each row and reverse the bit. And which bit to be reversed? Well, if we are in the ith row, then we reverse the ith bit. So, in the first row which is s 0, we reverse b_0 , in the second row. So, if you want go back, so this was 0, so we are here at 0 0 1 0. So, after flipping, it becomes 1 1 0 1 right. So, what we are doing is in 0th row, we are flipping b_0; in row s_1 we are flipping b_1; in s_2 we are flipping b_2, and so on.

(Refer Slide Time: 13:03)



So, now this gives us a new sequence which we can read off diagonally right. The sequence consists of the red numbers which we have got by flipping the number at the i-th position in the i-th sequence. What can be say about this sequence? Well, first of all it is an infinite 0, 1 sequence.

(Refer Slide Time: 13:20)



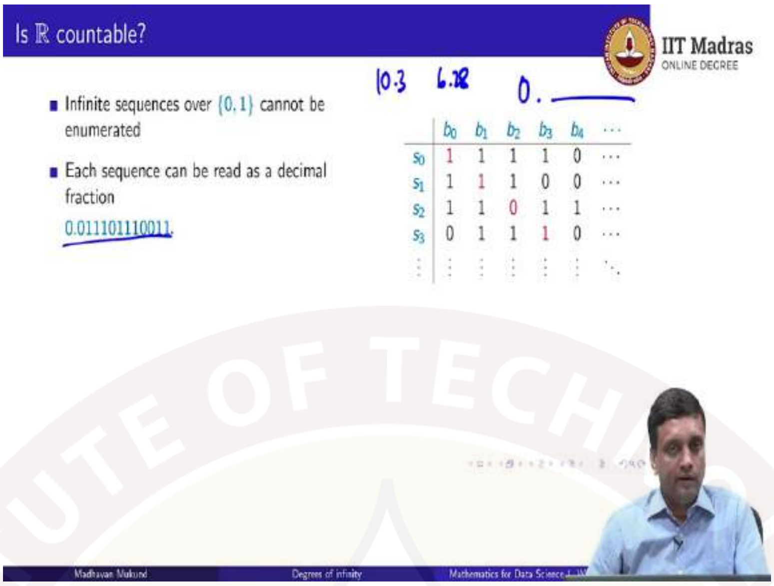
But this infinite 0, 1 sequence cannot be any of the rows in my table, because by construction if it is a row in my table it must be s_j for some j, but at position j, s_j has been flipped. So, this cannot be s_j because if I had s_j already in my table if the sequence is already in my table, the new sequence has the j-th bit flipped. So, diagonal sequence differs from each s_i at b_i, and therefore, this new sequence that I have constructed cannot be part of the enumeration.

Now, it is important that we are shown this regardless of what the enumeration looks like, we have not made any assumption about the order in which we are enumerating. We have said no matter what sequence you have in mind in terms of enumeration, you would have to be able to write down the sequences one after the other table in a sequence of rows.

However you write it down, I will be able to construct this new diagonal sequence by taking the i-th bit in the i-th row and flipping it. So, however you enumerate it, I get a new sequence which is not part of your enumeration. Therefore, there is no possible way of enumerating 0, 1 sequences.

So, as we said this is not the question we asked, the question we asked is are the real numbers enumerable, are real numbers countable? And what we have actually argued is that 0, 1 sequences, infinite 0, 1 sequences are not countable. So, from here how do we get to the real number?

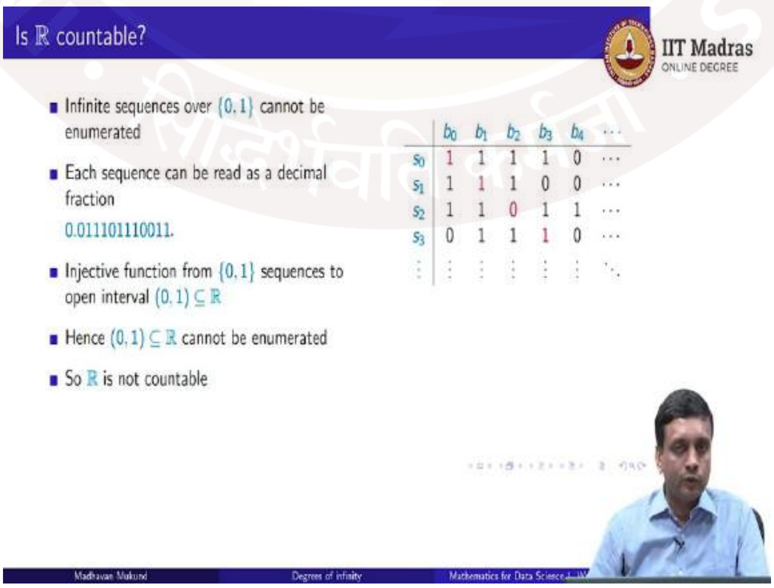
(Refer Slide Time: 14:38)



Well, it is one way to do this is to just think of these 0, 1 sequences as actually decimal fraction. Now, we know that we can write things like 10.3 and 6.28 and so on. So, now, we just restrict our self to writing in decimal fractions of the form 0 point something where everything on the right hand side of the decimal point is either a 0 or 1.

So, here is an example right. So, this is an example of a 0, 1 sequence represented as a decimal fraction. So, since each sequence is different, each such decimal fraction represents a different number.

(Refer Slide Time: 15:20)

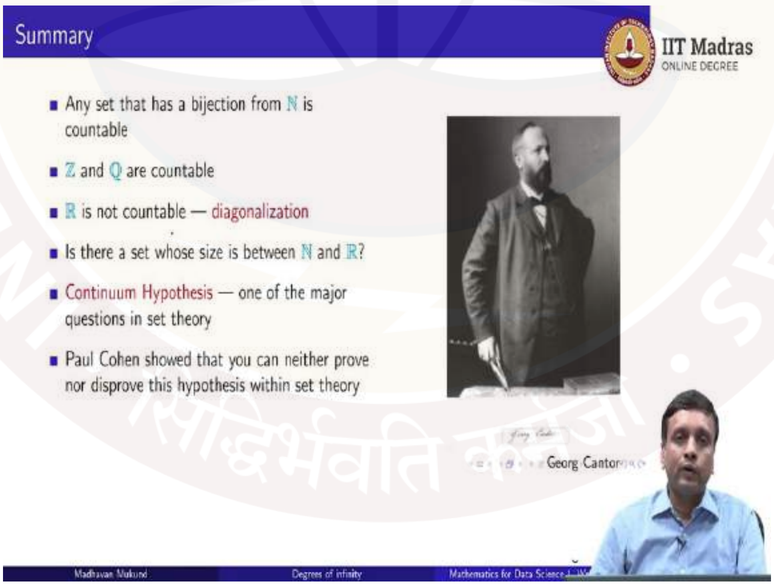


And these are all real numbers between 0 and 1, because they all have an integer part which is 0, and then we have something which is of course, we could have exactly 0 if we have all 0s ok. So, we definitely do not have, all we do not cannot get to 1, but we can think of these as numbers between 0 and 1.

So, each such sequence represents a different point in the interval 0 to 1. So, this is an injective function right. So, this is an injection that is a one-to-one function from infinite sequences 0, 1 to the interval (0, 1). Now, the interval (0, 1) is a very small fraction of the reals.

So, what this argument tells us is that in fact even this very small fraction of the reals is not countable because the set of underlying 0, 1 sequences not countable. So, if this even this small fraction of the reals is cannot be enumerated, then R itself cannot be countable right. So, this is an indirect argument saying that not saying that R itself is not countable directly, but saying that there is a small part of R, which is not countable. And since R is much more than that, if the small part cannot be counted we have no hope of counting the whole thing.

(Refer Slide Time: 16:26)



So, to summarize any set that has a bijection from N is what we call a countable set. And we showed that the set of integers in the set of rationals are countable by describing a strategy to enumerate the sets. Now, this argument is due to Cantor which builds this diagonal sequence called diagonalization and has been used in many other proofs involving infinity after that. So, the proof of diagonalization by Cantor shows that the set of real numbers is not countable.

So, notice that the set of real numbers is not countable and the set of rationals is countable. What it does to the rationals to create the real numbers? We added the irrational numbers. So, actually the set of irrational numbers that we have added to the rationals must be itself uncountable, because we cannot take two countable sets and add them up and get an uncountable set. So, in other words, there are vastly more irrational numbers than there are rational numbers that is what it tells us.

Now, one question that we could ask is, is there anything in between? So, these are sets that we have been using intuitively. So, we have counted them. But can we construct something for instance which is not countable, but which is smaller than the reals right? So, is there such an infinite set?

Now, it turns out that this is a very non-trivial question. This question was actually posed when Cantor came up with this proof in the late 1800s, and it remained a very central opened question it was called the continuum hypothesis.

So, if you look at cardinal numbers in the finite sense, we have 1, 2, 3, 4, 5. So, we have a kind of small jumps between them, but we have a continuous sequence of numbers. Now, we seem to have this big jump between the real number the integers of the natural number and the real numbers, is there something in between or is it so, is there a continuum of these infinite numbers or these big jumps?

And this continuum hypothesis was a very important open question in set theory. And in the 1960s Paul Cohen actually showed that this is a question which cannot be proved or disproved. So, this is what is called independent. So, this is a fact which is independent of set theory using the axiom of the set theory, no way that you can either prove or disprove it.

So, both the fact that there is a such a set, and there is not such a set are consistent. So, these infinite sets lead to a lot of interesting questions, some of them are quite mindboggling, and they are quite counterintuitive. But if you are interested in these things, it is well-worth looking into them.

IIT Madras

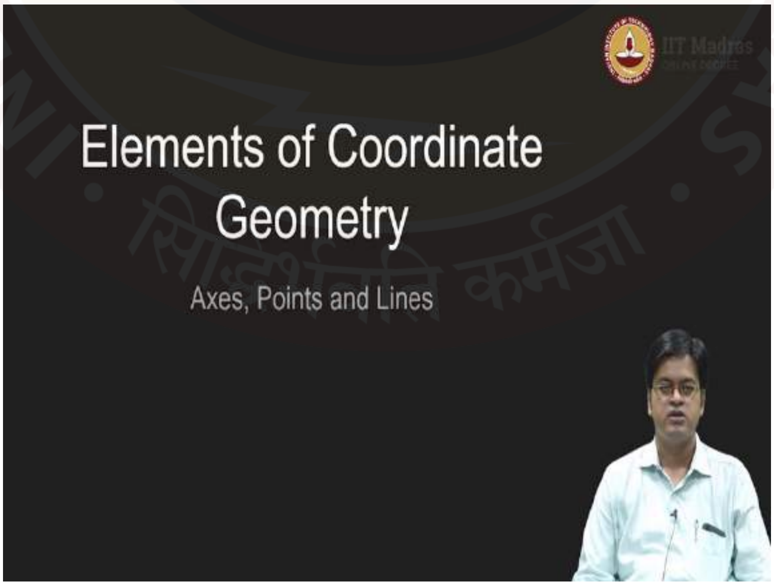
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 12
Rectangular Coordinate system

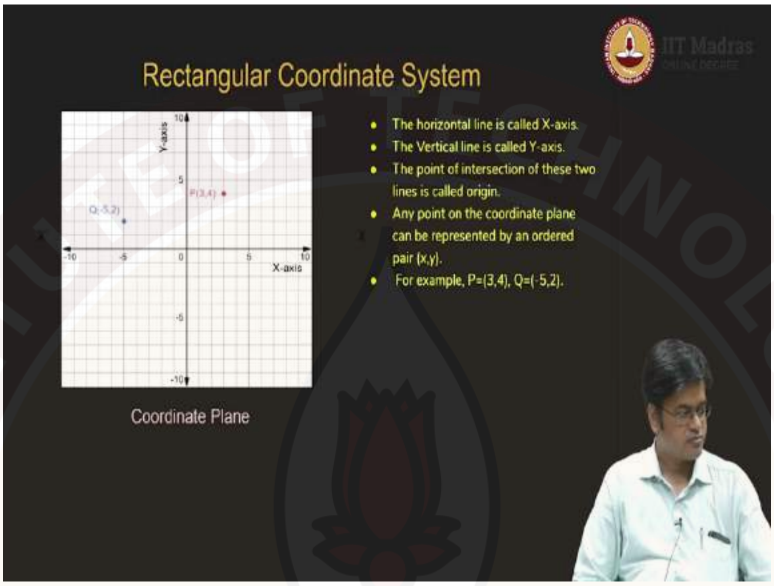
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, hello students, today we are going to see some elements of coordinate geometry. Now, let us try to identify these elements as axes, points and lines. We have already seen in basic geometry what are points, lines and planes. So, we will further study this and we will study some algebraic properties using coordinate geometry of these particular geometric objects.

(Refer Slide Time: 00:40)



So, in that context first we need to revise our Rectangular Coordinate System; why is rectangular coordinate system important and how we can study. Given a point on a plane; given a point on a plane you want to describe how this plane be how this point behaves or what is the location of this point. Now, if I want to consider this point and I want to describe the position of the point as of now I cannot say anything more than, this point is slightly towards the right top of the plane.

Now, if I introduce a horizontal line over here, then I can say the point is in the upper half of the plane. This gives a slightly better visibility to the point or slightly better description of a point. Now, if I consider a real number system associated with this line then I can say the point lies in 0 to 5, if I plot two perpendicular lines between 0 to 5 then I will get this point.

This is much better. Now, these perpendicular lines can also be replaced with one perpendicular line which is this which has a real number system associated with it. Now, when a real number system is associated with this point, then what you can actually see is if I can consider this, this particular structure or this particular square which is enclosed within 5 on the vertical line and 5 on the horizontal line; I am giving a much better description of a point.

Then I can enhance this further by putting up the grid lines. These grid lines now typically in this case locate the exact location of the point. So, what is the exact location of the point over here? If you look at this exact location of the point is on the horizontal line if you travel 3 units in one direction, horizontal direction and 4 units in the vertical direction then you will reach this point.

So, I can also name this point as in the horizontal direction I have to travel 3 units and in the vertical direction I have to travel 4 units. So, I can name this point as 3 comma 4 that will be a precise description of this point. So, in turn what we have seen just now is a reference system through which we are able to specify the location of a point in a specific manner. Let us analyze this reference system that we have introduced.

Now, in horizontal direction I have to travel 3 units and in vertical direction I have to travel 4 units; that means, I am actually specifying the coordinates in X direction and coordinates in vertical direction. So, in particular these horizontal directions and vertical directions are called X axis and Y axis respectively.

So, if you look at this horizontal direction, you can see the vertical line cuts the horizontal line into two parts; positive part of X axis and negative part of X axis. Similarly, the vertical line is cut by the horizontal line into two parts. On the upper side we have a positive part of Y axis and on the lower side we have a negative part of Y axis.

So, this is a typical structure which is called coordinate plane ok. Now, let us come to the nomenclature of this particular coordinate plane. As I mentioned if I am travelling 3 units in horizontal direction; I will call that as X coordinate and if I am travelling 4 units in vertical direction, I will call that as Y coordinate. Hence, the name coordinates.

These two lines X axis and Y axis meet each other at a 90 degrees angle; that means, both the lines are perpendicular to each other. Therefore, the name rectangular; recta means right in Latin so, rectangular means 90 degrees coordinate system; that means, a rectangular coordinate system. So, let us revise what we have studied just now in words.

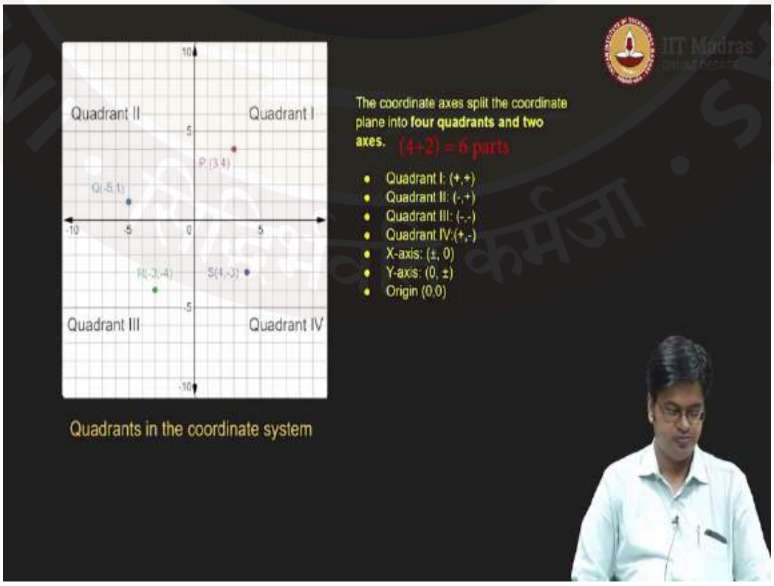
The horizontal line is called X axis, it allows you to move from left to right. The vertical line is called Y axis which allows the movement up and down, then there comes a point of intersection of these two axes which is called origin. The point of intersection of these two axes is called origin and if you look at the coordinates of these, then any point on this particular plane can be denoted by a ordered pair (x, y).

You can see one blue point is also popping up now. Now, how to describe a point using a coordinate plane? So, for example, given a point (3, 4) how will I locate this point? So, if you look at this (3, 4), we have already seen how to locate it. We have travelled 3 units in horizontal direction and 4 units in vertical direction therefore, (3, 4).

Now, suppose you are given another point which is (-5, 2), then this x coordinate corresponding x coordinate is negative; that means, I have to go to the left of the vertical line. That means, I have to travel here a 5 units distance which is - 5 and on the positive side of Y axis I have to travel that is up upper up upper half divided by X axis I have to travel 2 units which will give me the point (-5 ,2).

So, this is how we can uniquely describe points using coordinate plane. Now, when I was when we were studying these two points (3, 4) and (-5, 2), you can easily see with respect to this coordinate axes you can have 4 parts of the coordinate plane.

(Refer Slide Time: 07:24)



Let us study those parts in detail in the next slide. So, next slide is this coordinate plane. Now, I have identified 4 points in all 4 parts of the coordinate plane. So, you can see the first point P which lies in the positive side of X axis and positive side of Y axis has positive x and y coordinates which is given by quadrant I. So, any point in this plane, in this particular quarter will have positive X and positive Y axis.

Now, in general as a mathematical psychology we move in a anti-clockwise direction. So, now, I can move in a anti-clockwise direction to the next one fourth part, next quarter of the coordinate plane. And, see that my X axis has negative values and my Y axis has positive values. All points which have this form of values are called points on the second quadrant or the quadrant the one the quarter of this particular coordinate plane is called quadrant II.

Next we come in a anti-clockwise direction to the third side that is this. So, if you look at the point R which is lies in this particular quadrant is (-3, -4); that means, the x value is negative and the y value is negative. Therefore, (-3,-4) is a point which lies in quadrant III.

Remember it is easy to remember this that quadrant I and quadrant II, quadrant III that is odd quadrants have same parity of x and y coordinates. And, quadrant II and quadrant IV have opposite parity of x and y coordinates. So, let us go to quadrant IV, you can see a point S lies in quadrant IV which has coordinates 4 and -3. Now, this 4 and -3 which denotes x coordinate is positive and y coordinate is negative such a classification comes in quadrant IV.

So, this is how a coordinate plane is come split into four quadrants. Now, a question may arise in your mind; suppose I have this point which is (5, 0). Now, in which quadrant this point lie? The answer is this point does not lie in any of the quadrants. This point lies on the X axis. Similar question can be asked for a point (0, 5). The point does not lie on any quadrant, but lies on the Y axis.

So, based on this particular understanding, a coordinate plane its subdivided into first is four quadrants, two are axes. Let us try to see what are the typical features of the quadrants and these axes. Quadrant I, you will have x and y coordinates which are positive. Quadrant II, you will have x coordinate which is negative y coordinate which is positive. Quadrant III, you will have both negative values. Remember odd quadrants will have same parity that is quadrant I is positive, quadrant III is negative.

Now, quadrant IV will have positive and negative, x coordinate which is positive, y coordinate which is negative. Then comes the split into axes. So, on the X axis you will have points which can either take positive values or negative values for x coordinates and 0 for y coordinate. On the Y axis you will have points which can take positive and negative values for y coordinates, but 0 for x coordinate. Now, there remains only one point which is the point of intersection which is identified as origin ok.

So, this completes our understanding of the coordinate system. Why quadrants, quadrant system is helpful? Sometimes you have been given several points to plot. Now, those points if you look at them closely, you need not have to divide the system in a equally distance manner, like this manner. You may have many points in quadrant I, in that case you can scale this, you can bring this to the bottom right corner, bottom left corner and just focus on quadrant I.

So, if you have a good understanding of quadrants, you may be able to graph the functions better, graph the points better; that is why the coordinate system is important. This ends our discussion on coordinate system.

IIT Madras

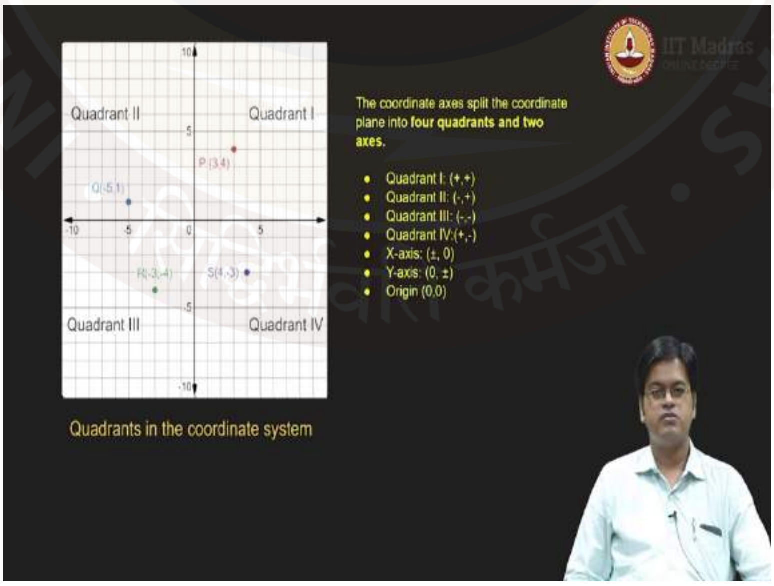
ONLINE DEGREE

Mathematics for Data Science-1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 13
Distance formula

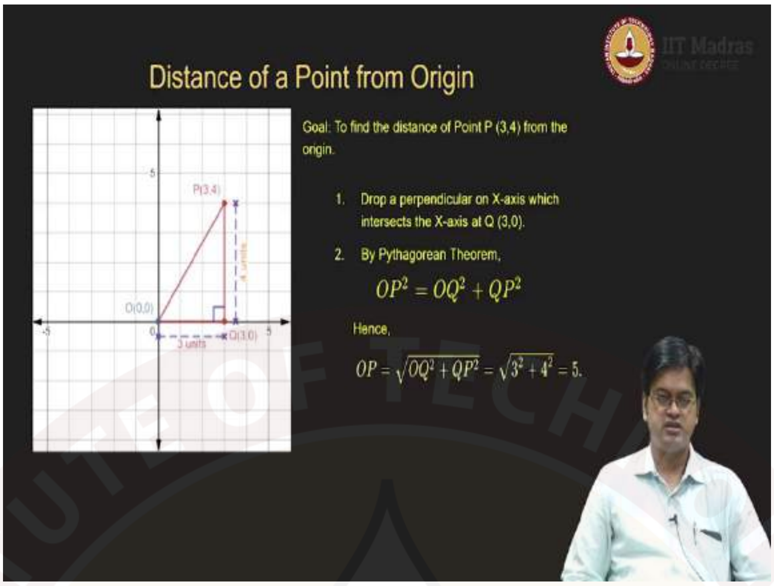
(Refer Slide Time: 00:05)

(Refer Slide Time: 00:14)



So, after coordinate system, let us try to identify one classical problem.

(Refer Slide Time: 00:20)



That is if I have a point and somebody ask me a point is located here; let us say point is (3,4). And somebody ask me what is the distance of this point from the origin? So, in this particular slide our goal is to find the distance of a point P which is (3,4) from the origin.

That essentially reduces to finding the length of this line segment which is joining points O and P. So, is there any classical tool that is of my help? Suppose, now if this point is either lying on X axis or Y axis, let us say if this point is say (3,0) ok. If this is the point that is of interest to me; do I know how to find the distance of this point? The answer is yes I know, I just need to calculate the units that are in horizontal direction.

Suppose the point is on Y axis, then do I know how to calculate the distance of this particular point from Y axis? The answer is again yes I know, I just need to calculate the number of units that I need to travel to reach this point. So, if the point lies on X axis and Y axis, I know how to calculate the distance of a point. Now, if the point is lying anywhere in the coordinate plane, how to find a distance is a question.

For that, let us try to understand the situation, that if I know if somehow I can understand this with respect to this coordinate axis. This particular position with respect to these coordinate axis then I will be able to give the answer to find the distance between the two points. So, let us try to do one thing that is let us try to get the image of this point (3, 4) on X axis.

So, how will I get the image of this point (3,4) onto the X axis? The easiest way is you drop a perpendicular on X axis, that intersects the X axis at point (3,0). Once this is done then you can actually drop a perpendicular and see that it forms a right angled triangle with X axis in place and a vertical line in place; you have a right angled triangle. Do you know any theorem in our conventional geometry that relates this particular structure?

You know Pythagoras Theorem or Pythagorean Theorem, that relates this particular structure. In a right angled triangle the hypotenuse length of the hypotenuse is given by square root of its adjacent sides; square root of squares of the lengths of the adjacent sides. So, we will try to use this for finding the distance of a point from the origin. So, by the Pythagorean Theorem, I know $OP^2$is actually equal to OQ^2+QP^2.

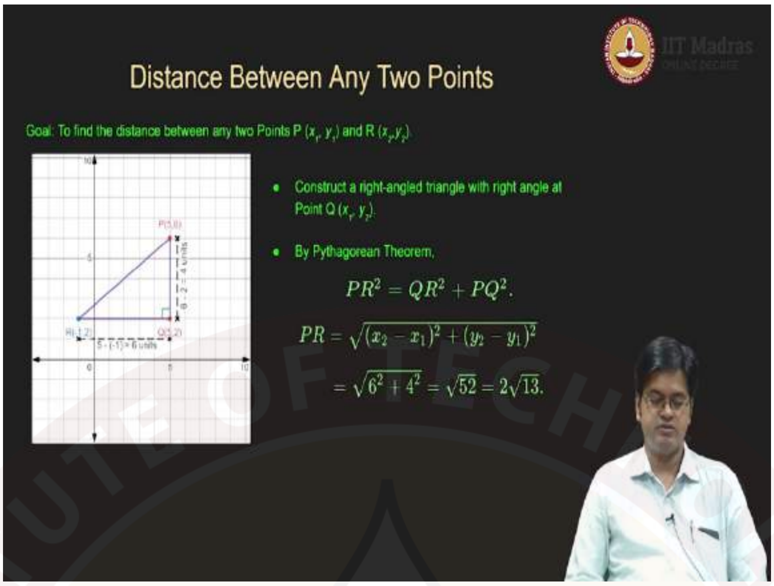
Now, the exercise that we did orally just before starting this problem will help us to understand what is OQ^2. So, what is OQ? OQ is a part of X axis, OQ is a line segment which is a part of X axis. What is the length of OQ? We have already discussed that, that length is 3 units. Similarly, if you look at QP; what is QP? QP is parallel to Y axis. So, it is as good as projection of Y axis projection onto Y axis.

So, what is the length of this particular line segment which is QP? That is 4 units; so I know the length of OQ and I know the length of QP. Therefore, by Pythagorean Theorem, I know the length of OP. So, what will be the length of OP? It will be $\sqrt{\square}$. So, $3^2$is 9, $4^2$is 16 therefore, this will give me 25; 16+9 and positive square root of it will give me number 5.

Now, has it anything special to do with point (3,4) or can I generalize this? The answer is yes, it has nothing special to do with point (3,4). I could have started with point P which is (x , y) and then projected this onto X axis or I figured out the image onto X axis which will be (x , 0). And therefore, the length of OQ will be x and the length of QP will be from 0 to y units; that means, y units.

So, length of QP will be y units and therefore, the formula OP = $\sqrt{\square}$ would have been possible. So, let us try to take this particular example and try to generalize this problem to finding the distance between any two points.

(Refer Slide Time: 06:02)



So, distance between any two points. So, again the setup is pretty common. Our goal is to find the distance between any two points P $(x¿¿1, y_1)¿$ and R $(x¿¿2, y_2)¿$. How will you find the distance between any two points? Let us see the points on the graph, then the things will be more specific. My $(x¿¿1, y_1)¿$ is (5, 6) and $(x¿¿2, y_2)¿$ is (-1 , 2).

Now, if I look at these two points, I want to find the distance between these two points. So, once easy way to find a distance between these two points is to construct a right angle triangle. But, now because the point is not located on X axis, this (-1 , 2) is not located on any of the axis; I cannot say drop a perpendicular to X axis.

So, the actual way that I should do here is I will drop a perpendicular to X axis which will intersect at (5 , 0). And, then to this line I will drop a perpendicular from the point R minus (1,2) and which will intersect this, this particular line which will be the perpendicular to X axis, at where the y coordinate will be 2 and x coordinate will be 5. So, this point will be (5,2) and then I will get a right angled triangle.

By skipping these steps, we can straight away say that you construct a right angled triangle with a right angle at point Q which is $(x¿¿1, y_2)¿$. Just relate this$(x¿¿1, y_2)¿$, if you use this terminology is the point (5,2). So, you can draw a right angle triangle using (5,2) ok.

So, this way we need not have to specify steps that you have to draw two perpendiculars and all; because the point may as well lie in the third quadrant. And, in that case dropping perpendicular to X axis may not help, you have to extend the perpendicular beyond X axis. So, it is always better to consider this kind of structure, that is construct a right angled triangle with right angle at point Q which is $(x¿¿1, y_2)¿$.

Then it does not matter where the point actually lies. Now, once the right angle triangle is in place, the same theory that we used Pythagorean Theorem will come into play. And, by Pythagorean Theorem if I want to find the length of PR, I know $PR^2 = QR^2 + PQ^2$. Can I calculate the length of QR and PQ? The answer is yes I can calculate, because, the line segment QR is actually parallel to X axis and the line segment PQ is parallel to Y axis.

Therefore, this is as good as computing the length on X axis and this is as good as computing the length on Y axis; hence what we will get is. So, how to compute the length? It is basically the change in x coordinates. So, how far the x coordinates have changed? So, while computing the length parallel to X axis always remember you should go from left to right, that is when you are subtracting you should take the highest value first that is 5 – (-1).

So, the length of this will be 6 units and while subtracting or while finding the length in a vertical direction go from bottom to up. That means, you subtract the value that is highest in Y direction to the value that is lowest in Y direction. So, here 6 - 2 will give me 4 and in the X direction 5 – (-1) will give me 6 units. So, this is how we will calculate the length of these two line segments.

And therefore, I can easily find the length of PR; while calculating the length because we are in this particular case, we are considering squares. It does not matter whether you consider x_1 first or $x_2$first because, anyway we are squaring even if you get the negative value, you will be squaring it.

So, in particular in this case where the coordinates are $(x¿¿1, y_1)¿$and $(x¿¿2, y_2)¿$, I will take $(x¿¿2 - x_1)^2¿$; does not matter which one is bigger. And $(y¿¿2 - y_1)^2¿$, does not matter which one is bigger.

And I will take a positive square root of it. Therefore, my length PR for this particular example will be 6^2+ 4^2; $6^2 = ¿$ 36, $4^2 = ¿$ 16 together they will give 52 is $2\sqrt{❑}$. So now, we have established a general formula which is called distance formula for finding the distance between any two points on a coordinate plane.

IIT Madras

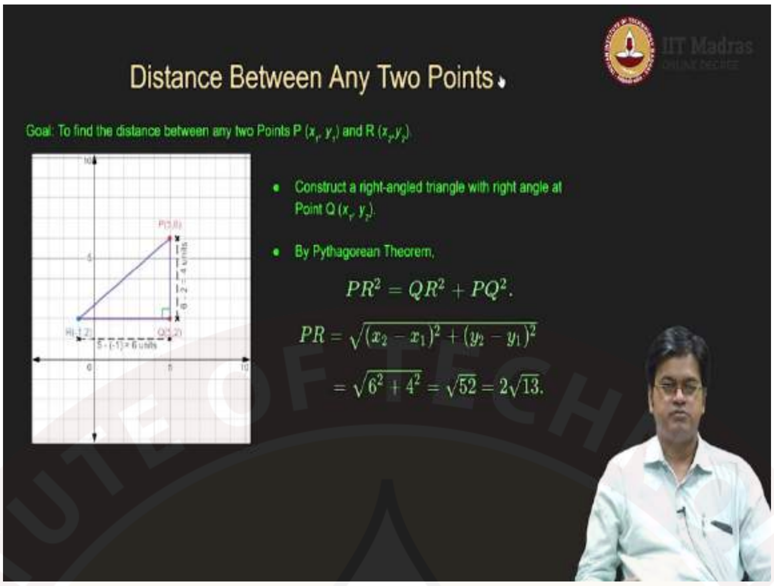
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 14
Section formula

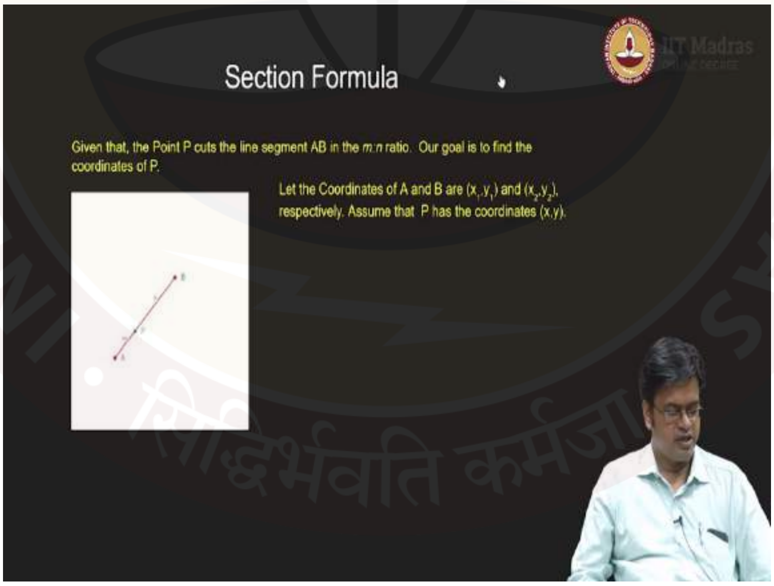
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



Now, let us take up the next concept. Now, we have handled two points; now let us take three points.

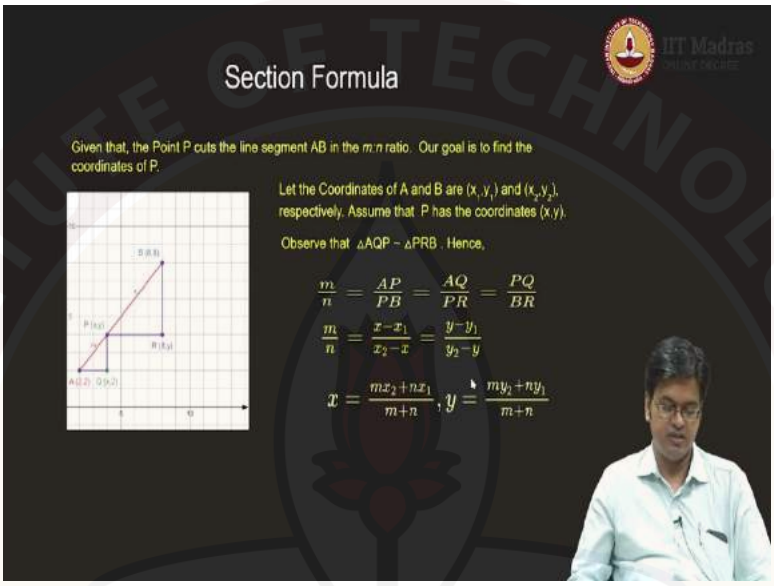
(Refer Slide Time: 00:22)



And, let us say those three points lie on a line and given that the point P cuts the line segment AB in the ratio m : n. Our goal is to know the coordinates of point P; this will give us the Section Formula. So, this is the graphical representation of the points. So, there are two, there is a line segment AB and point P cuts this line segment in the ratio m : n.

How will you find the coordinates of point P? This is the question; let us bring in our coordinate system. So, let the coordinates of A and B are ¿, y_1¿ and ¿, y_2¿, the coordinates of A are ¿, y_1¿ coordinates of B are¿, y_2¿,. I do not know what P is, let us assume it has some coordinates which are x and y ok. So, let us bring in them in the coordinate system which is this.

(Refer Slide Time: 01:29)



Let us try to understand this particular coordinate system by putting up some triangles around. So, what I have done is I have actually constructed two triangles using the same logic that I used in the distance formula. If the coordinates of point P are (x,y), then I will construct a right angled triangle in this direction; where the x coordinate will be x and y coordinate will be the coordinate of y coordinate of A that is y_1.

Similarly, I will do the same thing with respect to point B; I will drop a perpendicular which will meet at this particular point. So, basically I will drop a perpendicular which will meet the X axis and again I will draw a perpendicular here. But, let us for sake of simplicity we have constructed a right angle triangle, where the y coordinate of this point will be y and the x coordinate of this point will be the x coordinate of point B which will be x_2, (x_2, y).

With this understanding we can proceed further and see that the triangles, these two triangles are similar to each other. How? First of all let us see this line is parallel to X axis and this line is parallel to X axis as well. Therefore, these two are parallel lines and this is a transversal that is passing through these two parallel lines. Therefore, these two angles the angle A and angle P will be same or equal.

Next these two are right angles, then we know the sum of the angles in a triangle is 180 degrees, therefore this angle, angle B must be equal to angle P. Therefore, triangle AQP must be similar to triangle PRB by angle test that essentially means I have their sides in some ratio, correct. So, for simplicity I have plotted these points with some coordinate references.

So, this is A is (2,2), B is (8,8); then whatever I mentioned the coordinates of Q are (x,2) and coordinates of R are (8,y). So, now these two things will be in some ratio that is $\frac{AP}{PB}$, these are the hypotenuse of these two right angle triangles is equal to $\frac{AQ}{PR}$and this thing is $\frac{QP}{RB}$ right or you can see $\frac{AP}{PB}$ is equal to $\frac{AQ}{PR}$which is equal to $\frac{PQ}{BR}$.

Now, I already know $\frac{AP}{PB}$ have a ratio m : n. So, their ratio is m by n that is already known to us, that is given to me. Now, can I calculate the length of AQ and PR? The answer is yes, because AQ is parallel to x axis. It is just subtracting the highest x coordinate from the low.

So, it will be x -2 in the figure and in our theory it is x - x_1. Similarly, you can look at PR; it will be 8 - x or in our theory it will be x_2- x. For y axis or the lines that are parallel to y axis PQ and BR you can see you will go to the highest value that is y - 2 or y - $y_1$and the other one BR will have $y_2 - y$right.

So, together I will have a representation of this form: $\frac{m}{n}=\frac{x-x_1}{x_2-x}=\frac{y-y_1}{y_2-y}$. Now, take one equality at a time; that means, $\frac{m}{n}$is equal to let us say these the consider these x coordinates. So, we will just cross multiply them, rearrange them you will get what x is equal to.

In a similar manner just take $\frac{m}{n}$ is equal to this, these y coordinates ratio and then cross multiply and rearrange them. You will get the following values which are given by $x=\frac{m x_2+n x_1}{m+n}$. And, similarly $y=\frac{m y_2+n y_1}{m+n}$. This gives me the section formula, when a point divides the line in the ratio m : n.

Another interesting question is suppose I know the coordinates of x and y; can I find in what ratio the line divides? Obviously, yes because you know the coordinates of the line, you just need to use this formula for finding the ratio ok; that will be more clear when you solve more problems ok.

IIT Madras

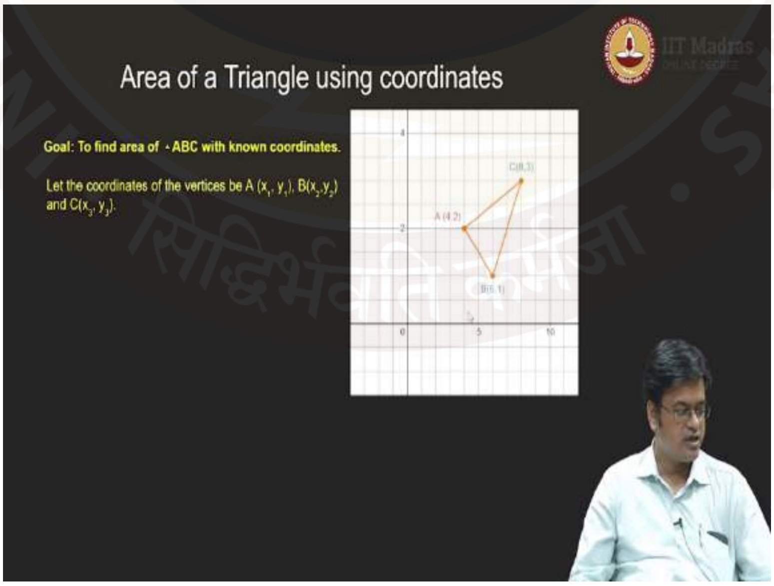
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 15
Area of triangle

(Refer Slide Time: 00:05)

(Refer Slide Time: 00:15)

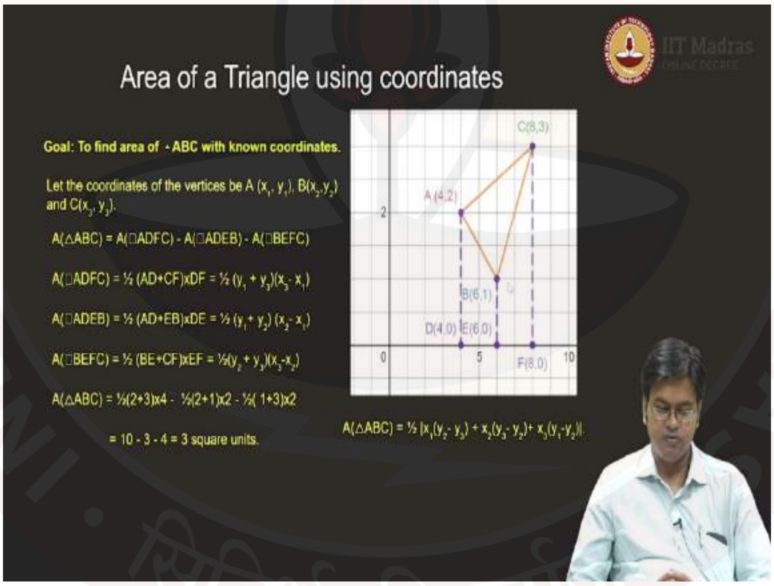


After section formula let us try to understand the three points when they are not on one line that is when they form a triangle. So, you have been given three points, and you know they are not collinear points and therefore, they will form a triangle. And the question can be how to find the area of triangle using the coordinate system.

So, let us try to see that using the coordinate system. So, there is some triangle ABC and I want to find the area of triangle ABC. Let the coordinates of that triangle be (x_1, y_1),(x_2, y_2) and (x_3, y_3). Once I have these coordinates, I can plot it here. You can see on the right there is an image of a triangle.

Now, how to find the area of this triangle? Now, whatever I discussed so far everything actually relied on dropping a perpendicular to X - axis and finding the area of the geometric object that is formed. In earlier cases, it was just a triangle. Now, if we follow that theory then you can easily see that I need to do something like dropping a perpendicular to x axis. So, I have dropped perpendiculars to X - axis.

(Refer Slide Time: 01:35)



Now, I have generated some figures. What are the figures that I have generated? In particular, I have generated 3 trapeziums, trapezium ADFC that is the biggest one encompassing everything. Then, you can look at trapezium ADEB, then you can look at the trapezium BEFC.

Now, my triangle is trapped in between these trapeziums. So, let us try to make our understanding crystal clear. If I want to find the area of triangle ABC, then I need to first consider the biggest possible quadrilateral or trapezium that is ADFC and eliminate the areas of two smaller trapeziums that is ADEB and BEFC. And whatever I am left with is the area of triangle ABC.

Now, do I know how to find the area of trapezium? Yes, I know. The formula is half times sum of parallel sides into the height of the trapezium. So, we need to quantify how will these quantities be calculated? Let us consider trapezium ADFC, if I consider a trapezium the ADFC then what are the parallel sides of this trapezium? Side AD and side FC.

So, I will take average of these two parallel sides that is half of AD plus FC. Then, what is a height? Height should have a perpendicular distance, so that is X - axis. So, I know the distance will be DF.

So, let us take the general coordinate system rather than using this coordinate system. What are the coordinates of A and D? So, A has coordinates $(x_1, y_1)$and after dropping a perpendicular on X – axis the y coordinate will vanish and therefore, the coordinate of D will be $(x_1, 0)$. So, what will be the length of AD? It will be purely in terms of y that is y_1.

Similar, thing is applicable for CF. So, it will be nothing but y_3; so, area of ADFC,

$$Area(ADFC) = \frac{1}{2}(AD + FC) \times DF = \frac{1}{2}(y_1 + y_3) \times DF.$$

Now, what is the length of line segment DF or FD? Highest minus the lowest. So, in this case our F is $(8, 0)$ or $(x_3, 0)$and the point D is $(x_1, 0)$. So, it is $(x_3 - x_1)$.

$$Area(ADFC) = \frac{1}{2}(AD + FC) \times DF = \frac{1}{2}(y_1 + y_3) \times (x_3 - x_1).$$

In a similar manner, I can actually see a smaller trapezium that is ADEB, smaller quadrilateral that is ADEB and the height of that quadrilateral will be the length of ED which is 2 in this case or $x_2 - x_1$in the coordinate system. So, this is what our understanding of the length is. In a similar manner, the sum of lengths of parallel sides is $y_1 + y_2$.

$$Area(ADEB) = \frac{1}{2}(AD + EB) \times DE = \frac{1}{2}(y_1 + y_2) \times (x_2 - x_1).$$

In a similar manner you can compute BEFC.

$$Area(BEFC)=\frac{1}{2}(BE+CF)\times EF=\frac{1}{2}(y_2+y_3)\times(x_3-x_2)\cdot$$

Now, using this you can compute the area of the triangle which can be easily seen to be in this form. So, I have just taken this example and computed these values. So, the values are effectively in this particular case the length of AD was 2.

The length of CF was 3 so

$$Area(abc)=\frac{1}{2}(2+3)\times 4-\frac{1}{2}(2+1)\times 2-\frac{1}{2}(1+3)\times 2=10-3-4=3\,square\,units.$$

Now, if you look at this particular thing and rewrite this expression you will get a very nice expression. You can juggle with this expression and try to simplify it by taking a cross products and you will come up with the expression of this form.

$$A(\Delta ABC)=\frac{1}{2}\vee x_1(y_2-y_3)+x_2(y_3-y_1)+x_3(y_1-y_2)\vee ¿$$

The absolute sign is just to ensure that the area value should not be negative, but the calculation still remains same. And you need to consider one caution here that all the vertices of a triangle in an anticlockwise direction then only this formula is valid. So, I have considered area of a triangle.

Now what we have seen so far is given two points how to find the distance between two points, given three points if they are collinear, we have found the section formula that can help us to find their ratios or the coordinates of the middle point. Now, if the points are non-collinear, we have seen how to compute the area of the triangle using the coordinate system.

IIT Madras

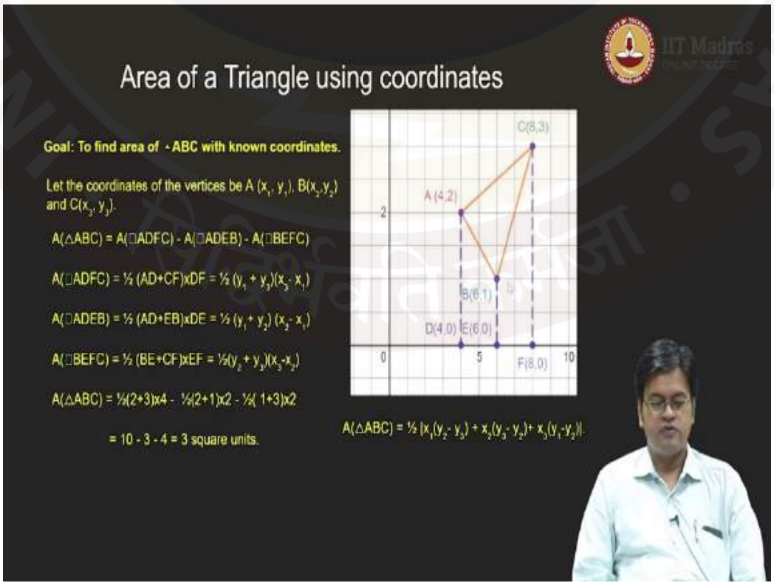
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 16
Slope of a Line

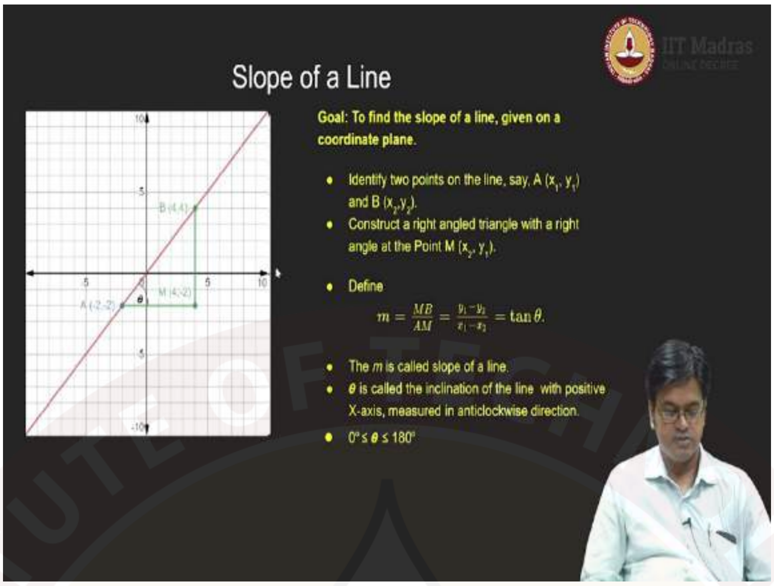
(Refer Slide Time: 00:05)

(Refer Slide Time: 00:14)



So, after looking at area of the triangle using coordinates, let us now focus our attention to again a two-point system and one-dimensional objects that is a line.

(Refer Slide Time: 00:24)



We have already seen in our basic classes that two points uniquely determine line. Now, if I want to characterize a line, and if I give you two points, I should be able to find a line passing through these two points. How is the geometric object algebraically related to the coordinate geometry? That is what we want to explore now, to explore that I need a concept of a slope of a line.

So, what essentially is the slope of a line? In a vague manner, what we understand by slope of a line? If you look at this coordinate plane which is displayed here. If I am moving some units in x directions; the question can be asked with respect to this change in x direction what is the corresponding change in y direction.

So, if I want to answer that question then I need to consider a ratio of change in y direction to change in x direction; some people call it as rise by run ratio, run is in the horizontal direction, rise is in the vertical direction. So, you can consider slope of a line as a rise by run ratio. So, let us try to make this work concept clearer by showing some examples.

So, now, here is a line with two points given onto it. Again, our standard conventional method we will construct a right-angled triangle using these two points. Now, the question that I posed is what is a rise by a run can be answered over here. For example, you look at this right-angled triangle, what is happening? This is the movement of a line in moving from one point to another point in y direction, this vertical length is the direction, is the movement of a line in moving from one point to other point in y direction and this horizontal line is a movement in x direction while moving from point A to B on a line.

So, essentially what I need to capture is the change in y direction that is from point $(4,-2)$ to point $(4,4)$, that is -6 and moving in x direction from $(-2,-2)$ to $(4,-2)$ that means, -6 here also. So, the slope of a line can be equal to 1. This we can make it more precise by giving some formal definitions.

So, if I want to find the slope of a line given the coordinate plane, I can always identify these two points as (x_1, y_1) and (x_2, y_2). I will construct a right-angled triangle which intersects the point at(x_2, y_1). And once I constructed, as I mentioned you know what is the change in x direction and what is the corresponding change in y direction, therefore, you can actually compute the ratio of this. But while computing the ratio, you can also think remember some concept from trigonometry.

For example, when I constructed this right-angle triangle there is some angle formed over here this arc denotes that angle. Let us call that angle as theta. Now, what I am saying is change in y upon change in x, but can you relate some quantity related to this trigonometric ratio that is tan of θ, right. So, what I can say is my m or the slope of a line is MB by AM which is $\frac{y_1 - y_2}{x_1 - x_2}$and which is also equal to$\tan\theta$.

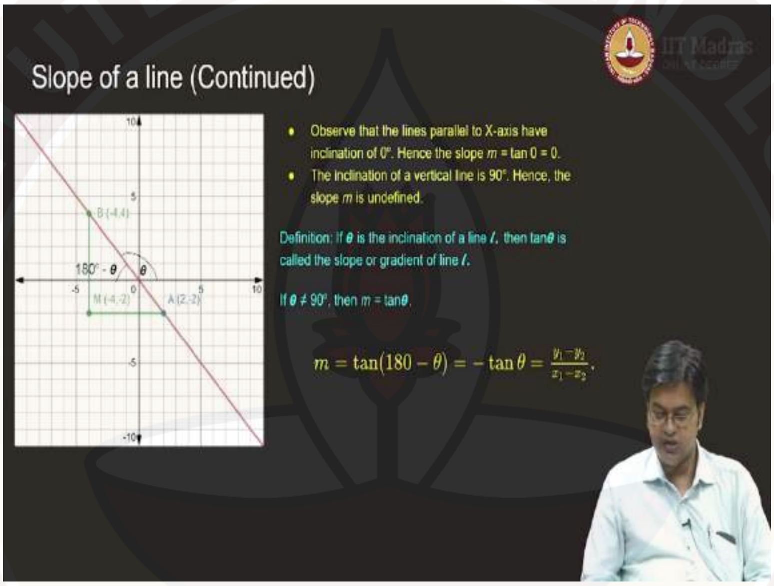
So, I have defined one thing that is m which is the ratio of these two, but which in turn turned out to be equal to$\tan\theta$. So, if it is $\tan\theta$, see here it does not matter whether I take $y_1 - y_2 \vee y_2 - y_1$whatever I am doing I should do synonymously. For example, if I have taken $y_2 - y_1$then I should take $x_2 - x_1$or if I have taken $y_1 - y_2$ then I should take $x_1 - x_2$.

So, it does not matter which order you are swapping because finally you are taking the ratio so whatever you are doing you do it asynchronously, so that there will not be any confusion. So,$m = \tan\theta$. Now, I have introduced two terminologies here m and θ. So, let us define them properly. This m is called slope of a line, which is the topic of this discussion. And then this θ is called inclination of a line with respect to positive X - axis measured in an anti-clockwise direction.

Now, somebody may say I have drawn this angle over here, but if you look at this particular line, this line is parallel to X - axis. And this line is intersecting X - axis here, that means even if I consider this angle, this angle also will be $\theta$from the basics of geometry, correct.

So, now the question can be asked how far the θ can go? So, to answer that question let us try to see if I am considering a θ then θ can be equal to 0, θ equal to 90 degrees tan is not defined. As you can see tan of 90 is not defined, but it can go up to 180 degrees. So, the variation of θ allowed is 0 to 180 degrees.

(Refer Slide Time: 07:42)



So, now let us have a look at the salient features of the slope of a line. In particular, let us see if the line is parallel to X - axis the angle of inclination is 0 degrees; therefore, the slope of a line should be 0. Now, if the angle is 90 degrees; that means, 90 degrees with respect to X - axis; that means, eventually I am on Y - axis or in fact, I am on Y - axis in such case tan 90 is undefined, right. Therefore, slope is undefined.

As you can see if I have an angle which is 90 degrees that is Y - axis; that means, x is equal to constant is the equation of the line. And you cannot have any movement in y direction or you can have infinite movement in y direction without any change in x direction. That itself creates a problem therefore, the slope is undefined for theta is equal to 90 degrees or the inclination is equal to 90 degrees.

So, with respect to inclination there is another definition of slope. If theta is the inclination of a line l then tan theta is called slope or gradient of the line. This is the second definition of our slope of a line which matches exactly with the original definition, but there will be some glitch, there may be some confusion, ambiguity.

So, let us try to resolve that ambiguity because this theta is the angle made with respect to positive X - axis. And theta not equal to 90 degrees I can define $m=\tan\theta$. That is perfectly fine and it is well-defined over there whenever it is not equal to 90 degrees. What is the ambiguity? The ambiguity can be shown in the figure. For example, now what is θover here? θ over here is actually this particular angle.

Now, if I you look at this particular angle which is θ you can see that this is an obtuse angle. Now, how to evaluate a tan of this angle? We already know some methods, but will that contradict with our definition of slope. That is the question. So, if I use the rise by run formula or the change in y to up on change in x formula, how will I figure out the slope? So, the answer is I will simply drop a perpendicular or I will construct a right-angle triangle with right angle at point M which is $(-4,-2)$.

In that case, I will be interested in this angle that is angle at A in our older definition or this angle is essentially equal to $180-\theta$. So, let us go further. This angle is equal to this angle. What is the measurement of this angle? It is$180-\theta$. That means, if I want to find a slope according to our definition that is $\frac{\delta y}{\delta x}$or change in y by change in x, then I need to consider the angle of this particular structure that is $\tan(180-\theta)$. So, $m=\tan(180-\theta)$.

Now, what is $\tan(180-\theta)$? If you use simple trigonometric formula you will get $\tan(180-\theta)$ is nothing, but $-\tan\theta$. But what is $-\tan\theta$? You can easily see what is $-\tan\theta$which will be $\frac{y_1-y_2}{x_1-x_2}$. So, in short, our formula for slope is consistent no matter which definition we use, therefore a slope of a line is uniquely determined given a line.

IIT Madras

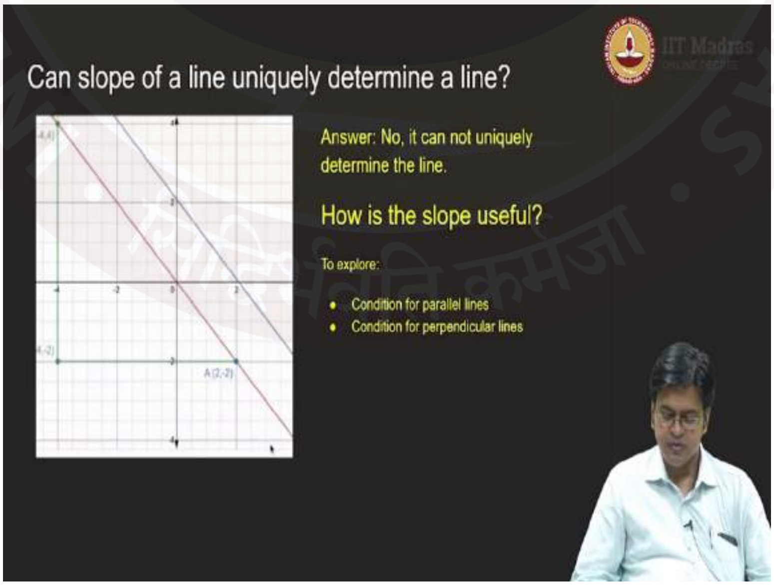
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 17
Parallel and perpendicular lines

(Refer Slide Time: 00:05)

(Refer Slide Time: 00:15)



Now the question can be asked that if a line is given to me, I can uniquely determine the slope, but if a slope is given to me can I uniquely determine a line? That is the next question that I will put up. In any sense the question asks can there be many lines with same slope? The answer can be seen in this GIF image.

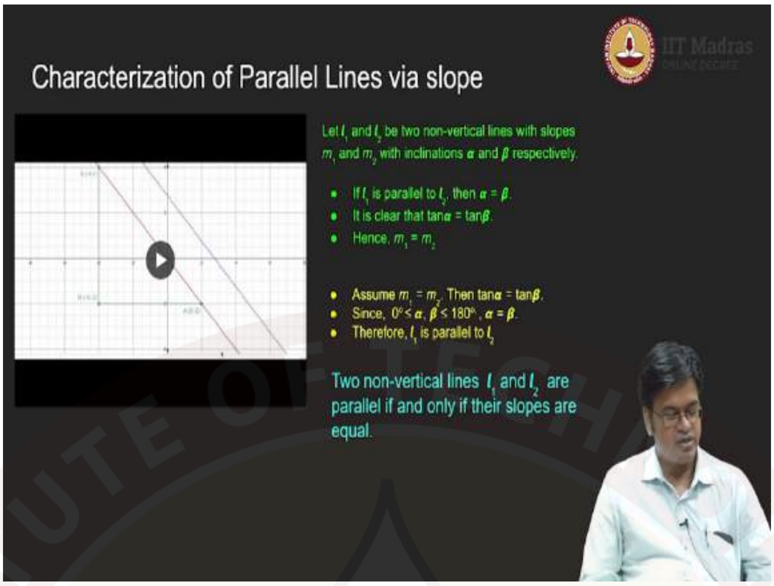
If you look at this image closely what we have done is? We have fixed one line and we know how to compute the slope of this line we have a it will be minus 1 based on the coordinates. Now, the blue line that is revolving around is actually having the same inclination as the orange line.

Now, the orange line and blue line have the same inclination; that means, tan of those inclinations will be same, will match and hence there can be infinitely many parallel lines which have a same slope. So, the answer to this question, can slope of a line uniquely determine a line? The answer is no, you cannot uniquely determine a line given the slope of a line or the inclination of the line.

Now, why do we study the concept of slope or whatever we studied how it is helpful? The helpfulness of this concept is just what we discussed in this graphical image, what we are seeing is if the inclinations are same the line better be parallel. So, for parallel lines I can use this concept and derive a condition of slope. Similarly, I can do by rotating them by 90 degrees; that means, I we can consider the perpendicular lines and I can consider general two lines intersecting each other and see what condition I can derive based on the slope.

So, I want to explore the usefulness of slope. So, to explore this I will first figure out the condition for parallel lines and I will figure out the condition for perpendicular lines, in due course we will find the relation between slopes of two lines and their intersection and their angles of intersection. This is what we will do in next few minutes.

(Refer Slide Time: 03:15)



So, let us go to the next characterization of parallel lines via slope. Now as you can see in this image there are two parallel lines, they have same inclination, but they are not unique that is what we figured out. So, if I play this video you can see again, this is similar to what we have seen in the last video.

So, I have something which is moving around and there can be infinitely many lines, what remains constant is the inclination, the inclination is same if I have parallel lines. So, let us try to see whether we you can derive something. So, let to put it in a proper context.

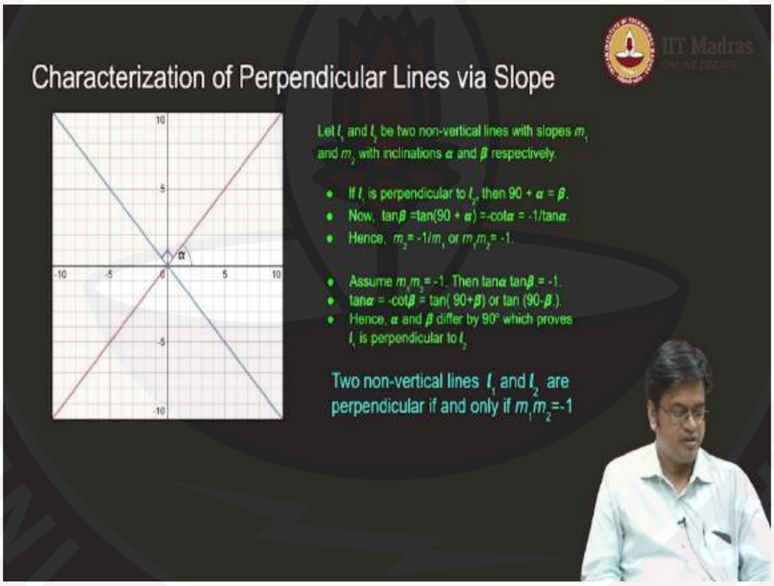
Let orange line be $l_1$and the blue line be l_2 be two non-vertical lines. Why non-vertical lines? Vertical lines have angle of 90 degrees for which the concept of slope is undefined, inclination 90 degrees for which the concept of slope is undefined. So, what I need is non-vertical lines. So, considered two non-vertical lines with slopes $m_1 \wedge m_2$given the slopes their inclinations $\alpha$and βrespectively.

Now, if you have been given that $l_1$is parallel to $l_2$then $\alpha = \beta$, inclinations are same that is what we have seen in the figure and that is what we discussed in the last slide also. So, if $\alpha = \beta$ then naturally $\tan\alpha = \tan\beta$, once $\tan\alpha = \tan\beta$; what is $\tan\alpha$? It is the slope of line1_1 that is m_1 and $\tan\beta$is the slope of line $l_2$which is m_2. Therefore, clearly the slopes are equal, $m_1 = m_2$.

The converse that is assumed that, if the slopes are equal then $\tan\alpha = \tan\beta$ by a definition. Now, $\tan\alpha = \tan\beta$ does that imply α is equal to β? In our case because we are restricting the inclinations to vary from 0 to 180 degrees the value of tan is uniquely determined. And therefore, because $\alpha \wedge \beta$ lie in 0 to 180 degrees $\alpha = \beta$ which resolves the problem; that means, their inclinations are same. That means the two lines are parallel. So, l_1 is parallel to l_2.

So, what is a characterization of parallel lines? That means, if I want to say two non-vertical lines l_1 and l_2 are parallel then it suffices to check whether their slopes are equal or not. If they are parallel then the slopes better be equal and if the slopes are equal then we have parallel lines. Now similar characterization we are searching for in perpendicular lines. So, let us go and try to figure out this characterization for perpendicular lines.

(Refer Slide Time: 06:42)



Let us try to visualize, what are the perpendicular lines? So, here are two perpendicular lines one l_1 and l_2 let us take the orange line as l_1 and blue line as l_2. So, l_1 will have slope m_1, l_2 will have slope m_2 angle of inclination of l_1 is α then inclination of β, if it is perpendicular to line l_1 is $90 + \alpha$ which is β. And, then you may play with the tangent function of it and you can get something which is very interesting.

So, let us try to figure out what is that interesting thing that we are getting. So, to put it formally let l_1 and l_2 be two non-vertical lines because I cannot work with vertical lines θ equal to 90 degrees, the concept of slope is not defined which slopes m_1, m_2 inclinations α

and β respectively, no problem in this. If l_1 is perpendicular to l_2 as is the case in this a figure I have β is equal to$90+\alpha$.

So, if I want to figure out the relation between the slopes of 1 1 and l_2 then it is a good idea to take tangent of β. So, let us take that. So, $\tan\beta=\tan(90+\alpha)$, but $\tan(90+\alpha)$ if you use that simple formula that is available to you is $-\cot\alpha$ which also can be written as $\frac{-1}{\tan\alpha}$.

But what is $\tan\alpha$? $\tan\beta$ is the slope of a line l_2 which is m_2 and $\tan\alpha$is the slope of a line l_1 which is m_1. So, what we have just now derived is $m_2=\frac{-1}{m_1}$ or $m_1 m_2=-1$. That means, if you take two slopes if you take slopes of two lines take a product of them and if you get the quantity to be equal to -1; that means, you have got a perpendicular line.

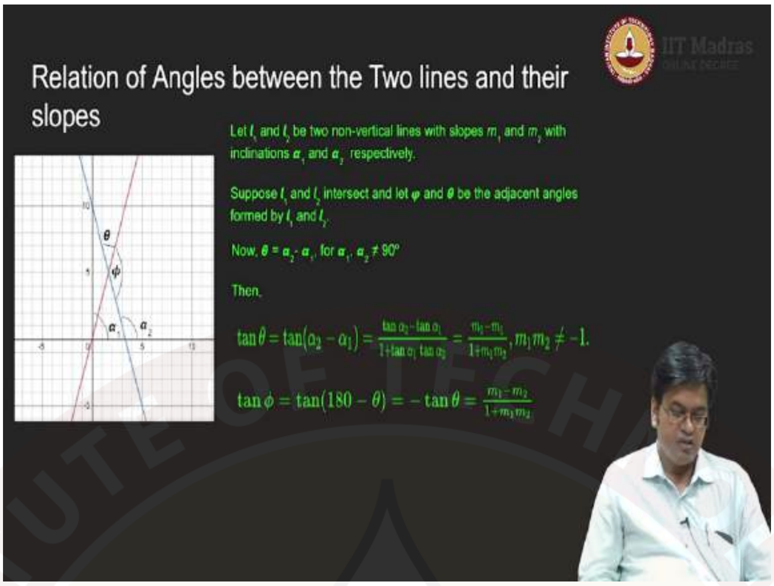
But right now, we have not proved that result, what we have proved just now is if l_1 is perpendicular to l_2 then the product of the slopes better be -1. Now I want to prove if the product of the slopes is -1 then the lines are perpendicular, how will I go about this? Exactly the way we went for parallel lines.

So, $m_1 m_2 is -1$ then I; obviously, $\tan\alpha\tan\beta=-1$; that means, $\tan\beta$will be equal to $\frac{-1}{\tan\alpha}$ or $\tan\alpha=-\cot\beta$but what is $-\cot\beta$? $\tan(90+\beta)$or either it will be this way or it will be the other way so, $\tan(90-\beta)$. So, $-\cot\beta$ is either $\tan(90+\beta)$ or $\tan(90-\beta)$, in any case the difference between α and β is 90 degrees.

Therefore, l_1 is perpendicular to l_2. Hence, we have proved a characterization that if two non-vertical lines are perpendicular to each other, the product of their slopes is equal to -1 which can be written in this form. Two non-vertical lines l_1 and l_2 are perpendicular if and only if $m_1 m_2=-1$ or you can verbally write product of their slopes is equal to - 1.

So, this is the characterization of the perpendicular lines via slope. So, what we have seen so far is the characterization of parallel lines by slope and characterization of perpendicular lines via slope, what if they are not parallel or perpendicular and they intersect just like that? If they are not parallel then they better intersect each other.

(Refer Slide Time: 11:37)



So, in general if I want to have an intersection of two lines and I know the slopes of those two lines. Can I talk about the angle of intersection of these two lines? The answer is yes. So, here is the relation of angles between the two lines and their slopes. So, what I want to say if once I show the figure it will be clear.

As of now let us understand I have two non-vertical lines with slopes m_1 and m_2, inclinations $\alpha_1$and α_2 respectively. And, l_1 and l_2 intersect each other, they are not parallel so they will intersect somehow and they are not perpendicular also. So, they intersect in angles ϕ and θ are the adjacent angles that are formed by l_1 and l_2, if they intersect in a perpendicular manner the adjacent angles will be 90 degrees each. So, that is not an interesting case because we have resolved that case.

So, now, if they intersect at any angle then this figure will look like this; let us first understand this figure. So, there are two lines l_1 and l_2. So, l_1 has angle of inclinationα_1, l_2 has inclination α_2 these two lines intersect over here near y coordinate $\phi$and they have two angles; one is θ, another one is ϕ.

So, these two angles are adjacent angles. What can you say about the angle θ that is formed? As you can see the angle α_2 is obtuse and α_1 is slight acute. So, the angle θ is actually α_2 minus α_1 provided α_1 and $\alpha_2$are not equal to 90 degrees. Why? Because I cannot consider vertical lines as simple as that. So, the angle is 90 not equal to 90 degrees, $\theta = \alpha_2 - \alpha_1$.

So, if I want to talk in terms of slopes of these lines, I better take tangent function and apply it to the angle θ. So, let me do it. So $\tan\theta = \tan(\alpha_2 - \alpha_1)$. Take a standard trigonometric formula of $\tan(\alpha_2 - \alpha_1)$, you will get $\frac{\tan\alpha_2 - \tan\alpha_1}{1+\tan\alpha_1 \tan\alpha_2}$. But what is $\tan\alpha_2$? $\tan\alpha_2$is nothing but the slope of line l_2 which is m_2 and $\tan\alpha_1$ it is slope of line l_1 which is m_1.

Therefore, the answer to this is $\frac{m_2 - m_1}{1+m_1 m_2}$. So, I know what is $\tan\theta$, now you can look at the angle ϕ which is $180-\theta$. So, I can similarly derive a relationship for $\tan\phi$which is $\tan(180-\theta)$, we have already seen, this is $-\tan\theta$.So, that $m_2 - m_1$ will be swapped to $m_1 - m_2$ denominator remains the same, the condition $m_1 m_2 \neq -1$remains the same because they should not be perpendicular.

In this case we have figured out what is the relation of tan of that angle with respect to the slopes of the lines. So, this finishes our discussion on two lines. Now another interesting question that comes is, what if the three points are collinear, then how will the slopes be interpreted? Imagine three points are collinear then what happens is their slopes must be equal because they are all lying on the same line right and there is one common point.

So, if A, B, C are collinear slope of AB is equal to slope of BC and therefore, all of them must be collinear. So, if there is any common point in between those three points the slopes are equal, the points are collinear, that is called the relation of collinearity using slopes.

IIT Madras

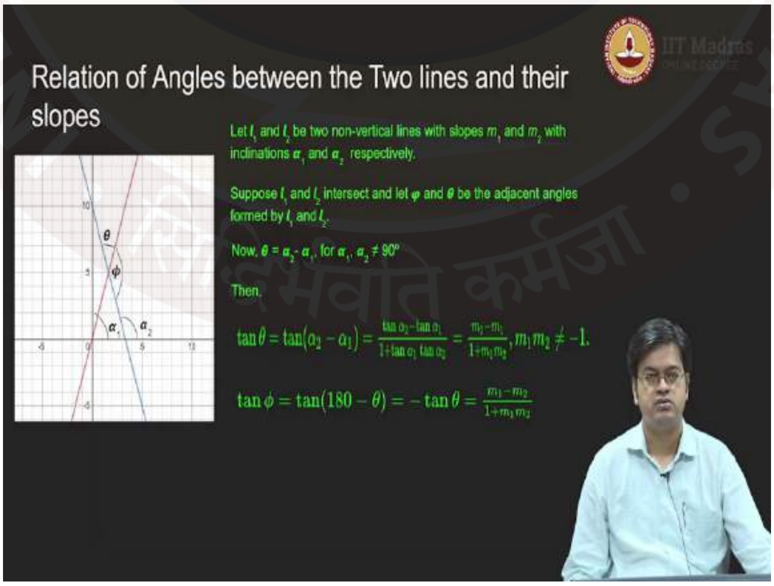
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 18
Representation of a Line-1

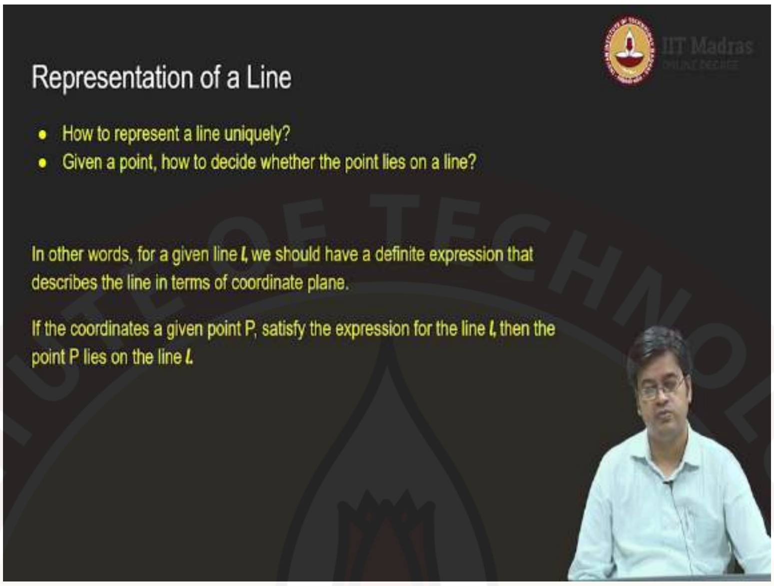
(Refer Slide Time: 00:06)

(Refer Slide Time: 00:14)



So, what we have seen so far is, what is a relation of the slope with respect to line and we have exploited certain how we can use the slope to determine whether the lines are parallel perpendicular. And, if I know the slope of the line then how will I find a slope of two non - vertical lines, then how will I find the relation between angles and other properties.

(Refer Slide Time: 00:41)



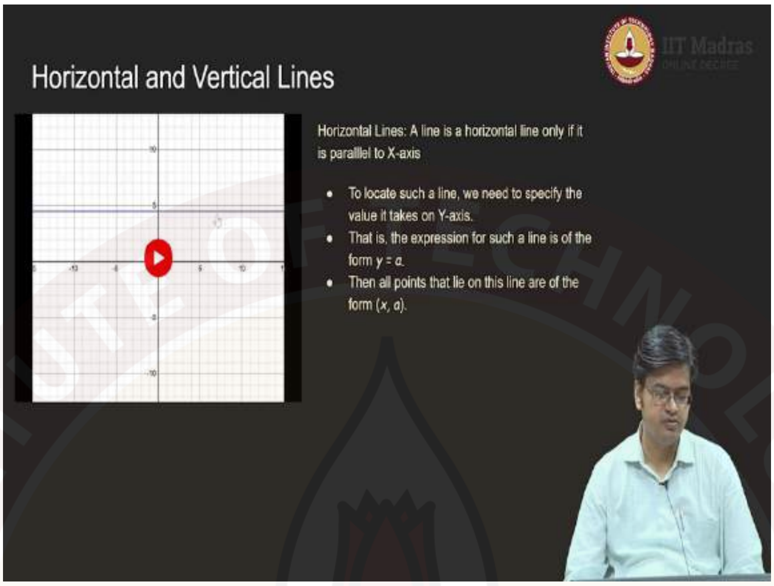
Now, we will come to the Representation of a Line, as we have already seen slope cannot represent a line uniquely. So, what is it that, that is required for representing a line uniquely? So, this raises two questions, how to represent a line uniquely? And the second question is, given any point of how will you decide whether that point lies on the line or not?

So, in order to answer these two questions, let us take the first question first and rephrase it. So, if I want to represent a line uniquely, then what I need to figure out is, I need to figure out a condition or a definite expression which will describe the line in terms of its coordinate plane. So, for a given line l I should be able to find a definite condition or expression which describes the line in terms of coordinate plane. That is in terms of the coordinates or to be more precise what should be the condition on the coordinates in order to describe the line l.

If I can understand what is this condition then the second question is automatically answered because if the coordinates of P are given to me and they satisfy the condition or expression for the line l then they must lie on the line l otherwise they do not lie on the line l, then it is just a simple job of checking whether that condition is satisfied or not. So, with this in mind we will try to answer the first question that is how to represent a line uniquely?

Now, what kind of lines we have seen so far? We have seen lines which are similar to X - axis, lines which are similar to Y - axis; those are typically horizontal and vertical lines.

(Refer Slide Time: 02:37)



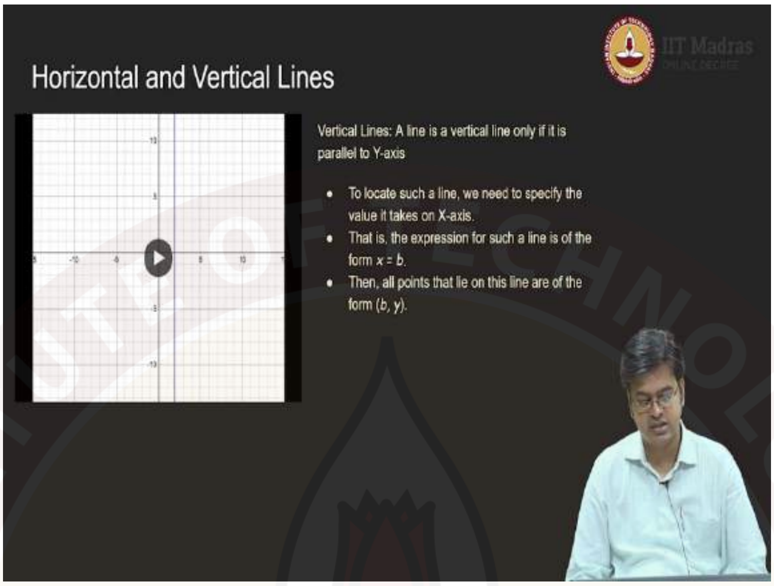
Let us first understand, what is a horizontal line. So, a line is said to be a horizontal line if it is of this form, now this line can be infinitely many. So, you can have infinitely many horizontal lines as can be seen from the video. Now, how to represent this line uniquely is my question. So, let us say I need to find this line or the condition for this line, how can I find the condition for this line?

So, let us first define this line as a horizontal line and let us say horizontal line is a horizontal line if and only if it is parallel to X - axis, this is our definition of a horizontal line. Now, if I want to specify this line uniquely what do I need to know? I need to know the distance of this line or the location of this line from X - axis, that is I need to know the y coordinate of this line you can see here. So, I want to locate this line or the value that it takes on Y - axis if I want to specify this line.

Let us say this value is given to be a then I know it is a horizontal line. So, all points will lie at a same distance from X - axis therefore, all points will satisfy the condition $y=a$. You take any point on this line it will satisfy the condition $y=a$.

So, in case of horizontal lines what I have done is I have identified the condition that is $y=a$. So, what will be the condition on points? The points will be of the form (x,a), x can be any value, but the y coordinate of that point will be fixed that is a. In a similar manner we can consider vertical lines.

(Refer Slide Time: 05:02)

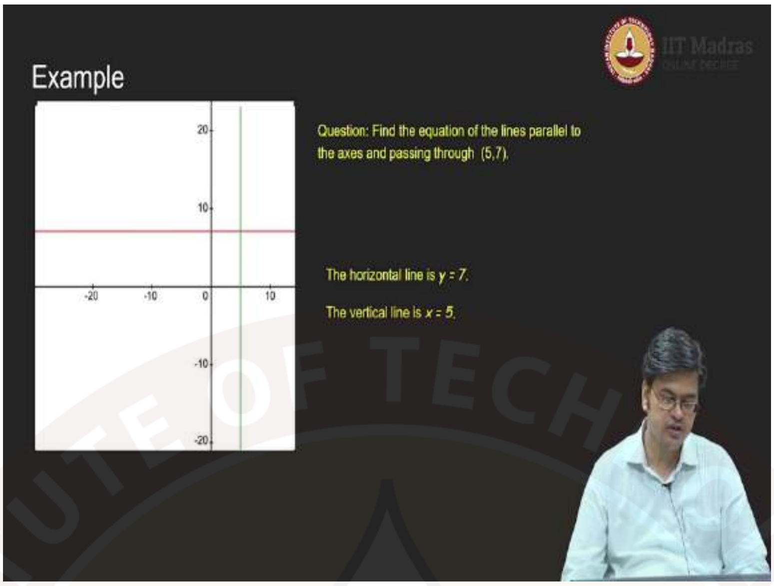


So, what is a vertical line? You can see this in this image and in this video, we can see all these kind lines are called vertical lines. So, how will I identify these vertical lines? First, I will define the vertical lines a line is a vertical line if and only if it is parallel to Y - axis. Now, to specify the location what do I need? To locate a line, I need to know the distance of this line from Y - axis ok. So, that essentially means what value it takes on x coordinate or X - axis.

So, how will you identify this? You just need to identify the one point in this particular line let us say this is the point and I need to see what is the distance of this point from X - axis, if that is b, then all points of the form (b, y) will be lying on this line; all points of the form (b, y) will be lying on this line. And therefore, the equation of the line the expression for the line will have a form $x=b$.

I mentioned the all points will be of the form (b, y). So, if I get two points where the y coordinate where the x coordinate is fixed and I know it is a line then I know it is parallel to Y - axis or it is a vertical line right. In a similar manner the other one is parallel to X - axis and it will be a horizontal line. Let us make it more crystal clear by solving one example.

(Refer Slide Time: 06:49)



So, here is an example where a question is given to you want to find the equation or expression for the lines parallel to the axis and passing through point $(5,7)$. Now, the lines are passing through point $(5,7)$ and it is also given that they are parallel to axes. So, a line which is parallel to X - axis is known as horizontal line, a line which is parallel to Y - axis is known as vertical line. So, essentially this question asks you to find one horizontal line and one vertical line.

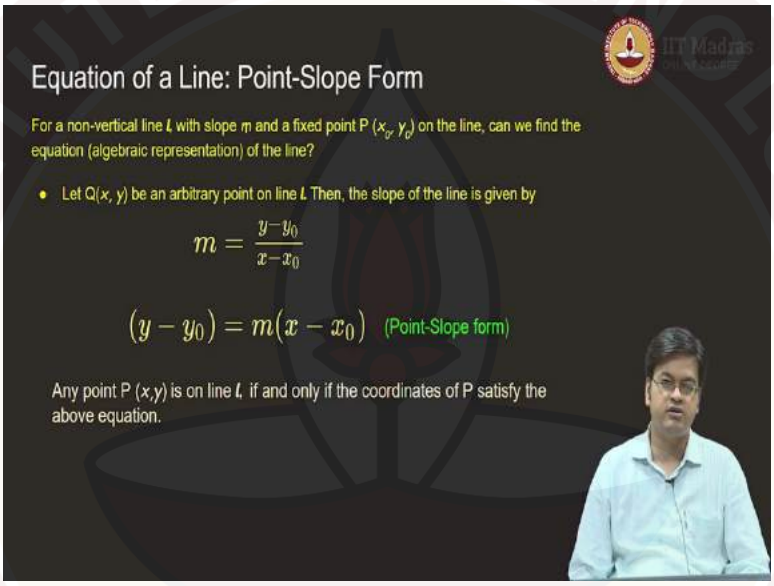
So, let us go to the coordinate plane, this is the coordinate plane let us locate the point $(5,7)$, it will be somewhere here. Now let us first focus on identifying the horizontal line. What is a horizontal line? A line which is parallel to X - axis is a horizontal line. So, a line which is parallel to X - axis, then what do I need to know? Its distance from X - axis, the distance is 7 according to this particular expression because $(5,7)$ is a point on that line.

So, the distance is 7 so, the line must appear somewhere here, now further the next question is I want to find a vertical line that passes through point $(5,7)$. So, now I need to know the distance of a line from X - axis. So, I will locate point 5 over here and all points on the line on that particular line will be of the form $(5,y)$. So, $(5,7)$ will also fall on that line. So, this is the line; so, this is how we will find the lines.

Now what are the typical equations of the line? So, the horizontal line will be $y=7$ and the vertical line will be $x=5$. This is how we will study horizontal and vertical line. So, what is a vertical line? Vertical line has inclination as at 90 degrees, and therefore, the slope of this line is not defined remember this in mind.

Another point which is horizontal line it never intersects actually X - axis, but the inclination of this line with respect to X - axis is 0 degrees therefore, it will have a slope 0. So, we have eliminated the cases where the slope does not exist or slope is 0, now we need to identify similar kind of expressions for lines which are not vertical. So, let us go further and identify such expressions.

(Refer Slide Time: 09:48)



So, here as we already know that slope cannot uniquely determine line, then the question is which slope if I give you some more information can you determine the line? So, here in this case what we are identifying is we are giving a point and giving a slope and then we are asking a question can we solve this problem or can we find a unique expression for a given line? So, the question is for a non-vertical line l vertical line we do not have to consider because the slope is not defined. So, for a non-vertical line l with slope m and a fixed-point P (x_0, y_0) on the line can we find the equation or the algebraic representation of a geometric object that is line is the question.

So, here what are the things that we know? We know slope and we know a point on the line. So, in order to answer this question, we know that two points uniquely determine a line. So, let us take another point Q (x, y). I do not know the coordinates of these points, but I assume that this point lies on line l. Now, I know from the definition of slope that I have defined change in y by change in x the slope of a line is given by. So, what are the two points now? Q and P.

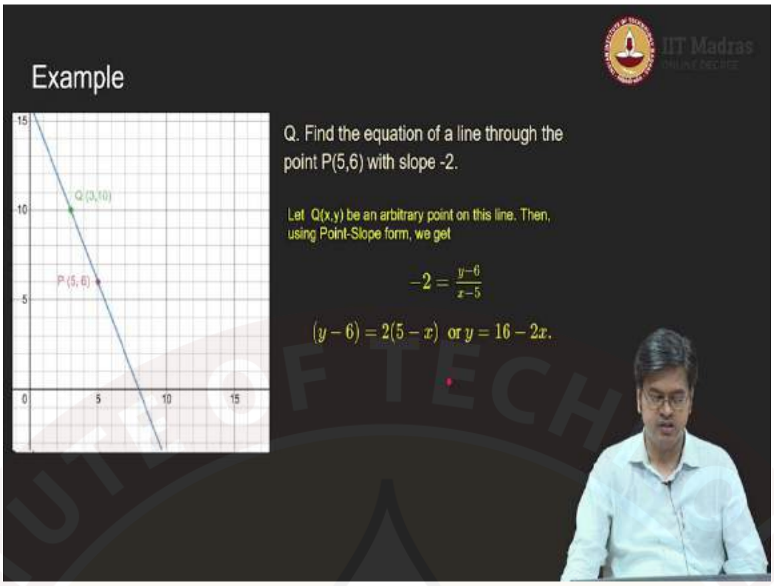
So, change in y will be $y-y_o$ and change in x will be $x-x_0$. So, I know $m=\frac{y-y_0}{x-x_0}$, this is what I know from my definition. It has nothing to do with $\tan\theta$ even if you have it you can find out what is $\tan\theta$, but since nothing is known in specific we cannot find the $\tan\theta$, but $\tan\theta$is anyway given to you in terms of slope.

So, now I have $m=\frac{y-y_0}{x-x_0}$. So, how will I find the condition on x and y? Just cross multiply this $x-x_0$, you will get an expression which is $y-y_0=m(x-x_0)$. This condition uniquely identifies my line, there cannot be any other line satisfying this condition.

So, therefore, any point that lies on this particular line that is P (x, y) that lies on this particular line, it must satisfy the condition that is given here. This form of expression is called point slope form. So, this is a point slope form of equation which essentially says that give me one point and slope of the line I will give you the equation of a line.

The beauty is the geometric object now can be represented in terms of the equation, initially when we started, we tried to represent a point which is a geometric object in terms of coordinate plane and the coordinates of the point. Now we are giving infinite set of points having certain condition that is a geometric object of line how you can represent it algebraically using the equation of a line. So, this is point slope form. Let us try to see how we can use the point slope form in our problem solving.

(Refer Slide Time: 13:40)



So, now, I have been asked to find the equation of a line which passes through point $(5,6)$ and has slope of -2. Here the interesting thing is slope is negative. So, let us identify the point $(5,6)$ on the coordinate plane and now I want to identify the line that passes through this with slope of -2. So, now I have a formula for point slope form, I can use that formula and I can straight away derive it for let us try Q (x,y) is an arbitrary point.

So, I need two points to identify a line. So, Q (x,y) with the arbitrary point on this line then using point slope formula we simply substitute $-2=\frac{y-6}{x-5}$, slightly rearrange the terms; what you will get is $y-6=2(5-x)$. If you simplify this you will get the expression $y=16-2x$.

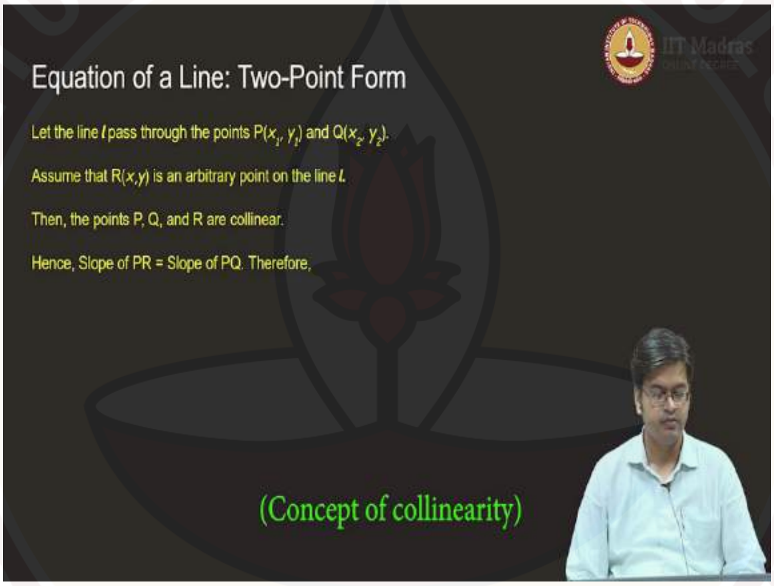
Now, let us try to see, if I want to know this value of x what point what value of y will satisfy this equation. Let us put x is equal to 3 here if I put x is equal to 3 here then I get y is equal to 10 after simplifying this I will get $y=10$.So, that means, the point $(3,10)$ should lie on this particular line. So, let us see that $(3,10)$ is here and now you know from basic geometry that two points uniquely identify a line. So, you can just draw a line using your ruler passing through these two points this is the line that we are expecting.

So, the question did not ask you to draw a graph, but drawing graph always verifies whether you have found a correct answer or not. So, it is better to cross check using graphs. So, the answer to the question is the equation of the line passing through point $(5,6)$ and slope -2 is

$y = 16 - 2x$. Now, suppose somebody decides not to give me slope and somebody says that now you have been given only two points; can you find the equation of line?

The answer is; obviously, yes because given two points I can always determine the slope right for example, in our earlier case when we defined slope I need to figure out what is change in y and what is change in x using these two points and that will give me slope to be equal to -2. And therefore, I can always use this formulation to find the equation of the line, but you can use this knowledge and derive another form that is equation of a line two - point form.

(Refer Slide Time: 17:00)



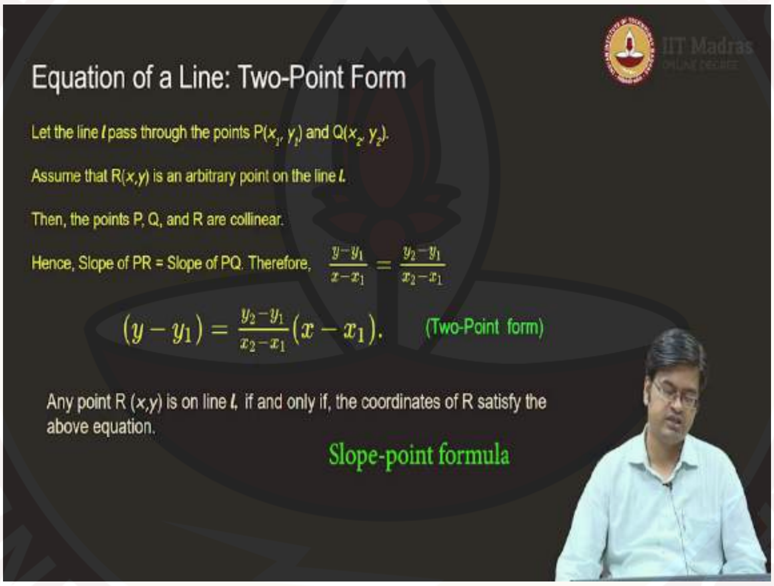
So, given two points the question is can you determine the line uniquely which should be possible and through our basic knowledge of geometry we already know that two points uniquely determine a line, now we will see that in our coordinate geometry. So, the assumption is let the line l pass through points P and Q with coordinates (x_1, y_1) and (x_2, y_2).

To start with this, I will take another point R which is arbitrary point because I want to find the condition in terms of coordinates. So, whenever I want to find the equation of line I will start with an arbitrary point. So, R (x, y) is an arbitrary point on the line l. Now, look at these three points P, Q, and R they all lie on one line therefore, the points P, Q, and R are collinear yes; so, points P, Q, and R are collinear.

Therefore, suppose I consider only these two points P and R, using these two points P and R, I can easily figure out the slope of a line. If I consider points P and Q, I also know the slope of a line; now because these points are collinear what can you say about slope of line PR and slope of line PQ, both must be same or equal? So, slope of PR is equal to slope of PQ because they are collinear.

So, if this is the case, then what is slope of PR? You can easily figure out P is this $(x_1, y_1)$and R is (x, y). So, the slope of PR first you consider change in y, $y-y_1$ upon change in x that is $x-x_1$that is slope of PR. What is slope of PQ? PQ is(x_1, y_1) and (x_2, y_2). So, y change in y is y_2-y_1 and change in x is x_2-x_1.

(Refer Slide Time: 19:21)

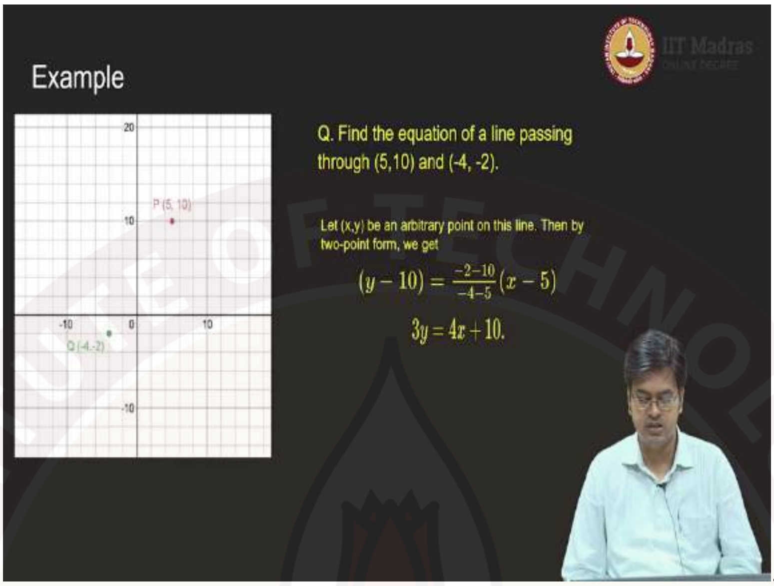


Therefore, I will get the equation of this form $\frac{y-y_1}{x-x_1}=\frac{y_2-y_1}{x_2-x_1}$.

So, again if you look at it closely this particular thing is nothing, but the slope of a line m and we are doing things which are very similar to slope point form. But instead of counting it explicitly we are counting it as a ratio and then you rearrange the term and you will get this expression because you just take this denominator on the other side and you will get this expression.

Now, this line is again uniquely characterized and therefore, any point that lies on this line must satisfy this condition. So, if your point is R lies on this line then it must satisfy this condition and this form is called two - point form. So, remember these are the formulas that we are deriving; first was slope line formula, second is two - point form.

(Refer Slide Time: 20:38)



Let us understand this formula better by solving some examples. So, let us take one example where I want to find the equation of a line that is passing through two points $(5,10)\wedge(-4,-2)$. Let us identify these two points on a coordinate plane $(5,10),(-4,-2)$. I want to find the equation of this line.

So, I will use another point Q which is an arbitrary point and it has a coordinate(x,y), I will use the two - point form. So, using two- point form what should I get? So, I am taking, this point P. So, $(y-10)=\frac{-2-10}{-4-5}(x-5)$. So, always remember this order does not matter I can always start with this as well. So, change in y is $10-(-2)$ and change in x is $5-(-4)$ in both cases my answer to this particular fraction will be $\frac{12}{9}$ which is $\frac{4}{3}$.

So, it does not matter whether you take this as (x_1,y_1) or you take this as (x_1,y_1), you will always get the same answer. So, if you simplify this you will get the expression of a line because as I mentioned the slope was $\frac{4}{3}$. So, you just simplify this you will get the expression of a line$3y=4x+10$. So, this will be the line that is passing through these two points.

IIT Madras

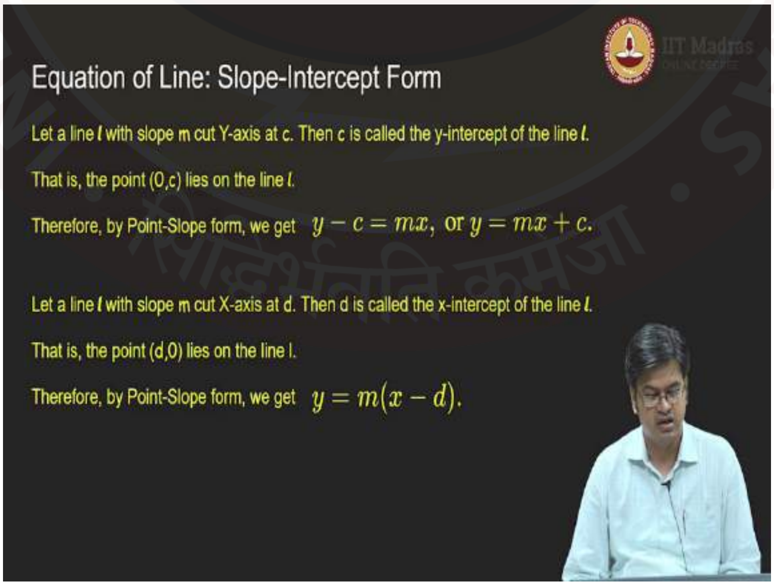
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 19
Representation of a Line-2

(Refer Slide Time: 00:05)

(Refer Slide Time: 00:14)



Now let us go ahead and try to figure out some spatial variations where the calculations become extremely easy for. These two forms are primary two-point form and slope-point form. So, when you consider slope point form you can also consider a special case that is slope-intercept form. So, this is the methodology that we will use for considering slope-intercept form, before that let me define what is an intercept.

So, let l be the line with slope m that cuts Y - axis at point c. Then this c is called y intercept of the line l. So, what is the meaning that it cuts Y - axis at c? The y coordinate of that point is c and the x coordinate is 0; that means, any point that it cuts through Y - axis of line l will be of the form $(0, c)$ and that $(0, c)$ will lie on line l.

Now we have our slope point form instead of having any point (x, y) you have a specific point which is $(0, c)$. So, I apply the slope point form or point slope form in this expression. What you will get instead of $y-y_0$ you have $y-c$which is equal to m, m is the slope of the line m times $x-x_0$. What is x_0? Zero.

So, so we will get $y-c=mx$ and therefore, I will get a form $y=mx+c$, this is a standard form that we generally deal with when we are dealing with straight lines. So, you have got a slope-intercept form which is of the form $y=mx+c$.

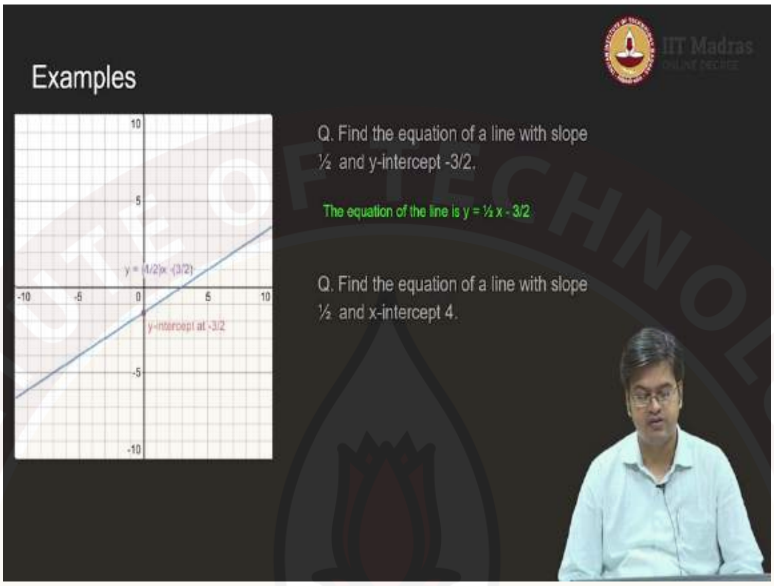
The interesting fact is the calculations are very simple whenever you are given the slope-intercept form. For example, now if you know the y intercept is at c and the slope is m you do not have to do any calculations, but straight away write this expression that is y is equal to take the slope m, take the intercept c; $y=mx+c$ will be your answer.

Therefore, the calculations simplify significantly when you are considering a slope-intercept form. If the intercept is not available then you may have to go to that point slope form and figure out what it is. Now there can be if the line cuts Y - axis the line can as well-cut X - axis. So, there can be another variation of this formulation that is if a line l with slope m cuts X - axis at point d. Then d will be called as x intercept of the line l.

If d is called as x intercept of the line l then how will this point lie on the line l or what are the coordinates of the line that intersects X - axis and line l? So, what is the point of intersection? That will be $(d, 0)$ and this $(d, 0)$ lies on line l. So, I will again use the point slope form of the line.

So, if I want to use point slope form $y-0=m$¿) will be the answer. So, that will be the form $y=mx-md$. So, let us try to use this and solve some problems for finding the equations of the line using slope-intercept form.

(Refer Slide Time: 04:22)



So, typically some example like this. So, I want to find the equation of a line with slope is $\frac{1}{2}$ and y intercept is $\frac{-3}{2}$. Remember here things are very easy because you just need to know $mx+c$. So, what is m? m is $\frac{1}{2}$ and c is $\frac{-3}{2}$. So, upfront I can tell you orally this, the equation of the line will be $y=\frac{1}{2}x-\frac{3}{2}$. Let us verify the result using the graphics and all other things.

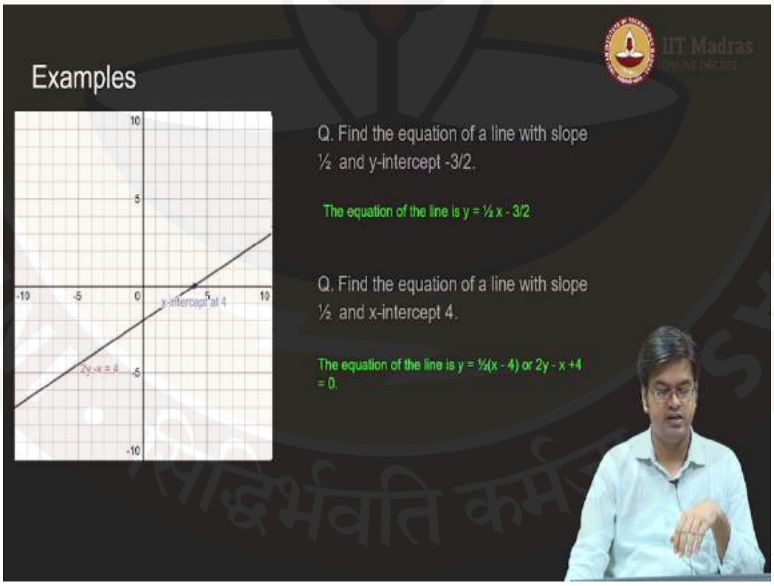
So, here is the y intercept of this particular line. So, here the y intercept is at point $\frac{-3}{2}$. Now slope is half correct. So, the equation of line you can easily see is $y=\frac{1}{2}x-\frac{3}{2}$. So, let us try to figure out what is the x intercept of this line. So, $y=\frac{1}{2}x-\frac{3}{2}$. So, the x intercept of this line is 3. So, the question could have been asked that find the equation of a line with slope half and x intercept equal to 3 that also can be a question and the answer will be same.

So, let us see what is the next question that is find the equation of a line with slope half, but x intercept is 4 it is not 3. So, it is definitely not a same line because x intercept is 4, but the slope is half. So, can you relate it to some of the concepts? The slope is half; that means, the slopes are equal, we have seen that if the slopes are equal then lines must be parallel to each other.

So, therefore, I can easily see that the line must be parallel to this line with some different intercept which is at 4 for this the intercept is 3 so, intercept is 4. So, what can be the y intercept can also be an interesting question. We will answer it later. Right now, let us see how we can answer the question that is asked here. Find the equation of line with slope half and x intercept 4.

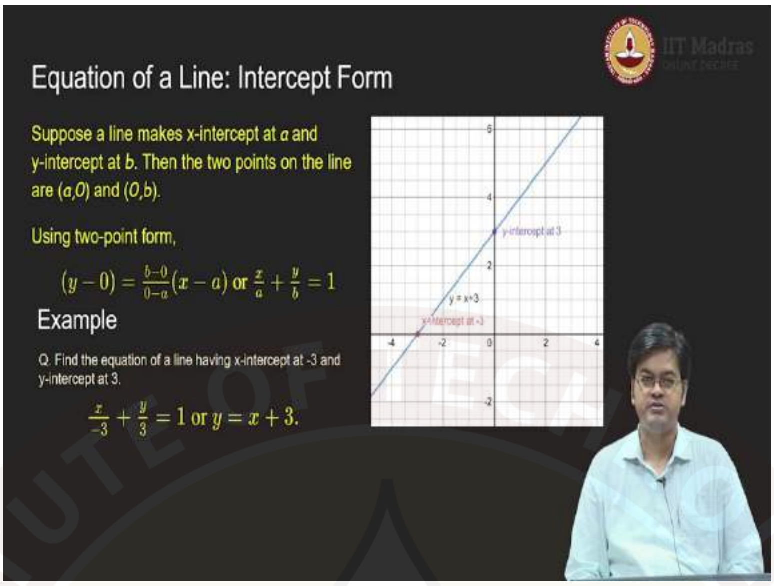
So, according to our formulation $y = mx - md$. So, where d is the intercept that is 4 so and this is half. So, $y = \frac{1}{2}x - 4$ is the equation of this line.

(Refer Slide Time: 07:00)



You can simplify this which will give you $2y - x + 4 = 0$. So, this will be the expression for the line. This is the slope-intercept form of the line, now we can go to two-point form that is suppose I have been given x intercept and y intercept how will I identify the line.

(Refer Slide Time: 07:33)



So, let us now go to the form of intercept that is intercept form, how to find equation of line when you have been given two intercepts x and y. So, let us formulate the hypothesis, suppose a line makes x intercept at a, y intercept at b, then naturally the coordinates of these two points are $(a,0)\wedge(0,b)$. So, we will use two- point form to derive the equation of line.

So, I will take this point as the first point therefore, the y coordinate is 0. So, $(y-0)=\frac{b-0}{0-a}(x-a)$.

Now, if you divide this expression throughout by b then you will get $\frac{y}{b}=\frac{-x}{a}+1$. Because this has a minus sign shift it to the left hand side and you will get this expression which is $\frac{y}{b}+\frac{x}{a}=1$, now you see how beautiful is this expression; x intercept is a so, below x you put a y intercept is b. So, below y you put b.

Therefore, there is nothing to memorize, it is just a simple trick that $\frac{x}{x-intercept}+\frac{y}{y-intercept}=1$ that is how you will get the intercept form. So, it is very easy to solve the problems if you remember this trick.

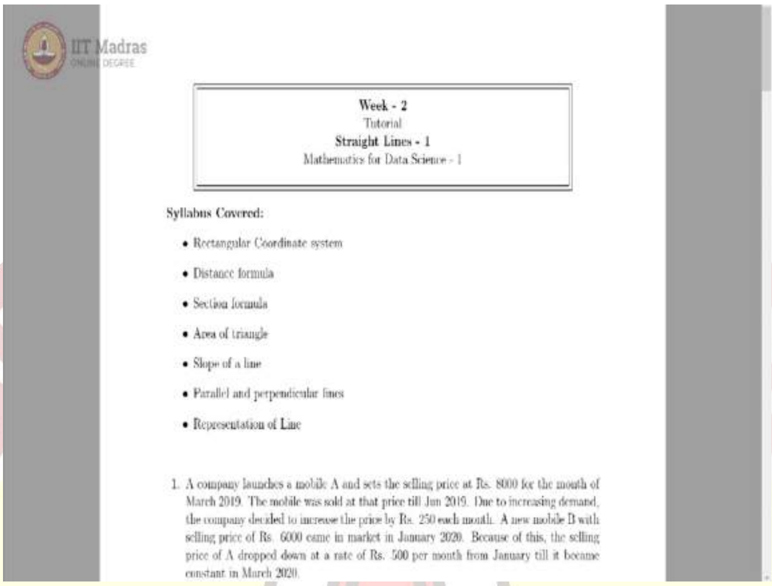
Now, let us take one example where we need to find this. So, find the equation of line having x intercept at -3 and y intercept at 3. So, you do not have to do any complicated calculations, you can simply say $\frac{x}{-3}+\frac{y}{3}=1$, multiply throughout by 3 you will get the expression $y=x+3$.

So, let us verify whether this satisfies because it is always better to verify using graph. So, x intercept is -3 y intercept is 3, the line that passes through these two points is $y=x+3$. This is what the intercept form is, it is very simple and you can practice more and more problems.

That is all for today.

Mathematics for Data Science 1
Indian Institute of Technology, Madras
Week 02
Tutorial 01

(Refer Slide Time 00:19)



Week - 2
Tutorial
Straight Lines - 1
Mathematics for Data Science - 1

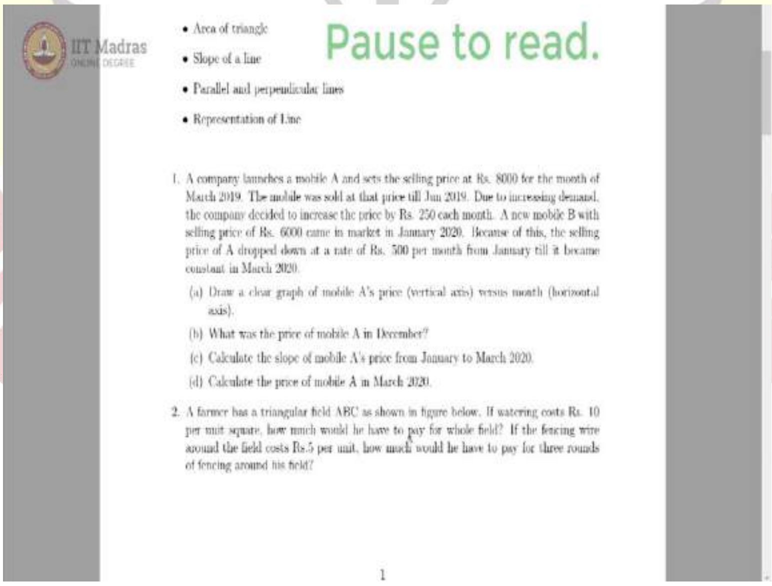
Syllabus Covered:

- Rectangular Coordinate system
- Distance formula
- Section formula
- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.

(())(0:18) In this tutorial we are going to look at the problems which are related to contents of week 2, that is to do with straight lines and all these topics here.

(Refer Slide Time 00:34)



- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.
   (a) Draw a clear graph of mobile A's price (vertical axis) versus month (horizontal axis).
   (b) What was the price of mobile A in December?
   (c) Calculate the slope of mobile A's price from January to March 2020.
   (d) Calculate the price of mobile A in March 2020.
2. A farmer has a triangular field ABC as shown in figure below. If watering costs Rs. 10 per unit square, how much would he have to pay for whole field? If the fencing wire around the field costs Rs.5 per unit, how much would he have to pay for three rounds of fencing around his field?

1

- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.
   (a) Draw a clear graph of mobile A's price (vertical axis) versus month (horizontal axis).
   (b) What was the price of mobile A in December?
   (c) Calculate the slope of mobile A's price from January to March 2020.
   (d) Calculate the price of mobile A in March 2020.
2. A farmer has a triangular field ABC as shown in figure below. If watering costs Rs. 10 per unit square, how much would he have to pay for whole field? If the fencing wire around the field costs Rs.5 per unit, how much would he have to pay for three rounds of fencing around his field?

- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.
   (a) Draw a clear graph of mobile A's price (vertical axis) versus month (horizontal axis).
   (b) What was the price of mobile A in December?
   (c) Calculate the slope of mobile A's price from January to March 2020.
   (d) Calculate the price of mobile A in March 2020.
2. A farmer has a triangular field ABC as shown in figure below. If watering costs Rs. 10 per unit square, how much would he have to pay for whole field? If the fencing wire around the field costs Rs.5 per unit, how much would he have to pay for three rounds of fencing around his field?

1

- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.
   (a) Draw a clear graph of mobile A's price (vertical axis) versus month (horizontal axis).
   (b) What was the price of mobile A in December?
   (c) Calculate the slope of mobile A's price from January to March 2020.
   (d) Calculate the price of mobile A in March 2020.
2. A farmer has a triangular field ABC as shown in figure below. If watering costs Rs. 10 per unit square, how much would he have to pay for whole field? If the fencing wire around the field costs Rs.5 per unit, how much would he have to pay for three rounds of fencing around his field?

1

IIT Madras
ONLINE DEGREE

- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.
   (a) Draw a clear graph of mobile A's price (vertical axis) versus month (horizontal axis).
   (b) What was the price of mobile A in December?
   (c) Calculate the slope of mobile A's price from January to March 2020.
   (d) Calculate the price of mobile A in March 2020.
2. A farmer has a triangular field ABC as shown in figure below. If watering costs Rs. 10 per unit square, how much would he have to pay for whole field? If the fencing wire around the field costs Rs.5 per unit, how much would he have to pay for three rounds of fencing around his field?

IIT Madras
ONLINE DEGREE

- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.
   (a) Draw a clear graph of mobile A's price (vertical axis) versus month (horizontal axis).
   (b) What was the price of mobile A in December?
   (c) Calculate the slope of mobile A's price from January to March 2020.
   (d) Calculate the price of mobile A in March 2020.
2. A farmer has a triangular field ABC as shown in figure below. If watering costs Rs. 10 per unit square, how much would he have to pay for whole field? If the fencing wire around the field costs Rs.5 per unit, how much would he have to pay for three rounds of fencing around his field?

1

IIT Madras
ONLINE DEGREE

- Area of triangle
- Slope of a line
- Parallel and perpendicular lines
- Representation of Line

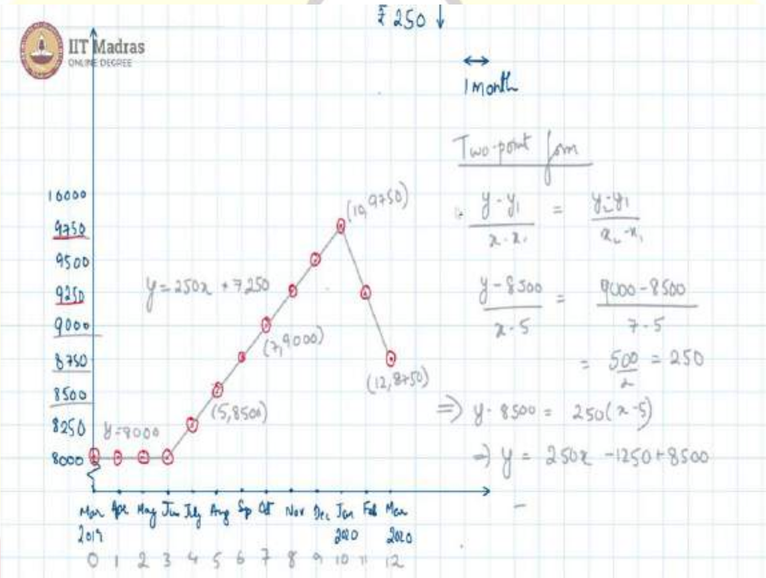
1. A company launches a mobile A and sets the selling price at Rs. 8000 for the month of March 2019. The mobile was sold at that price till Jun 2019. Due to increasing demand, the company decided to increase the price by Rs. 250 each month. A new mobile B with selling price of Rs. 6000 came in market in January 2020. Because of this, the selling price of A dropped down at a rate of Rs. 500 per month from January till it became constant in March 2020.
   (a) Draw a clear graph of mobile A's price (vertical axis) versus month (horizontal axis).
   (b) What was the price of mobile A in December?
   (c) Calculate the slope of mobile A's price from January to March 2020.
   (d) Calculate the price of mobile A in March 2020.
2. A farmer has a triangular field ABC as shown in figure below. If watering costs Rs. 10 per unit square, how much would he have to pay for whole field? If the fencing wire around the field costs Rs.5 per unit, how much would he have to pay for three rounds of fencing around his field?

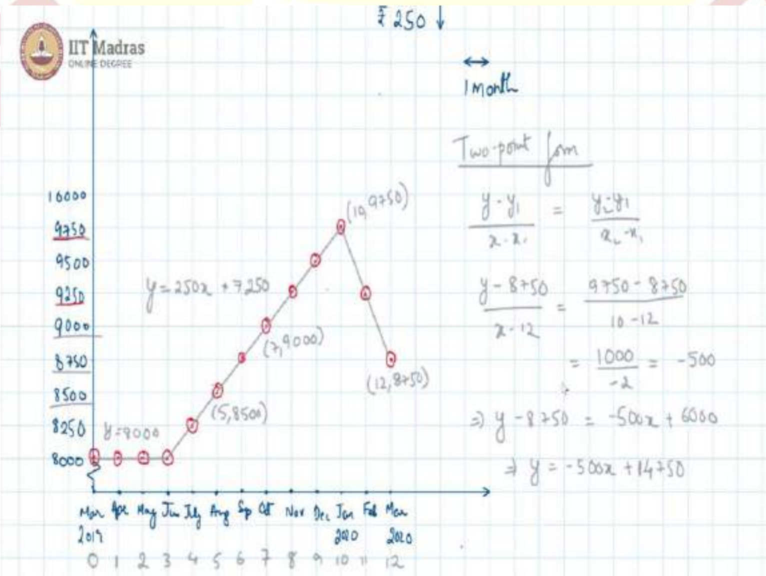
1

So, we will start with our first question. The data provided here is, there is a company which is selling mobile phones and it all begins in March 2019. In March 2019, the selling price was 8000 and it was sold at 8000 rupees, mobile A was sold at 8000 rupees from March until June. After that, due to increasing demand, the company decided to increase the price by 250 each month, so they are selling better.

So, they have decided to increase their price by 250 rupees every month. This went on until a new mobile B was launched at a lesser price, competition at a lesser price was launched in January. So, because of this the selling price of A dropped at a rate of 500 per month, from January till March 2020, so 2 months it had decreased. We are expected to demonstrate a clear graph of this. For that let us look at this graph.

(Refer Slide Time 01:58)





What we need to realize about situations like this is, the x and y axis do not necessarily represent the same units. So, we have along the x axis 1 unit is 1 month, however along the y axis 1 unit is, let us take about 250 rupees. So, 1 month and 250 rupees are not the same thing, so please remember this in situations like this. Now, because we are beginning from March let us take the starting month to be March, then this is April, May, and so on.

So, our entire problem deals with this 1 year span from March 2019 to March 2020, so this will be along our x axis. And now along the y axis, if we took each unit to be 250, then this is 250 and this is 500 and so on, the 8000 will be beyond our screen. So, to better represent our situation, we are going to introduce a zigzag here to indicate that a lot of values have been compressed into this little space. So, we are going to start from 8000 and this is going to be 8250, 8500, so on. And now we begin to mark out the points that we have, we know that in month of March the price was 8000, so this is the point for the month of March.

And then in April, May, and June the price stayed constant so it is been like this. And this portion can be represented using a horizontal line and this line is y is equal to 8000. Beyond that, the price had been increasing by 250 every month so in July we will be here, August here, September here, this will be October, this will be November, this is December, and this is January.

So, this segment can be indicated by this line, in order to find out the equation of this line we use the 2 point form, so we first write 2 points on this line segment. You could choose this one which is August, and for that let us number our months now, so March will be 0, April is 1, May is 2, June is 3, this is 4 and this is 5. So, our price point here it is $(5, 8500)$.

I am ready to take another month, so let us take October, this is the seventh month from March 2019, so this point becomes $(7, 9000)$. Using these 2 points, we can find the equation of the line by employing the 2 point form of the line equation, $\frac{y-y_1}{x-x_1} = \frac{y_2-y_1}{x_2-x_1}$, where x_1, y_1 and x_2, y_2 are two points on the line segment. So, here we can see it as $\frac{y-8500}{x-5} = \frac{9000-8500}{7-5}$. So, this would be equal to $\frac{500}{2} = 250$.

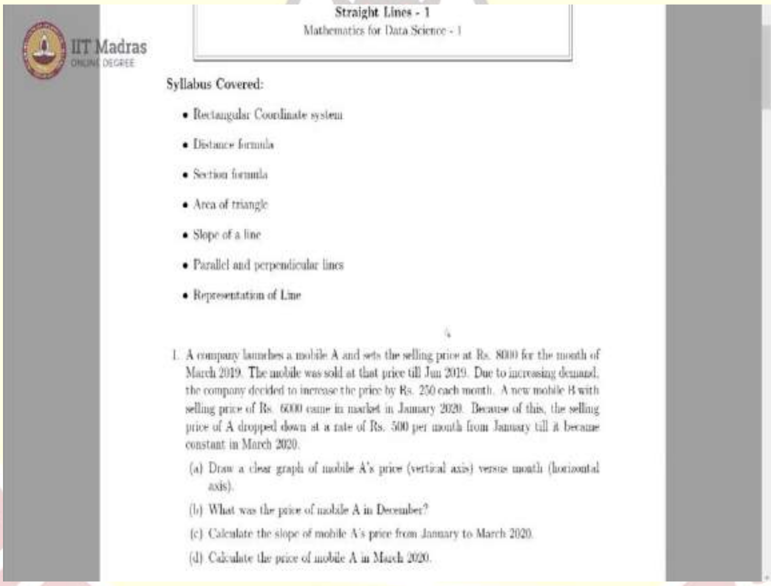
So that implies $y - 8500 = 250(x - 5)$, which finally gives us the line equation to be $y = 250x - 1250 + 8500$ plus this line is y $= 250x + 7250$. Moving on, the next 2 months, the price dropped by 500 each month. So, here we are at 9750, then for February we should be at

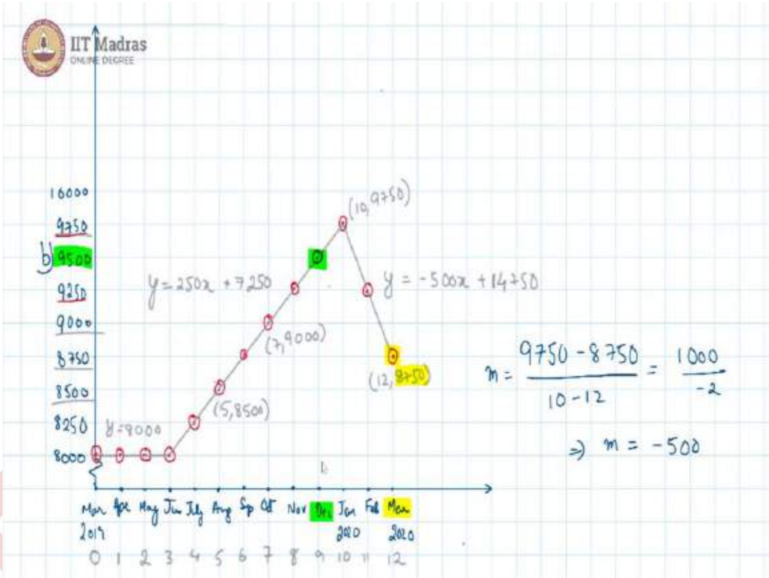
9250, so this will be our point for February and then the next month again 500 drop we will reach here, which is 8750. And this line segment also corresponds to a straight line, which also we can find using the 2 point form.

So, this point here is, let us number the months completely, this is 8, this is 9, this is 10, this is 11, this is 12. So, this point here, which is January is the tenth month, and the y axis gives us 9750 whereas this point here, this is the twelvth month, and it corresponds to 8750. And again, we would like to know the line equation for this and we use the 2 point form again.

So, this is $\frac{y-8750}{x-12} = \frac{9750-8750}{10-12}$ that gives us $\frac{1000}{-2} = -500$. Plus we have $y - 8750 = -500x + 6000$. That gives us $y = -500x + 14750$, so this is our new length. And this is a clear graph of the situation and the given question.

(Refer Slide Time 09:18)





For the part B of this question, it is asked, what is the price of mobile A in December. So, this is the December month, which would be this point here which has a price of 9500, so this is our answer for B. And then in C it has asked, calculate the slope of mobile A's price from January to March 2020, so we want the slope of this segment here and this slope we had already calculated, it was $m = \frac{9750-8750}{10-12} = \frac{1000}{-2} = -500$.

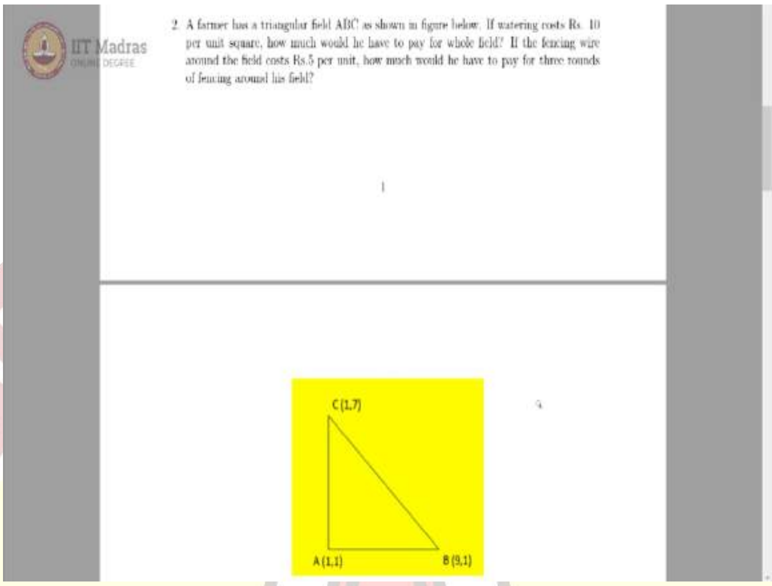
And because of the negative slope you can see that it is a decreasing function, which is what is happening, the price had fallen at 500 per month. Lastly, we have being asked, what is the price of mobile A in March 2020, so this is March 2020, this is a point and we have already found the price which is 8750 that is our part.

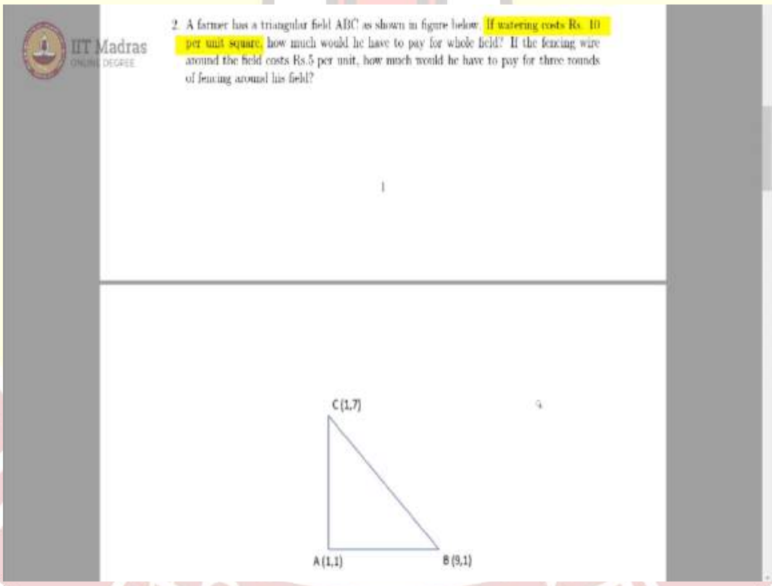
IIT Madras

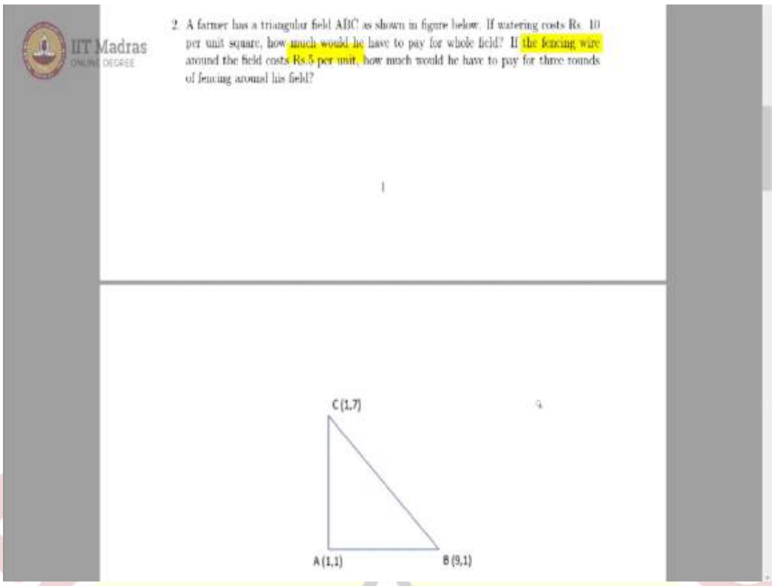
ONLINE DEGREE

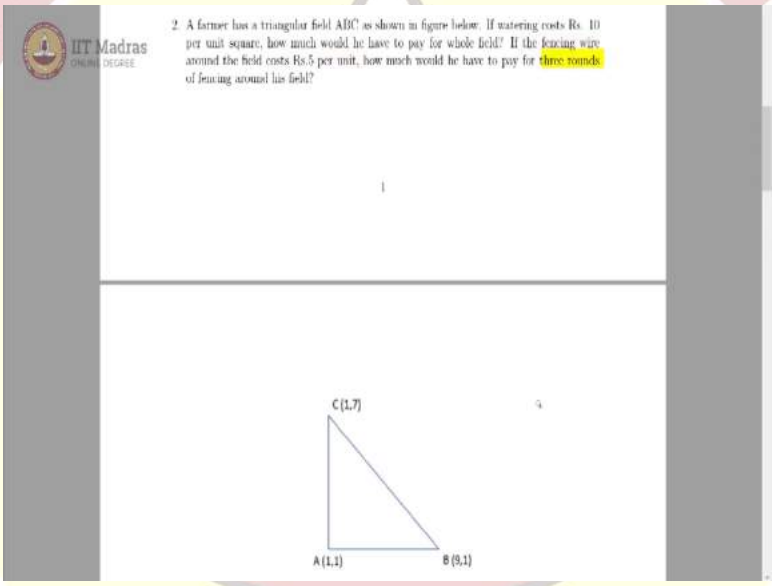
Mathematics for Data Science 1
Indian institute of Technology, Madras
Week 02
Tutorial 02

(Refer Slide Time 00:16)



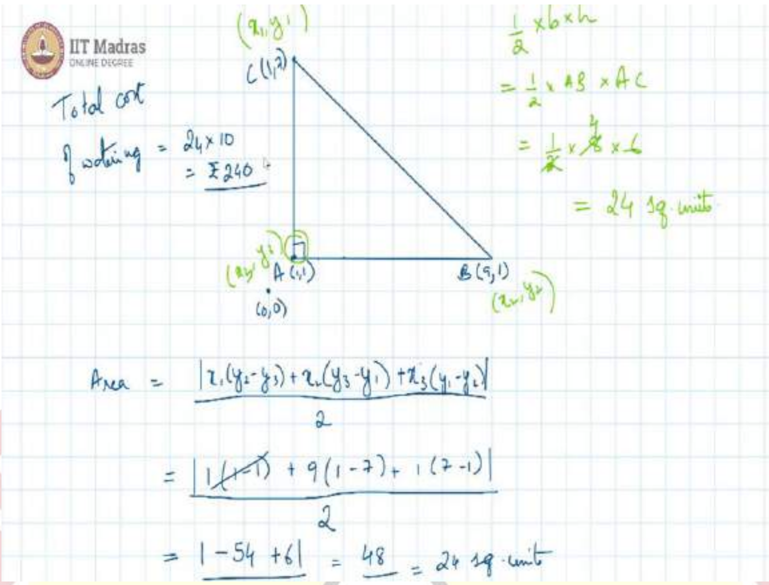






In the second question, the reserved triangular field ABC, whose coordinates are given. And if watering costs rupees 10 per unit square, so they are giving the cost of watering the field, and it is so and so amount per unit square that is area, how much would you have to pay for the whole field? So, we would like to find out the area of the field. And then if the fencing wire around the field costs rupees 5 per unit, how much should he have to pay for 3 rounds of fencing around this field that is find the perimeter, so find the area and perimeter of this particular field.

(Refer Slide Time 01:08)



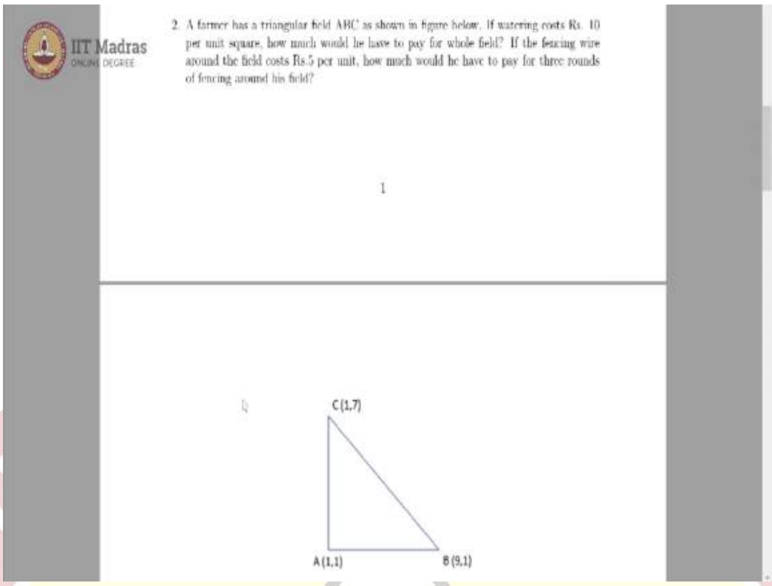
So, if we consider this to be our origin, the triangle is made up of these points, $(1, 1)$, this will be $(9, 1)$ and this is $(1, 7)$. So, this is our triangle, this is A, this is B, and this is C and you can see that AC is completely vertical, its x coordinate remains the same, it is 1, and AB is completely horizontal, its y coordinate remains the same, which is 1.

So, this is a right angled triangle. Now, we could use the area of triangle formula, the area will be $\frac{|x_1(y_2-y_3) + x_2(y_3-y_1) + x_3(y_1-y_2)|}{2}$, which in this case is going to be, you can take any of these points to be $x_1\ y_1$ and the others to, the next one to be $x_2\ y_2$ and x_1 to be $x_3\ y_3$. The, how you choose $x_1\ y_1\ \ x_2\ y_2$ and $x_3\ y_3$ does not matter, the order is what is important. Applying this formula, we get our x_1, y_2, y_3 is 1.

So, $\frac{|1(1-1) + 9(1-7) + 1(7-1)|}{2} = \frac{|-48|}{2} = \frac{48}{2}$, and that is 24 square unit. However, the same problem could be approached in a slightly different way which is, if I observe that this is a right angle triangle, I could just do half into base into height. And here the base would be the length AB, that is half into AB for which the height would be the length AC.

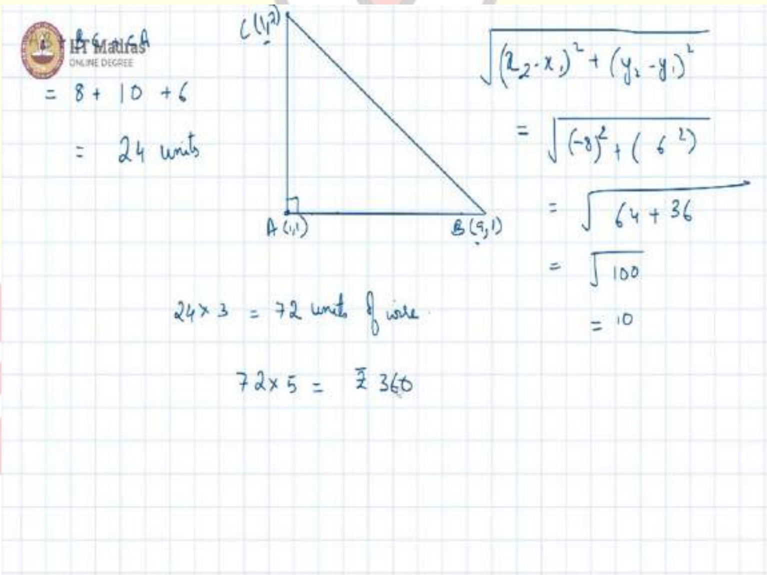
And now since AB is horizontal, you can directly take the length to be $x_2 - x_3$ which is the difference in the x coordinates, so $\frac{1}{2} \times 8 \times 6$ we have 24 square unit. So, this worked out because our triangle is a right-angled triangle. So now the cost of watering is supposed to be the area into cost of watering per square unit which is 10 rupees, so total cost of watering is equal to 24 into 10 that is rupees 240.

(Refer Slide Time 05:18)



For the second part of the question, we require the perimeter of this triangle because fencing is done along the perimeter, and they have to do 3 rounds of fencing at the rate of rupees 5 per unit. So, we first find the perimeter.

(Refer Slide Time 05:39)



Perimeter would simply be AB plus BC plus CA, which is AB is clear it is 8 units, CA is also clear which is 6 units, BC needs to be figured out and BC we find out using the Euclidean distance, that is the $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$ that is the square of the difference in x coordinates plus the square of the difference in y coordinates, the whole under root.

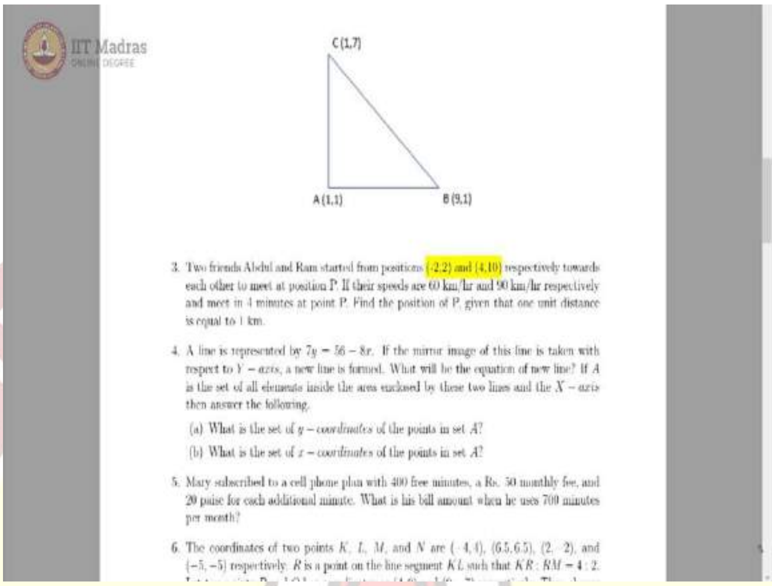
So, this gives us $\sqrt{(-8)^2 + 6^2} = \sqrt{64 + 36} = \sqrt{100} = 10$. So, we have 10. So, this quantity is 10 and thus our perimeter is also 24 units and we need wiring for fencing around 3 rounds. So, we will require 24 into 3 is equal to 72 unit of wire and then each unit has been fixed a price of 5 rupees. So, we have 72 into 5 is rupees 360 is the cost of fencing.

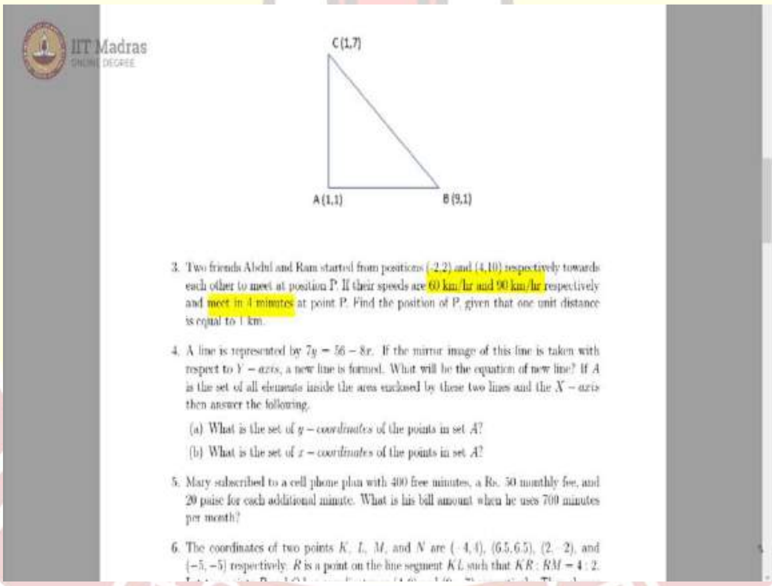
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Indian Institute of Technology, Madras
Week 02
Tutorial 03

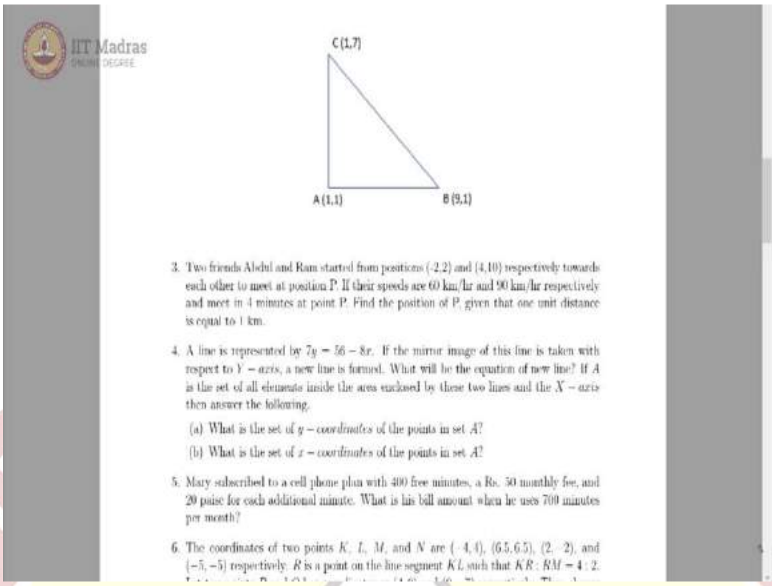
(Refer Slide Time 00:17)

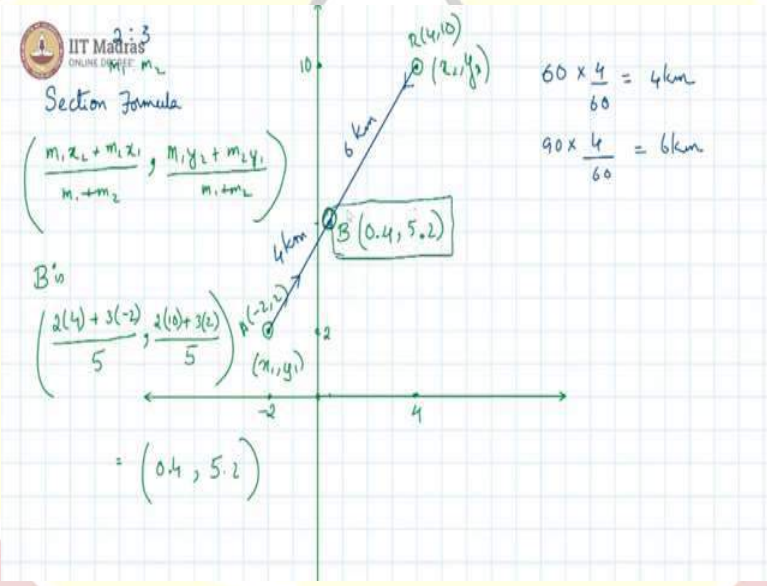




In the third question, the two friends positioned at these two locations and both of them go to a position P. The speeds are given, and the time of their meeting is given, then what should be this position P given that 1unit distance is equal to 1 kilometre. So first let us look at their positions.

(Refer Slide Time 00:55)





So, this point is the origin, and now among the 2 friends, 1 Abdul is at (-2,2), so this is -2 here, and this is 2 here. So, Abdul is here, A (-2,2). And we have the other one Ram at (4,10), which is this is 4 on the x axis and this is 10 on the y axis, so Ram is here (4,10). It says they are moving towards each other, so this is a path they take, where Abdul is moving this way and Ram is moving this way.

And what we know about their movement is, Abdul is moving at 60 kmph and Ram is moving at 90 kmph, so Ram is faster and they are meeting in 4 minutes. If 1 unit is a kilometre, we have 60×4/60 because it is 4 minutes and the units are in hours kilometre per hour, so we do 4/60 is equal to 4km.

So, Abdul is moving 4 km, whereas Ram is moving 90×4/60, which is 6 km, so they meet somewhere in this region and we would like to know that point. And that point we can achieve through the section formula; we do not actually need to find the distances. And for applying the section formula, what we need to know is the ratio of how this point cuts the line segment AR. And that ratio we can use it in this way.

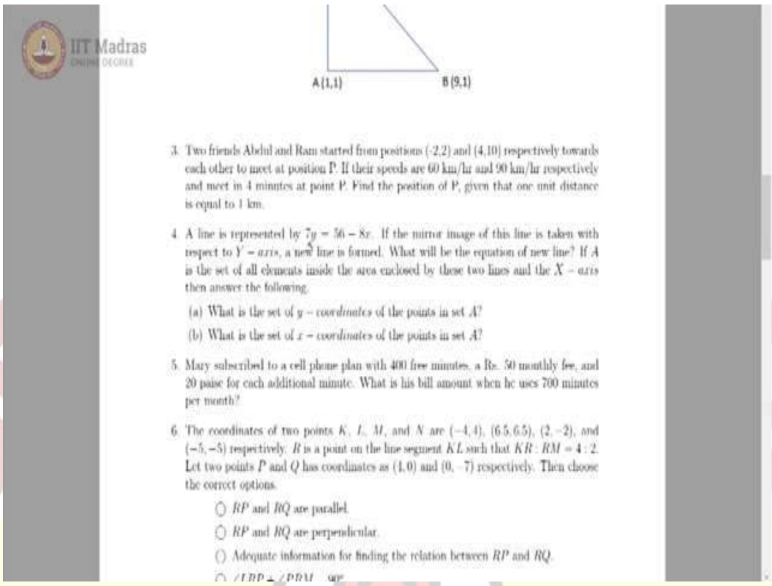
So, we know that this length is supposed to be 4 km and this length is supposed to be 6 km which means the ratio is 4:6 that is 2:3. So, we now apply the section formula, which is $(m_1x_2+ \; m_2x_1)/(m_1+m_1)$. This will be the x coordinate of that point and $(m_1 \; y_2+ \; m_2 \; y_1)/(m_1 + m_1)$ will be the y coordinate of that point. So, let us call this point B, so this is the formula for B, so we get the point B is applying m_1 is, this is the ratio m_1: m_2 and this is (x_1, y_1) and this is (x_2, y_2)

So, we have, m_1x_2 would be 2×4 + m_2x_1 would be 3×(-2) the whole by m_1+m_1 is 5 and (m_1 y_2 would be 2×10 + m_2 $y_1$would be 3×2)/5 again. So that gives us 8-6 = 2, 2/5 is 0.4, and 2×10 = 20, 3×2 = 6 or 26 /5 = 5.2. So, B is (0.4,5.2). We can check with our intuition, this point that we marked out actually has an x coordinate between 0 and 1 and a y coordinate between 5 and 6. So the point we are looking for is (0.4,5.2).

IIT Madras

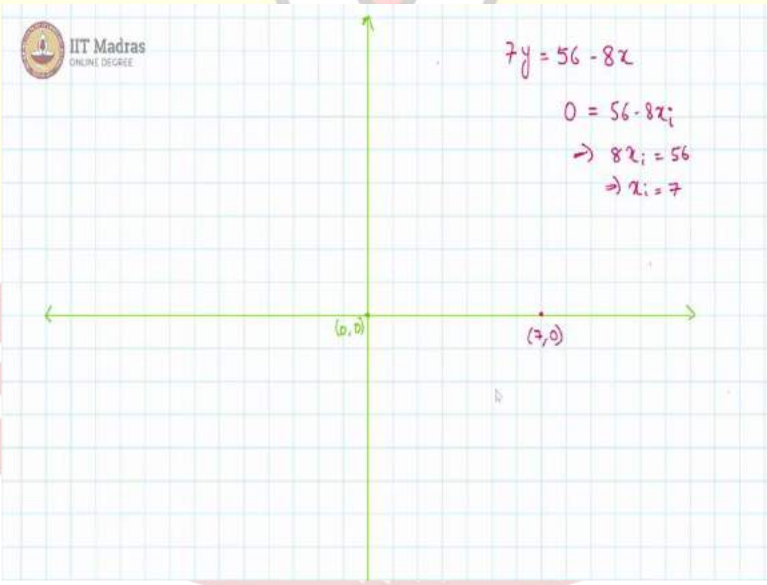
ONLINE DEGREE

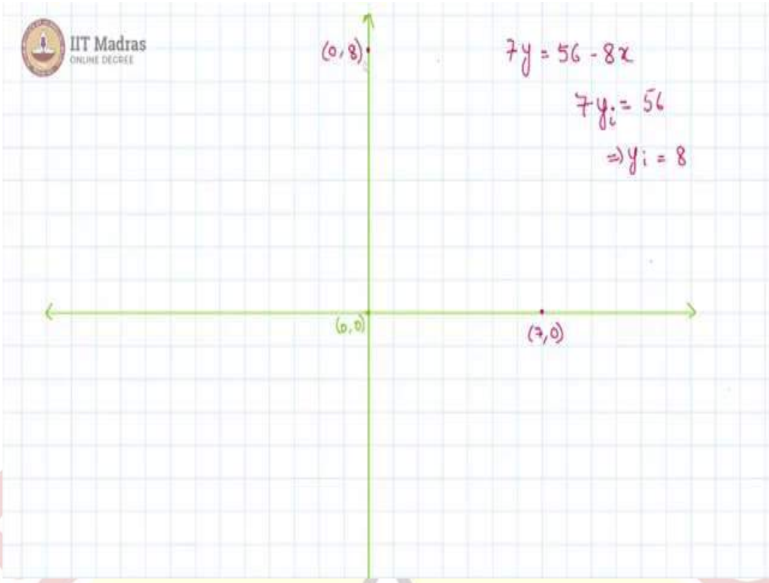
(Refer Slide Time 00:16)



Now, fourth question, there is a line which is represented by 7y = 56 - 8x.

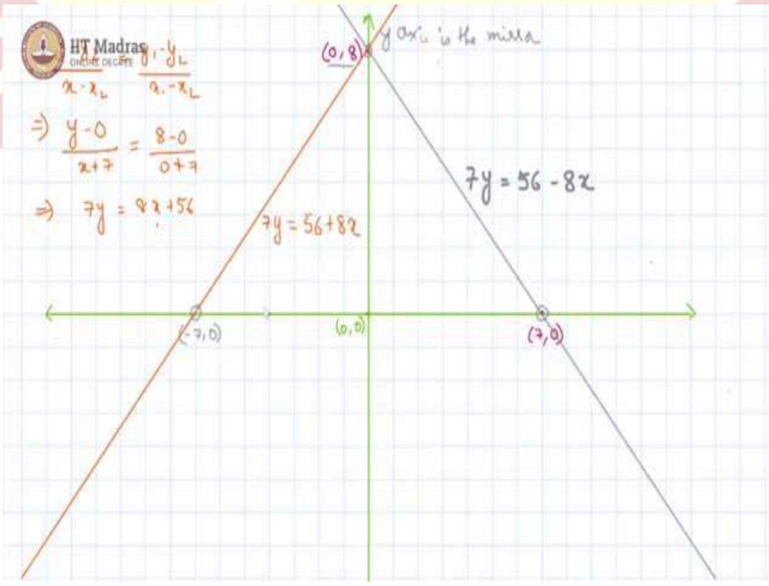
(Refer Slide Time 00:27)





Let us first draw this line, so this is our origin and our line equation is 7y = 56 - 8x. In order to draw this line, in order to find out the curve, we need two points, two points are enough. And the easiest way to find out these two points is to work with the intercepts, that is when this line cuts the X-axis and when it cuts the Y-axis. So when, it is cutting the X-axis, y will be 0, so we just take the Y-coordinate to be 0, and we write 0 = 56 - 8x and to denote that this is the intercept, I am going to call it x_i and that gives us $8x_i$ = 56 and that gives us x_i= 7. So, the x- intercept is 7 which is here. So, (7,0) is one point. And now, for the other point, we take x to be 0 and thus we can say 7y is equal to 56. Again, for the intercept, I am going to use y_i, 56 - 0, therefore y_i, , the y intercept is 8. So, this point here, which is (0,8), this is our y-intercept.

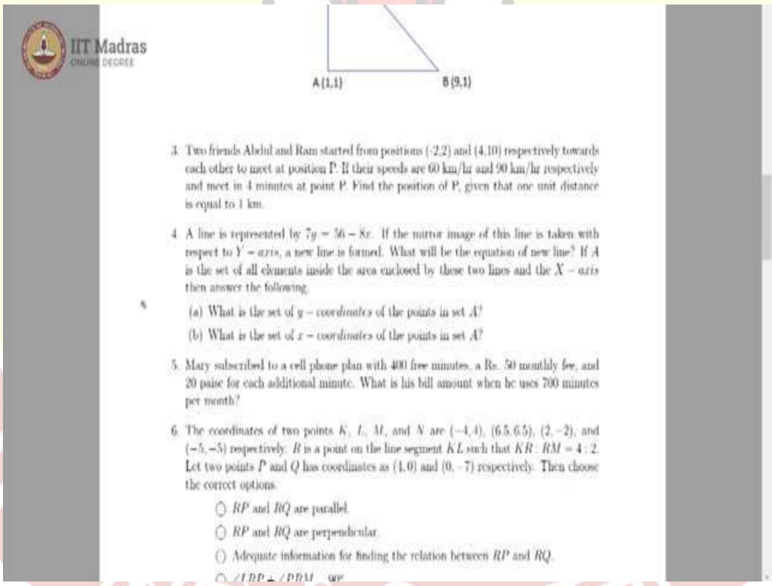
(Refer Slide Time 02:09)

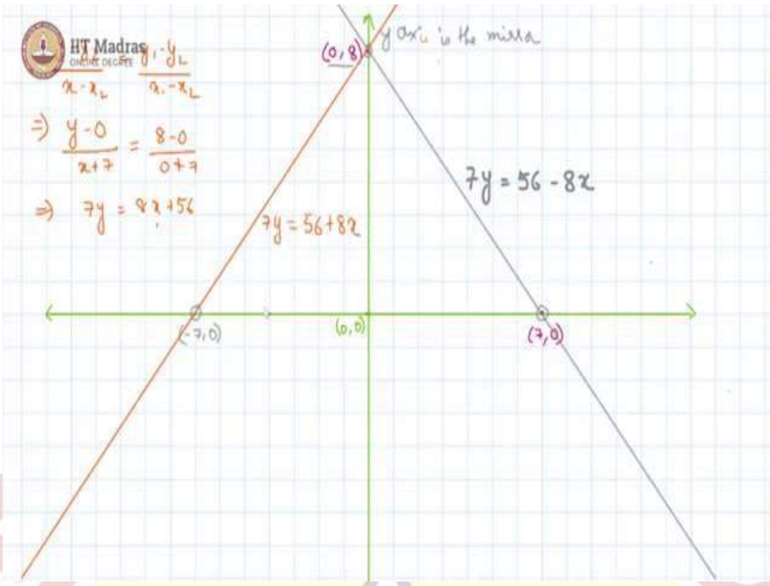


So, this is a straight line, we have been given 7 y = 56 - 8x. It passes through (7,0) and (0,8). Now, for a mirror image, what happens is, and here we are treating the Y-axis as the mirror, so Y-axis is the mirror, you are at the same distance from your mirror as your reflection. So, your reflection will be at the exact distance from the mirror on the opposite side as you, so for example, if we take our (0.7,0) on the other side, which is this point that is (-7,0), that would be the reflection of (7,0) with respect to the Y-axis as the mirror. However, (0,8), since it is already on the Y-axis, its reflection is going to coincide with itself, so this is the other point of the reflection.

And thus, the mirror image for this line is going to be this other line which passes through these two points, (-7,0) and (0,8). For finding the equation of this line, we can use the two point form. And when we apply the values, we get (y – 0) / (x +7) = (8 – 0) / (0 + 7), which gives us 7y = 8 x + 56. So, the mirror image line if you have to write it in the same form as the other one, 7y = 56 + 8 x.

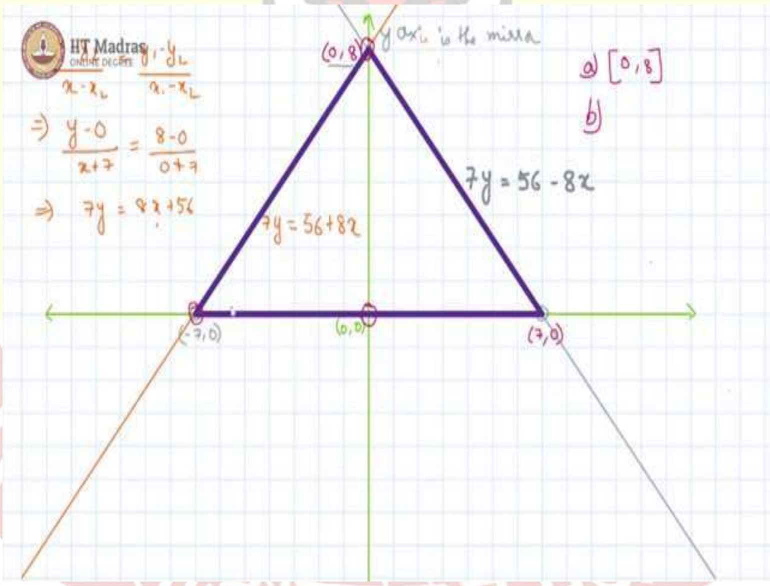
(Refer Slide Time 04:40)

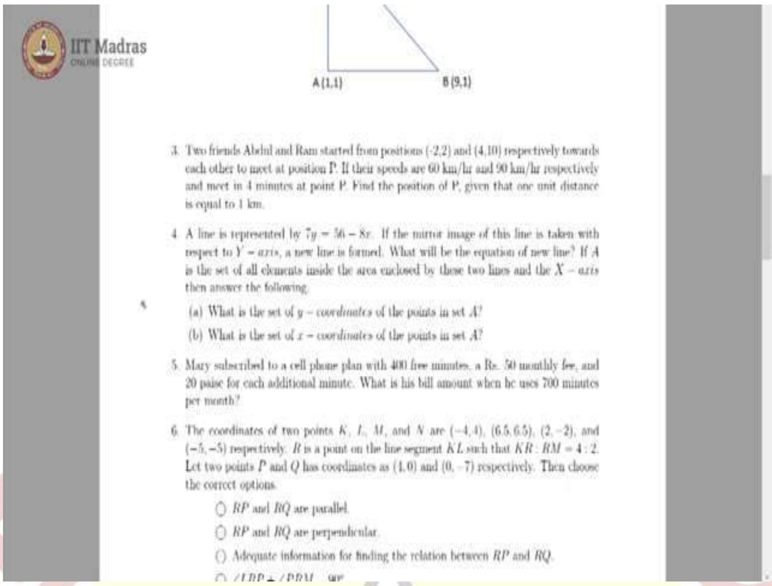




Now, in the next part of the question, they are asking if A is the set of all elements inside the area enclosed by these two lines and the X-axis. So, we are looking at this triangle, and in this triangle, we have being asked what is the set of Y coordinates of the points in set A.

(Refer Slide Time 05:03)





3. Two friends Abdul and Ram started from positions (-2,2) and (4,10) respectively towards each other to meet at position P. If their speeds are 60 km/hr and 90 km/hr respectively and meet in 4 minutes at point P. Find the position of P, given that one unit distance is equal to 1 km.

4. A line is represented by $7y = 56 - 8x$. If the mirror image of this line is taken with respect to $Y - axis$, a new line is formed. What will be the equation of new line? If A is the set of all elements inside the area enclosed by these two lines and the $X - axis$ then answer the following.

   (a) What is the set of $y - coordinates$ of the points in set A?

   (b) What is the set of $x - coordinates$ of the points in set A?

5. Mary subscribed to a cell phone plan with 400 free minutes, a Rs. 50 monthly fee, and 20 paise for each additional minute. What is his bill amount when he uses 700 minutes per month?

6. The coordinates of two points K, L, M, and N are $(-4,4)$, $(6.5,6.5)$, $(2,-2)$, and $(-5,-5)$ respectively. R is a point on the line segment KL such that $KR : RM = 4 : 2$. Let two points P and Q has coordinates as $(1,0)$ and $(0,-7)$ respectively. Then choose the correct options.

   ○ RP and RQ are parallel.

   ○ RP and RQ are perpendicular.

   ○ Adequate information for finding the relation between RP and RQ.

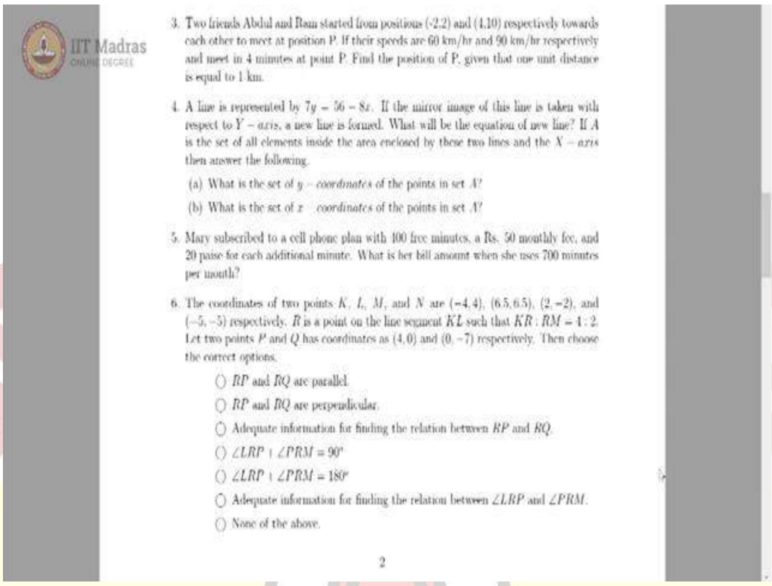
So, all possible Y coordinates in this set. So, every point within this triangle and on the triangle itself count, and as you can clearly see the least Y coordinate here is 0, and the maximum Y coordinate here is 8. So, the set of Y coordinates is going to be the closed interval [0, 8], because we are considering the triangle also to be part of this set, not just the points inside the triangle interior to the triangle, we are considering the triangle also to be part of the set. So, this is the answer for part A. And for part B we have what is the set of X coordinates of the points in set A, and again, we look for the least and the maximum here, the least is -7 and the maximum is 7. And every value in between is there so this would be again the closed interval [-7 ,7].

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Indian Institute of Technology, Madras
Week 02 - Tutorial 05

(Refer Slide Time 00:15)



3. Two friends Abdul and Ram started from positions (-2,2) and (4,10) respectively towards each other to meet at position P. If their speeds are 60 km/hr and 90 km/hr respectively and meet in 4 minutes at point P. Find the position of P, given that one unit distance is equal to 1 km.

4. A line is represented by $7y = 56 - 8x$. If the mirror image of this line is taken with respect to $Y - axis$, a new line is formed. What will be the equation of new line? If A is the set of all elements inside the area enclosed by these two lines and the $X - axis$ then answer the following.
   (a) What is the set of $y - coordinates$ of the points in set A?
   (b) What is the set of $x - coordinates$ of the points in set A?

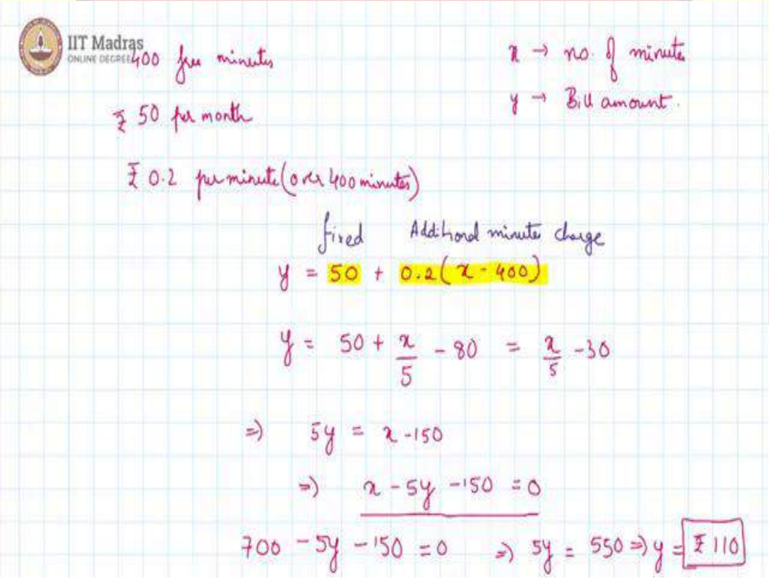
5. Mary subscribed to a cell phone plan with 100 free minutes, a Rs. 50 monthly fee, and 20 paise for each additional minute. What is her bill amount when she uses 700 minutes per month?

6. The coordinates of two points K, L, M, and N are $(-4,4)$, $(6.5,6.5)$, $(2,-2)$, and $(-5,-5)$ respectively. R is a point on the line segment KL such that $KR : RM = 1 : 2$. Let two points P and Q has coordinates as $(4,0)$ and $(0,-7)$ respectively. Then choose the correct options.
   - ○ RP and RQ are parallel.
   - ○ RP and RQ are perpendicular.
   - ○ Adequate information for finding the relation between RP and RQ.
   - ○ $\angle LRP + \angle PRM = 90°$
   - ○ $\angle LRP + \angle PRM = 180°$
   - ○ Adequate information for finding the relation between $\angle LRP$ and $\angle PRM$.
   - ○ None of the above.

2

Now 5th problem, Mary has subscribed to a cell phone plan with 400 free minutes, a 50 rupee monthly fee and 20 paisa for every additional minute over 400. And the question is, what is her bill amount if she uses 700 minutes?

(Refer Slide Time 00:42)



So let us put down our variables here. So there is 400 free minutes and there is a 50 rupee charge per month and we have 20 paisa that is 0.2 rupees per minute over 400 minutes. Now, our independent variable is the number of minutes, the bill is dependent on the number of minutes, so our x variable is number of minutes and the y variable is bill amount. And what we know is for every month the bill amount will always have a 50 rupee charge, and on top of that you are being charged 0.2 for every minute over 400, which means if x is the total number of minutes, then $(x – 400) (0.2)$ will be the charge for the additional minutes.

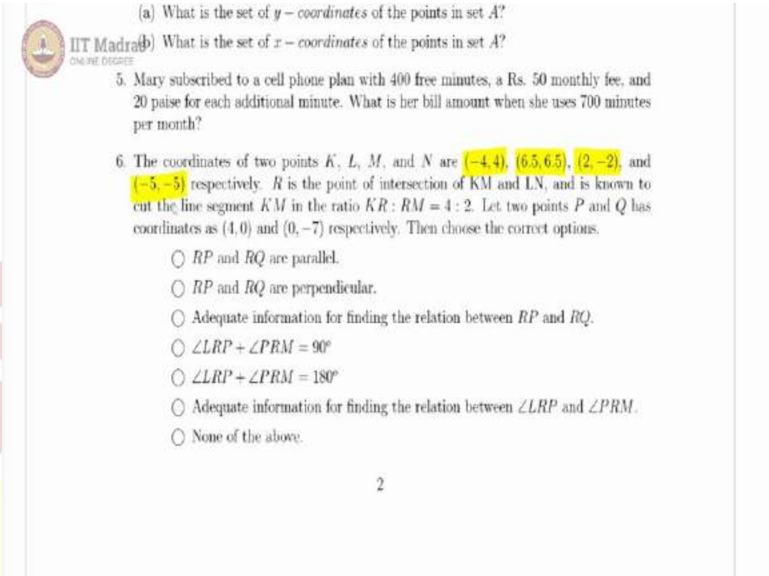
This is the fixed charge whereas this is the additional minutes charge, so we get a linear equation which is $y = 50 + x/5$(because 0.2 is 1/5) - 80 which is then $(x/5) - 30$. If we simplify it further, we get $5y = x - 150 \Rightarrow x - 5y - 150 = 0$. This is the equation that relates our bill amount to the number of minutes. So, Mary is using 700 minutes per month and we need the bill amount for that. So, if we substitute $x = 700$, we get, $700 - 5y - 150 = 0$, this gives us $5y = 550$ which implies $y =$ Rs. 110. This is the bill amount for Mary.

IIT Madras

ONLINE DEGREE

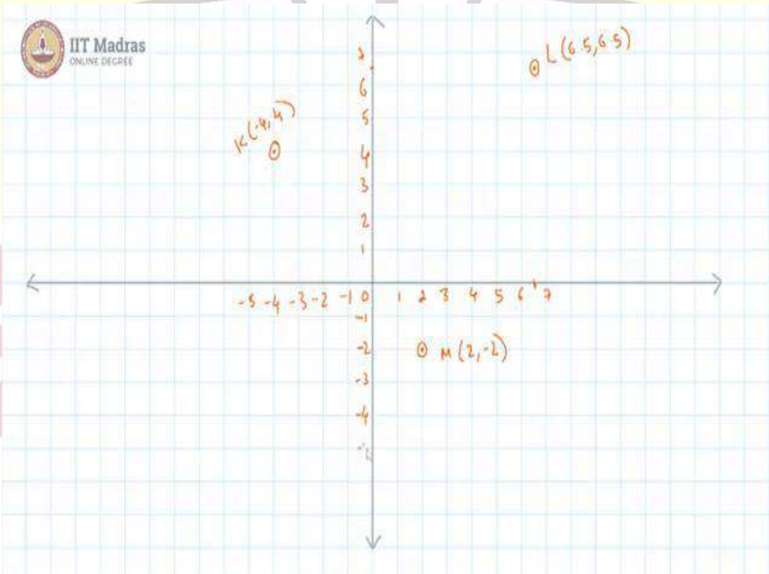
Mathematics for Data Science 1
Indian Institute of Technology, Madras
Week 02 - Tutorial 06

(Refer Slide Time 00:17)



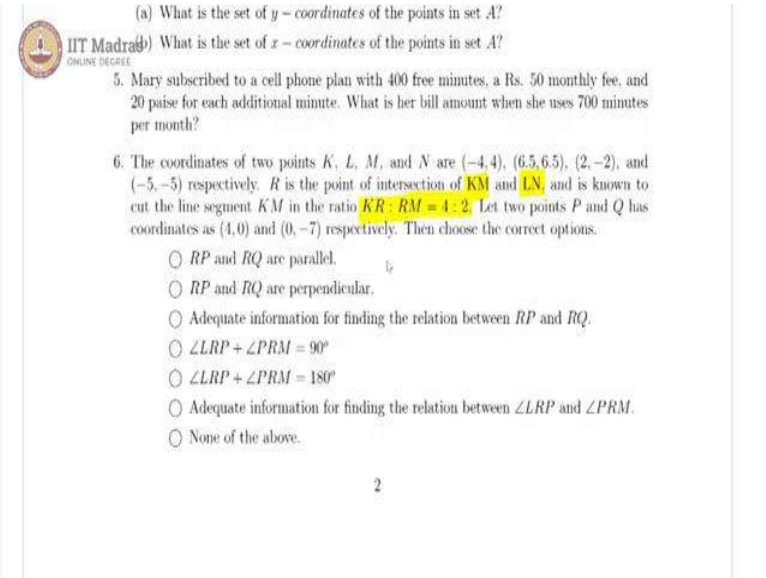
For our 6th problem we have these 4 points given to us. Let us first plot them out on a graph.

(Refer Slide Time 00:32)



And this will be 0, we have a (-4, 4), so this is -1, this is -2, this is -3, this is -4, this is 1, 2, 3 and 4, so this point here is our K (-4, 4). And then we have (6.5, 6.5), this is 1, 2, 3, 4, 5, 6, 7, this here is 6.5 and 5, 6 and 7, this here is 6.5, here we are with L (6.5, 6.5), then we have a (2, -2), -1, this is -2. So this point here is our M (2, -2). And lastly, we have (-5, -5), -4 and -5, this is our point.

(Refer Slide Time 01:48)



(a) What is the set of $y-coordinates$ of the points in set A?

(b) What is the set of $x-coordinates$ of the points in set A?

5. Mary subscribed to a cell phone plan with 400 free minutes, a Rs. 50 monthly fee, and 20 paise for each additional minute. What is her bill amount when she uses 700 minutes per month?

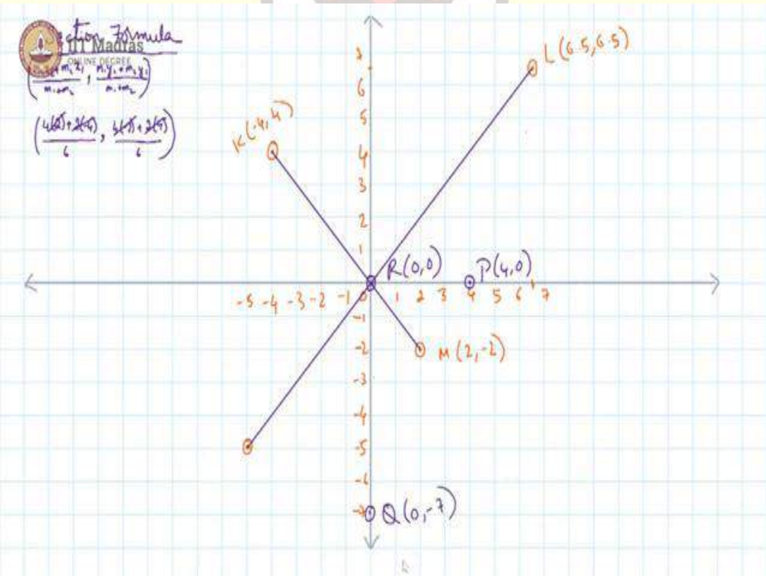
6. The coordinates of two points K, L, M, and N are $(-4,4)$, $(6.5,6.5)$, $(2,-2)$, and $(-5,-5)$ respectively. R is the point of intersection of KM and LN, and is known to cut the line segment KM in the ratio $KR:RM = 4:2$. Let two points P and Q has coordinates as $(4,0)$ and $(0,-7)$ respectively. Then choose the correct options.

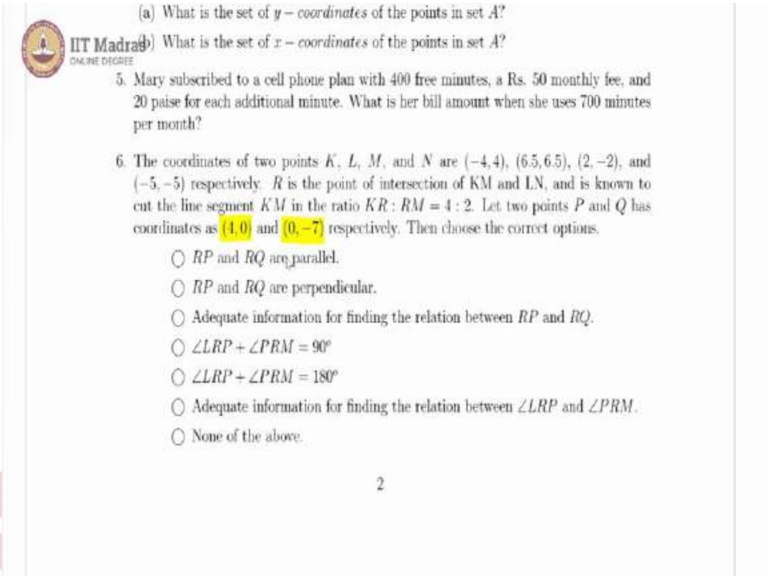
- RP and RQ are parallel.
- RP and RQ are perpendicular.
- Adequate information for finding the relation between RP and RQ.
- $\angle LRP + \angle PRM = 90°$
- $\angle LRP + \angle PRM = 180°$
- Adequate information for finding the relation between $\angle LRP$ and $\angle PRM$.
- None of the above.

2

Now we are told that R is the point of intersection of KM and LN, and it is known to cut the line segment KM in this ratio, 4 is to 2 ratio, so let us identify R.

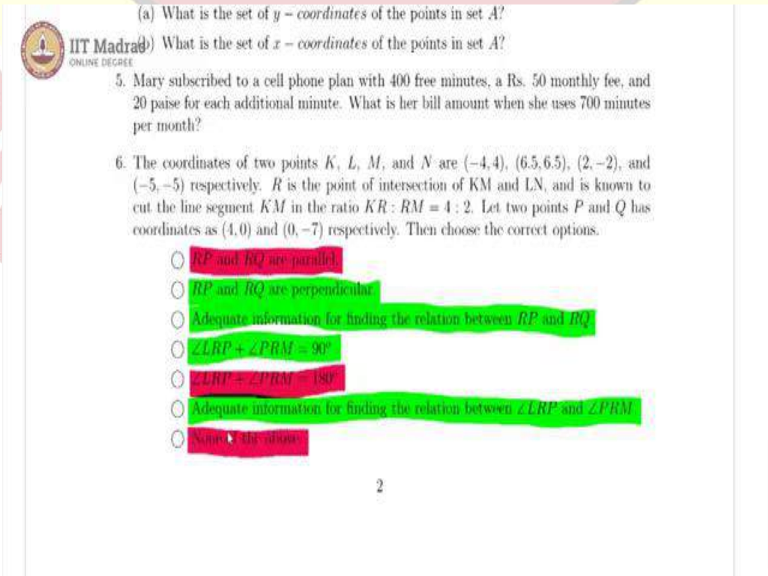
(Refer Slide Time 02:04)

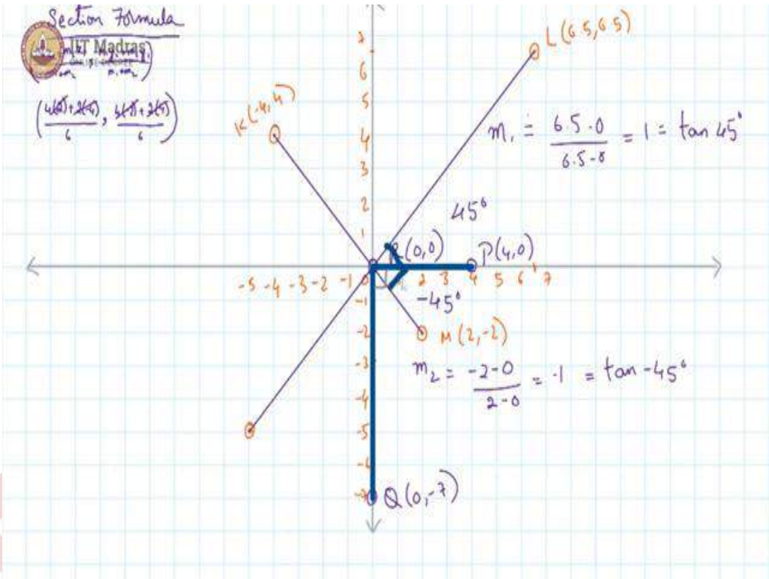




So from our diagram, it appears to be the origin. Lets verify this, so we need this to be in the ratio of 4:2. So when we use the section formula, which is the coordinates of a point cutting a line segment in a ratio, $m_1 : m_2$ would be this, $(m_1x_2 + m_2x_1) / (m_1 + m_2)$. And then we have $(m_1y_2 + m_2y_1) / (m_1 + m_2)$. So in this context, R is going to be ((4(2) + 2 (-4))/ 6 , (4(-2) + 2(4))/6). And these 2 cancel out because it is 8 - 8, these two also cancel because -8 +8. So it is true, R the point is the origin. Moving on then, we have two other points, P and Q given to be (4, 0), (0, -7), so these are on the axis. So this point here is P(4, 0), and this is -6, this is -7 so this point here would become Q, which is (0, -7).

(Refer Slide Time 03:58)





Lets look at the options, RP and RQ are parallel, this is one option, lets verify. Now clearly, this is 90°, PQ, PR is perpendicular to RQ and not parallel. So this is definitely wrong and this is definitely right. Is there adequate information for finding the relation between RP and RQ? Yes, we have just found the relation, so there has been adequate information.

Now let us look at ∠LRP +∠PRM. So we are interested in this angle plus ∠PRM. So this sum is the total ∠LRM, so we need to know what is the angle between LR and RM. So let us look at the slope of LR. So this slope if I call it m_1, this is equal to (6.5 - 0) /(6.5 - 0), which is 1, which is basically tan 45°, so this angle here it is 45°.

And now let us look at this angle here, which is PRM. Then, if we look at the slope here, which is m_2 that is (-2 - 0)/(2 - 0), which is -1, which is equal to tan -45°, therefore this angle here is -45° because we are going clockwise from the horizontal. So in sum, we know that ∠LRM is 45°+ 45°, leading us to see that this is 90°.

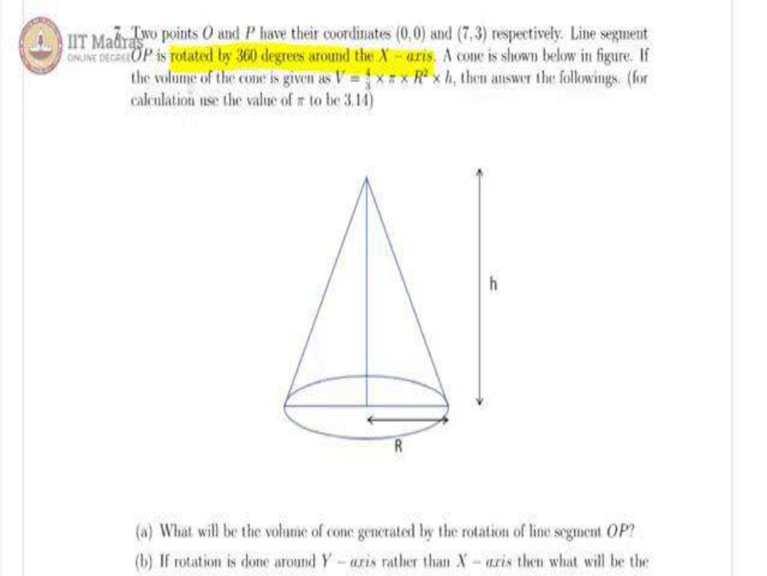
So this is true, which means the following statement is false, so this is false. And here we have again adequate information for finding the relation between LRP and PRM. Clearly, we have four options correct, so none of the above is not correct.

IIT Madras

ONLINE DEGREE

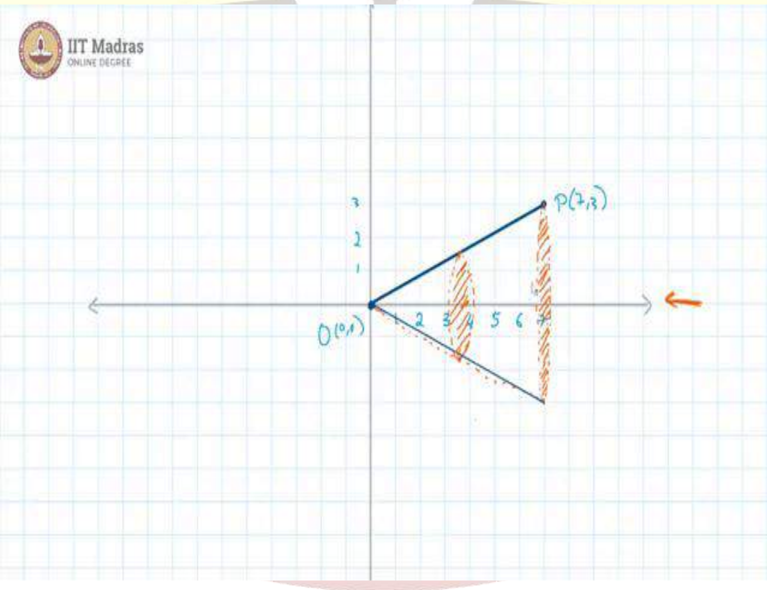
Mathematics for Data Science 1
Indian Institute of Technology, Madras
Week 02 - Tutorial 07

(Refer Slide Time 00:15)



In the 7th question we have two points, one is the origin O, and some other point P (7, 3). And this line segment OP is being rotated. So first, lets mark out these points.

(Refer Slide Time 00:32)

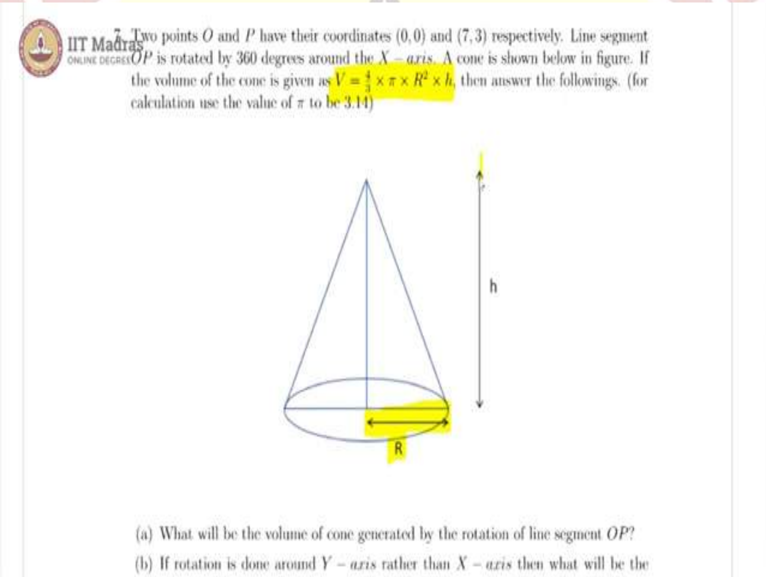


So first lets mark out these points, we have this is the origin and this is 1, 2, 3, 4, 5, 6, 7, this is 1, 2, 3. So our point P is here, this is P (7, 3) and this is the origin of course O. And we have this line segment OP given to us. Now, OP is being rotated by 360° about the x axis, lets see what that means. So every point on OP is going around the x axis in a circle, that is what rotation is, rotation is a combined circular motion of many particles, here we have this point let us take P, P goes around the x axis reaching this bottom point, then it circles back to itself.

So you would see the circle if you looked at it from the right. From the screen's perspective, this is what it will look like. And this is the case with every point on this line.

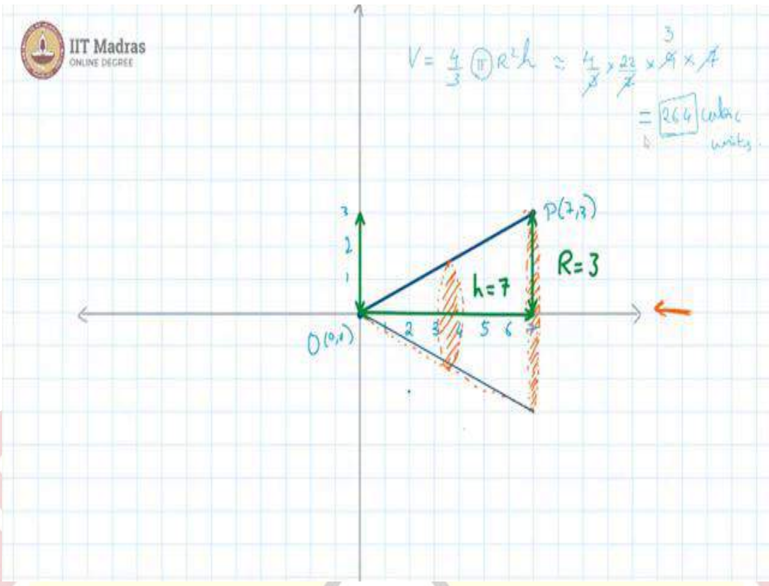
Suppose I took this point, this is just oppositely going to go till here in this circle and return back. So, every point is doing this circle, which means on this side we actually have the mirror image of OP with respect to the x axis which looks something like this. So, we have these circles being formed due to the rotation and as you can see, the final shape it appears to be a cone that is what has happened. Take a line segment and you rotate it about some central axis, you obtain a cone about that central axis.

(Refer Slide Time 02:53)



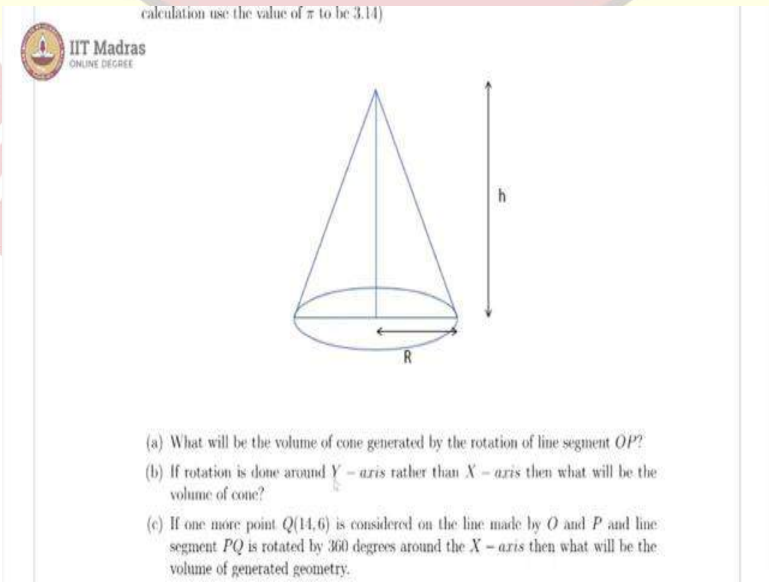
And they have given us the volume of a cone, volume of a cone is $(4/3)\pi R^2 h$, where R is the radius of the base circle and h is the height of the cone.

(Refer Slide Time 03:16)



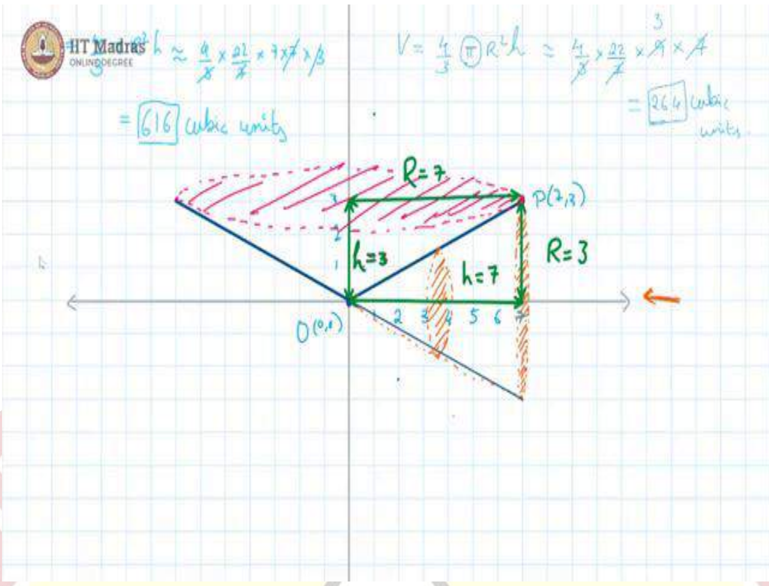
So in the cone we have obtained the radius, base radius is this quantity, which is R = 3 because it is a y coordinate of the point P that is a distance of point P from the x axis. And likewise, if you observe the height of this cone, that quantity h = 7 because the x coordinate of point. In this way, we can obtain the volume of our cone using the formula that is given V= $(4/3)\pi R^2h$. We are going to approximate pi to be 3.14 or 22 by 7. So, this is roughly equal to (4/3)(22/7)(9)(7) so 7 and 7 cancels of, 3 and 9 gives us 3. So we get 264 cubic units, so this is our volume of the cone.

(Refer Slide Time 04:54)



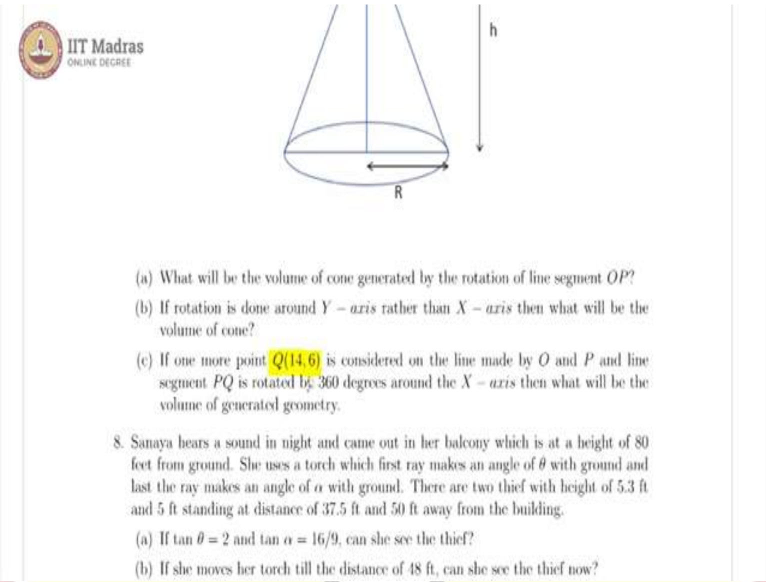
Now in the second part of this question, it is being said that the rotation is done around the y axis instead of the x axis, so what will this look like?

(Refer Slide Time 05:08)



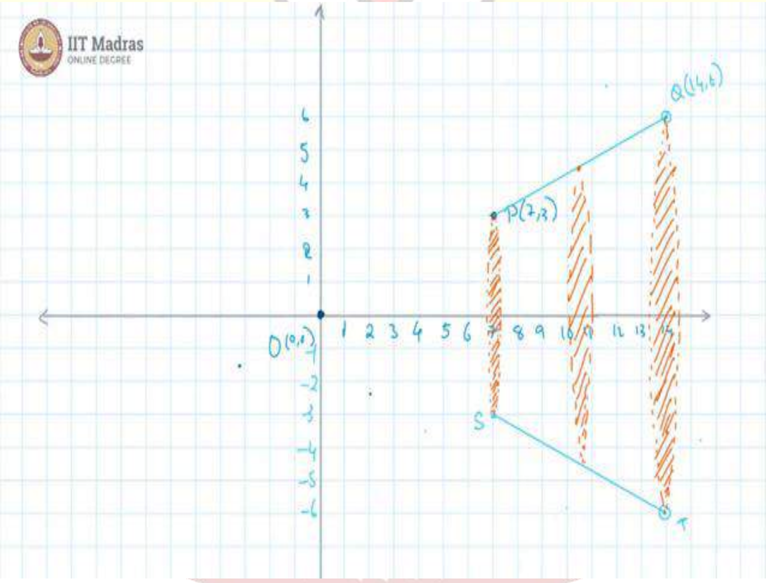
So, this OP is going to have a mirror image about the y axis which is going to look like this. So that means, our point P is going around in a circle to reach this opposite point here and it is coming back to itself. So this would become the base circle for our new cone which is obtained by rotation about the y axis. Now as you can see, this value is already the height so height is 3 now, whereas the radius is basically 7. So, our height is 3 and radius is 7 so these values have changed. So, if we call this quantity, this volume to be V_2, V_2 is $(4/3)\pi R^2h$, which will be roughly equal to (4/3)(22/7)(7)(7)(3). So, 3 and 7 cancel off here and we get this is equal to 616 cubic units. So, this is the volume if OP is rotated about the y axis 360°, so that cone's volume is 616 cubic units.

(Refer Slide Time 06:58)



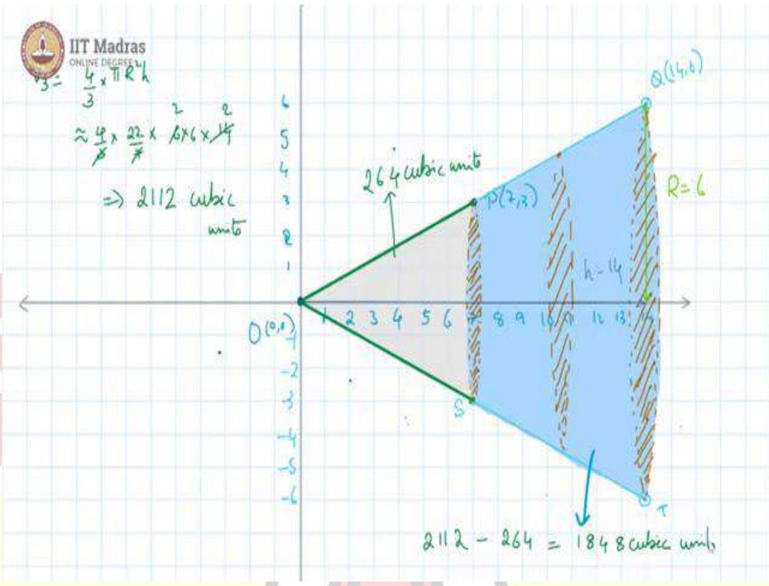
This problem gets progressively more complex, we are now adding the new point (14, 6) which is along the line segment OP, it is on the extension of OP, and then PQ is rotated about x axis by 360°

(Refer Slide Time 07:25)



So, lets see what is happening here, so our point (14, 6) Q, (14, 6) is here, which gives us PQ as this line segment. And now, they are saying that PQ is being rotated about the x axis which will result in the mirror image in this way. This is -1, -2, -3, -4, -5, and -6. So, we are here now, I think we can call this point T and this point is S. So, we have ST in this way again, so what we see here the rotated geometry. So, for reference we are going to take one more point here, which kind of moves around. So, what we are seeing here this is what is called the frustum of a cone. This is a cut-off portion of a larger cone which would be QO rotated about the x axis.

(Refer Slide Time 08:56)



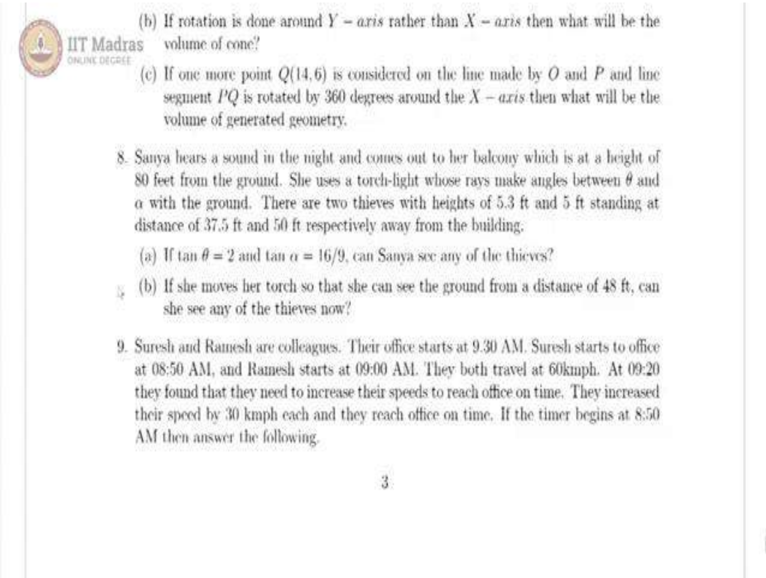
So, thus as you can see, the volume we require is the frustum of the cone which is this region and this volume is the result of subtracting this volume from the total cone. So, we already know this volume OP as that cone's volume to be 264 cubic units. So, the blue shaded region that would be the volume of the large cone that is of OQ rotating about x axis and that we can calculate as $V_3 = (4/3)\pi R^2 h$, where this is approximately equal to (4/3)(22/7)(6)(6)(14). So, 3 one's 3 two's, 7 one's 7 two's, so we have 2,112 cubic units. So the volume we require is going to be 2,112 – 264 = 1,848 cubic units.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 02
Tutorial 01

(Refer Slide Time 00:14)


(b) If rotation is done around $Y-axis$ rather than $X-axis$ then what will be the volume of cone?

(c) If one more point $Q(14,6)$ is considered on the line made by O and P and line segment PQ is rotated by 360 degrees around the $X-axis$ then what will be the volume of generated geometry.

8. Sanya hears a sound in the night and comes out to her balcony which is at a height of 80 feet from the ground. She uses a torch-light whose rays make angles between θ and α with the ground. There are two thieves with heights of 5.3 ft and 5 ft standing at distance of 37.5 ft and 50 ft respectively away from the building.

(a) If $\tan\theta = 2$ and $\tan\alpha = 16/9$, can Sanya see any of the thieves?

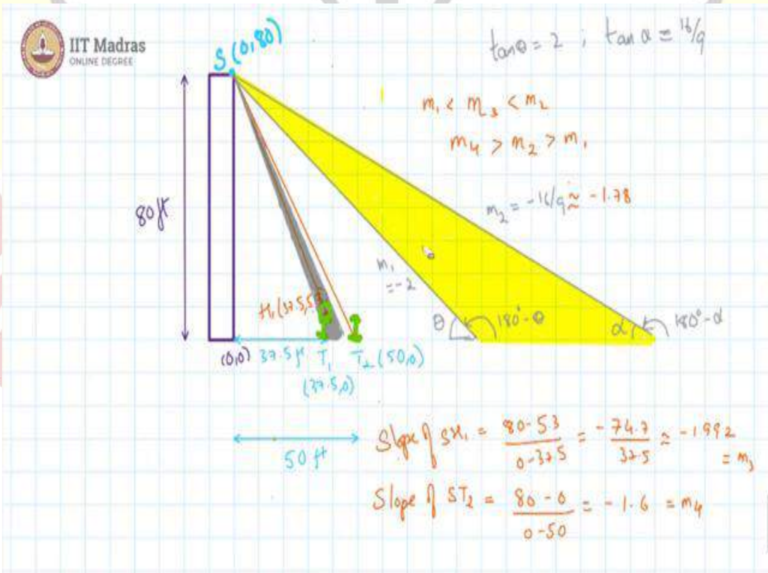
(b) If she moves her torch so that she can see the ground from a distance of 48 ft, can she see any of the thieves now?

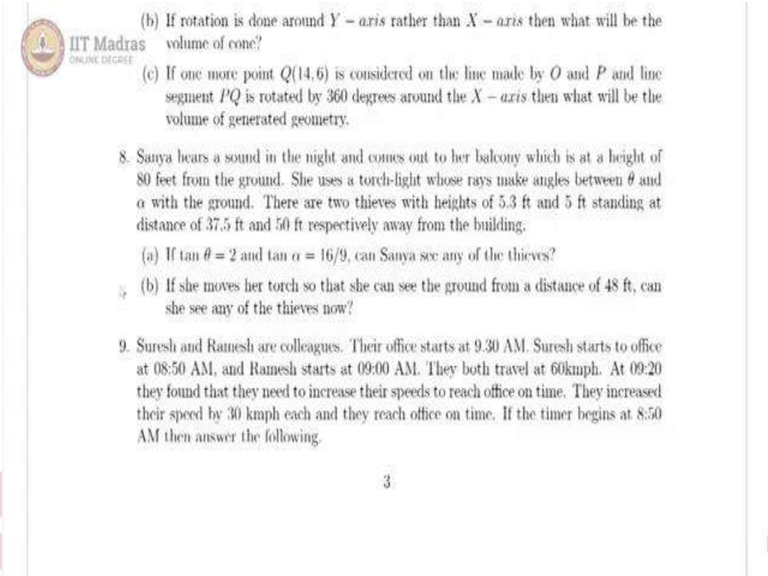
9. Suresh and Ramesh are colleagues. Their office starts at 9.30 AM. Suresh starts to office at 08:50 AM, and Ramesh starts at 09:00 AM. They both travel at 60kmph. At 09:20 they found that they need to increase their speeds to reach office on time. They increased their speed by 30 kmph each and they reach office on time. If the timer begins at 8:50 AM then answer the following.

3

The 8th problem is pretty interesting. So, we have Sania who hears a sound in a night, and she comes out to her balcony, which is at a height of 80 feet from the ground.

(Refer Slide Time 00:29)





(b) If rotation is done around $Y-axis$ rather than $X-axis$ then what will be the volume of cone?

(c) If one more point $Q(14,6)$ is considered on the line made by O and P and line segment PQ is rotated by 360 degrees around the $X-axis$ then what will be the volume of generated geometry.

8. Sanya hears a sound in the night and comes out to her balcony which is at a height of 80 feet from the ground. She uses a torch-light whose rays make angles between θ and α with the ground. There are two thieves with heights of 5.3 ft and 5 ft standing at distance of 37.5 ft and 50 ft respectively away from the building.

(a) If $\tan\theta = 2$ and $\tan\alpha = 16/9$, can Sanya see any of the thieves?

(b) If she moves her torch so that she can see the ground from a distance of 48 ft, can she see any of the thieves now?

9. Suresh and Ramesh are colleagues. Their office starts at 9.30 AM. Suresh starts to office at 08:50 AM, and Ramesh starts at 09:00 AM. They both travel at 60kmph. At 09:20 they found that they need to increase their speeds to reach office on time. They increased their speed by 30 kmph each and they reach office on time. If the timer begins at 8:50 AM then answer the following.

3

So, let this be our tower, which has a height 80 feet. So, if we take this point to be origin (0, 0), Sania is here, which would be (0, 80). And she uses a torch light which makes angles θ and α with the ground, so the rays from the torch light make angles between these two.

So, this angle here, this is θ and this angle here it is α . And the two thieves, their heights are given and they are standing at these distances from the buildings. So, thief T_1 is somewhere here and T_2 is here, what is given to us is this distance is 37.5. So, T_1 is (37.5, 0), and this distance is 50 feet, so this is 37.5 feet, this is 50 feet. So, T_2 will be the point (50, 0). And we are also given to understand that T_1 is standing at a certain height, T_2 is standing at a certain height, which are roughly the same; one is 5 feet, the other is 5.3 feet.

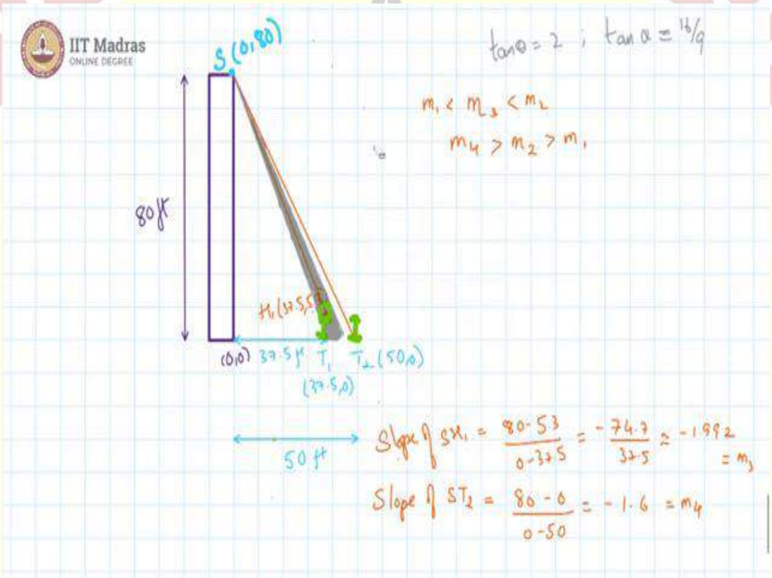
In our diagram, we have drawn the rays of light as though they are passing away from the 2 thieves, however that we need to find out. So, if tan θ is 2, and tan α is 16/9, can Sania see any of the thieves ? so it is given to us that $\tan\theta = 2$ whereas, $\tan\alpha = 16/9$.

So, that means we can find the slope of this line, which is the lowest ray of the torch, and this line which is the farthest ray from the torch, and these slopes would be $m_1 = -2$ and the minus is because the standard angle here, which is angle from the posture x axis is actually $180^o - \theta$. As you can see, it is clearly a line with the negative slope.

Likewise, this also is $180^o - \alpha$, thus this slope $m_2 = -16/9$. In our diagram, we have drawn it as though the 2 thieves are safe. But this is only a rough schematic diagram, we did not draw θ and α accurately. What we need to do now is to check if the line from Sania to the head of thief 1 or the line from Sanya to the foot of thief 2. If these 2 lines have slopes between m_1 and m_2, then the 2 thieves are likely to be seen. So, we need to calculate these slopes, let us call the head of thief 1 as H_1 and that point will be (37.5, 5.3). So, slope of SH_1 = (80-5.3)/ (0-37.5), which is -(74.7 / 37.5), which is roughly -1.992.

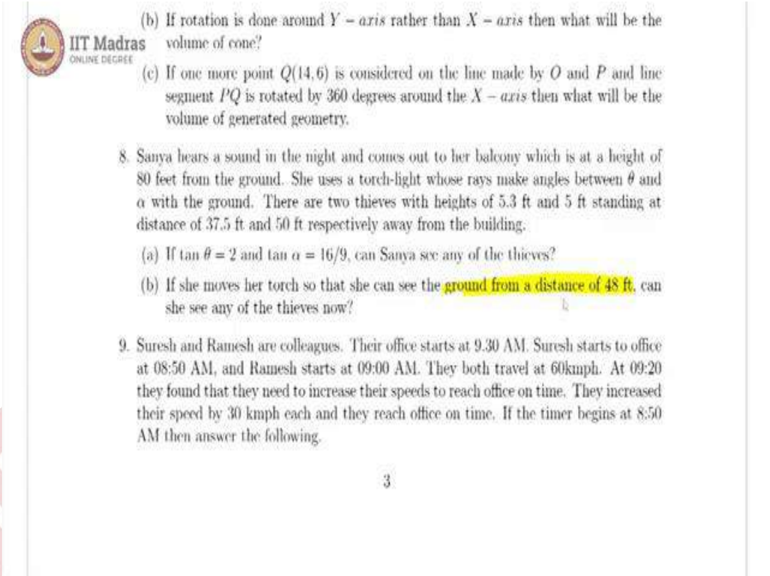
And slope of ST_2, which is to the foot of thief 2 is (80 - 0) / (0 - 15), which is equal to -1.6. So, I want to call this m_3 and this is m_4. And here m_2 is roughly equal to -1.78. So, clearly m_3 is greater than m_1 and lesser than m_2, but m_4 is greater than m_2 and also greater than m_1 which means m_4 that is the foot of thief 2 is not visible to Sania, the actual light cone looks something like this. So, thus we can say the head of thief 1 is visible whereas thief 2 is not visible.

(Refer Slide Time 7:19)



And this light cone that we have drawn earlier it is wrong.

(Refer Slide Time 7:28)


IIT Madras
ONLINE DEGREE

(b) If rotation is done around $Y-axis$ rather than $X-axis$ then what will be the volume of cone?

(c) If one more point $Q(14,6)$ is considered on the line made by O and P and line segment PQ is rotated by 360 degrees around the $X-axis$ then what will be the volume of generated geometry.

8. Sanya hears a sound in the night and comes out to her balcony which is at a height of 80 feet from the ground. She uses a torch-light whose rays make angles between θ and α with the ground. There are two thieves with heights of 5.3 ft and 5 ft standing at distance of 37.5 ft and 50 ft respectively away from the building.

(a) If $\tan\theta = 2$ and $\tan\alpha = 16/9$, can Sanya see any of the thieves?

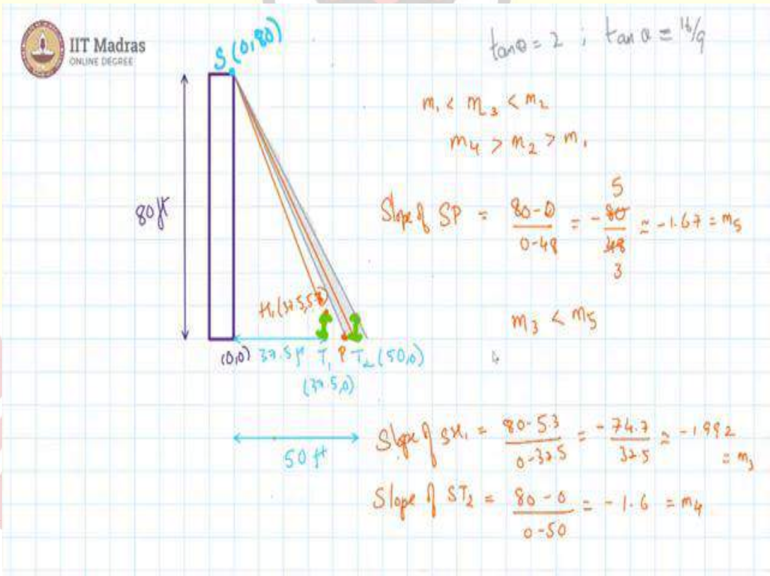
(b) If she moves her torch so that she can see the ground from a distance of 48 ft, can she see any of the thieves now?

9. Suresh and Ramesh are colleagues. Their office starts at 9.30 AM. Suresh starts to office at 08:50 AM, and Ramesh starts at 09:00 AM. They both travel at 60kmph. At 09:20 they found that they need to increase their speeds to reach office on time. They increased their speed by 30 kmph each and they reach office on time. If the timer begins at 8:50 AM then answer the following.

3

Now, she moves her torch so that she can see the ground from a distance of 48 feet. Can she see thieves or not?

(Refer Slide Time: 07:32)



That would mean she is able to see from some point here to some point beyond. From the diagram, it is pretty clear that thief 2 is going to be visible, we do not know if thief 1 will be visible though we have to check for thief 1's head. So, this point that we are talking about, which is let us call it P gives us a slope with S as SP, the slope is equal to (80 - 0) / (0 - 48) because point P is basically (48, 0).

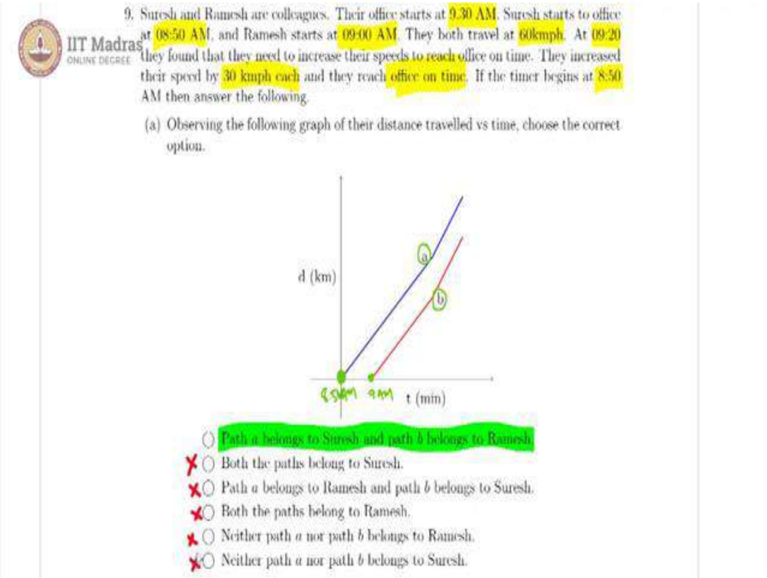
So, that gives us - (80 / 48) which is divisible by 16, both of them are divisible by 16. This would be 5 and this would be 3, so this is roughly -1.67. That would give us m_3 is let us call this now m_5. m_3 is lesser than m_5. And that means the head of thief 1 is not visible now, but thief 2 is visible.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week-02 Tutorial-09

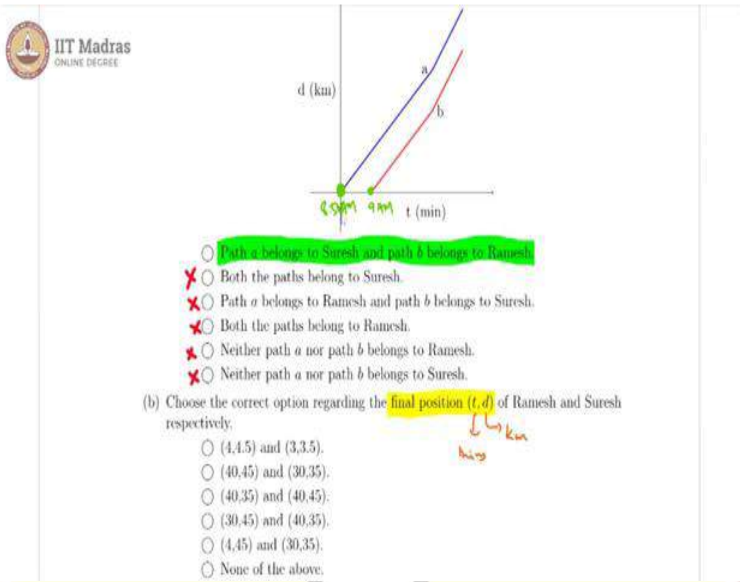
(Refer Slide Time: 0:12)



For our 9 th problem, we have 2 colleagues Ramesh and Suresh, and their office starts at 9:30 AM, Suresh starts at 8:50, Ramesh starts at 9, and they both go at equal speed. At 9:20 they decide to increase their speeds in order to reach their office on time, which is at 9:30 and this increase in speed was 30 kilometer per hour each, and they manage to reach the office on time. So, the timer begins at 8:50 AM, which means our origin is corresponding to 8:50 AM.

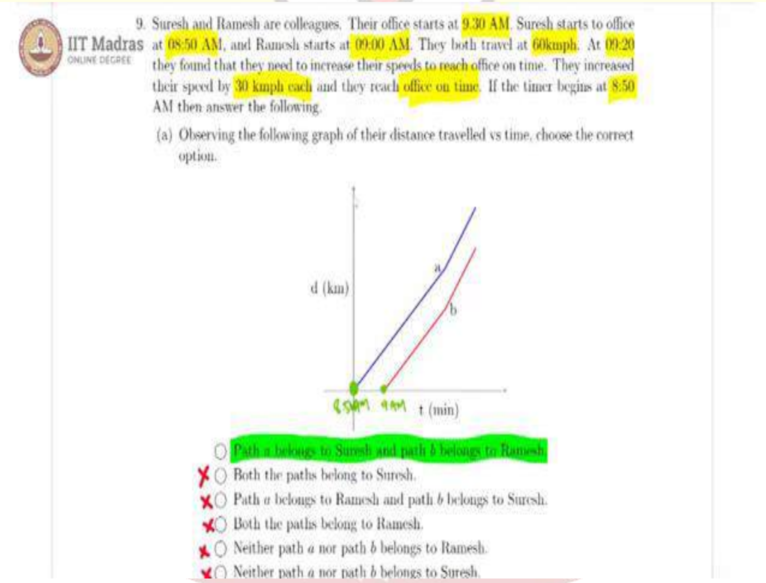
And since we know that Suresh started at 8:50 path A must belong to Suresh and Ramesh started a little late, so this here should be 9 AM. So, B, the path B corresponds to Ramesh's journey, which gives us option A is correct. Of course, this is wrong because both paths do not belong to Suresh. This is also wrong because path A does not belong to Ramesh, both paths do not belong to Ramesh and Ramesh has a path Suresh has a path so all of these options are wrong, only option A is right.

(Refer Slide Time: 1:54)



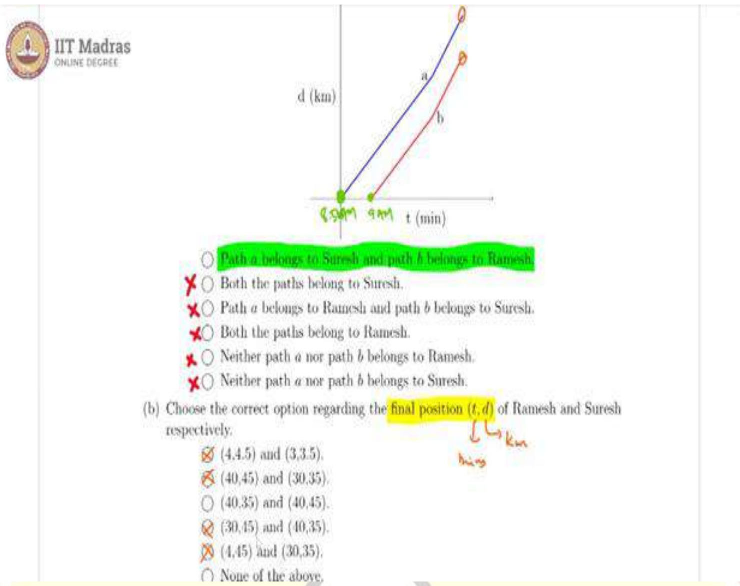
Now, in the second part, we are being asked the final position t, d, where t must be in minutes and d must be in kilometers. So, what, so this is not actually the position, is a coordinate in this particular graph regarding the final position of Ramesh and Suresh respectively.

(Refer Slide Time: 2:24)



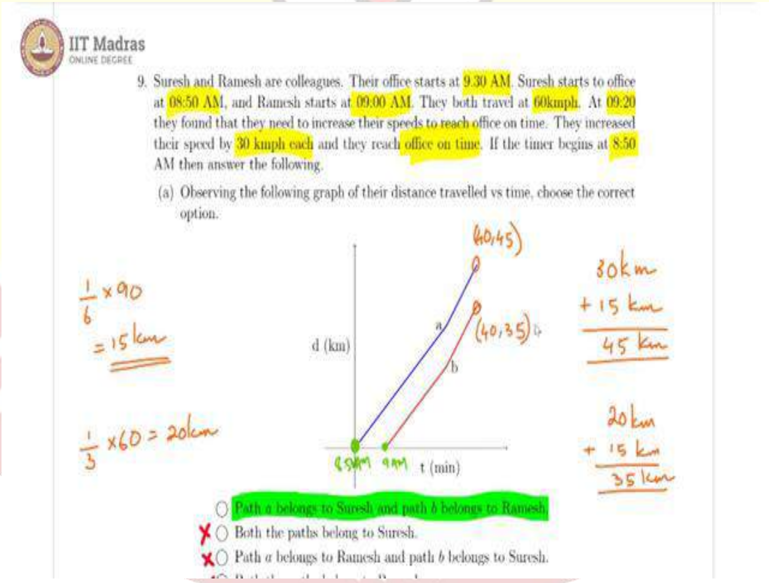
So, we know that, Suresh started at 8:50 and he traveled till 9:30. That means, Suresh traveled for 40 minutes, whereas, Ramesh started at 9 AM and reached office at 9:30 AM. So, Ramesh started for 30 minutes.

(Refer Slide Time: 2:50)



However, both of them reached at the same time, which means, this point and this point in the graph, both of them have the same x coordinate. Now, that clearly rules out this, this, this and this, because none of these have the same x coordinate.

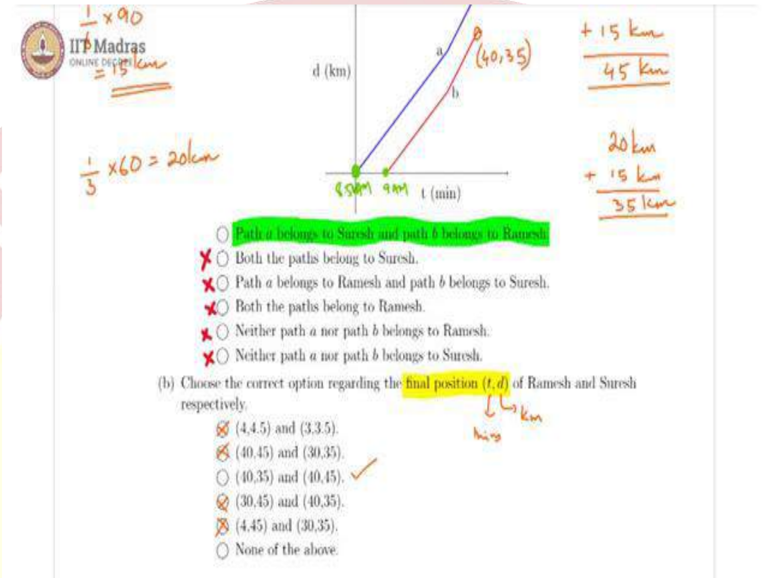
(Refer Slide Time: 3:18)



In terms of the number of kilometers traveled, Suresh goes at 60 kmph for $\frac{1}{2}$ an hour till 9:20. So, in $\frac{1}{2}$ an hour he must have covered 30 kilometers and then for 10 minutes, he goes at a speed of an additional 30 kmph, so, 90 kilometer per hour for 10 minutes. So, 10 minutes is $\frac{1}{6}$ an hour, $\frac{1}{6} \times 90$ gives us 15 km. So, overall Suresh covered 45 km, so this point it must be 40, 45.

Whereas, Ramesh also covered the same 15 km in those 10 minutes but in the initial time of established it is only 20 minutes, he did not cover 30, he instead covered 20 minutes is $\frac{1}{3}$ of an hour $\frac{1}{3} \times 60$ gives us 20 km. So, Ramesh covered 20 km + 15 km s giving us 35 kilometer overall. So, this point here it is 40, 35.

(Refer Slide Time: 4:48)



So, our correct option is this one.

IIT Madras

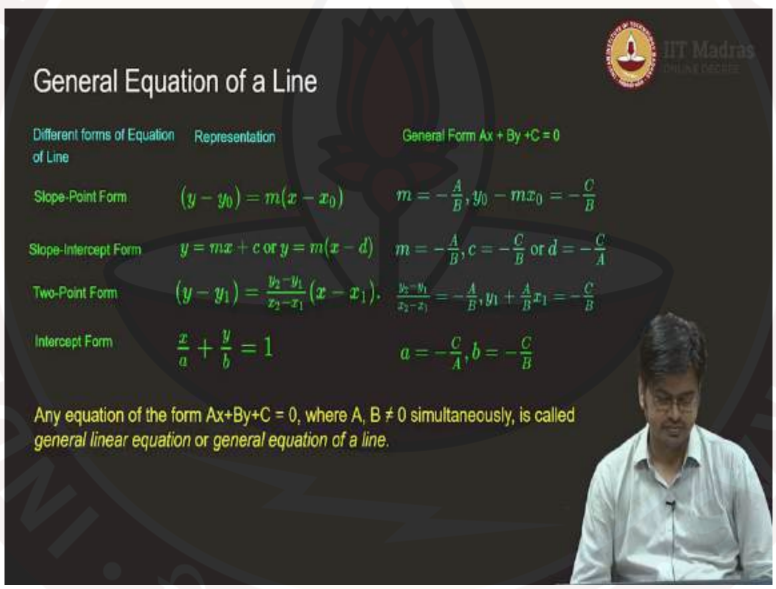
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology Madras – Chennai

Lecture-20
General Equation of Line

So, far in our journey we have studied how to represent on line which is a geometric object in algebraic manner using various forms of equations. This is a time to recollect; what are the forms of equations that we have studied and understand some common properties commonalities in that equation of line and give a general equation of line which will be helpful for further analysis. So, let us see what are the different forms of line; equations of line that we have studied.

(Refer Slide Time: 00:49)



So, in particular we had two forms one is two-point form another when its slope point form. So, first I will list the slope point form, a specialized version of this is slope intercept form where instead of a point you have been given x intercept or a y-intercept. Then we have also studied two-point form given two points how to uniquely determine a line and a specialized version of that is nothing but intercept form.

So we can quickly review these forms like slope point form we have a point (x_0, y_0) which is given to us and a slope m that is given to us. So, we come up with an equation when we give the algebraic representation of this line with slope y with slope m and point (x_0, y_0), we will come up with a representation as $(y-y_0)=m(x-x_0)$. When you come to slope intercept form suppose the x intercept is given to me if I have been given an x intercept then the y coordinate of that point will be 0.

So let us say x intercept is d, in that case my equation from slope point form as slope intercept form is a specialized version of slope in point form. My equation will become $(y-0)=m(x-d)$ if the intercept is at d. So, $y=m(x-d)$, in a similar manner so the y intercept is given to me and that intercept is at c then my y_0 will be replaced by c and x_0 will be replaced by 0 therefore I will come up with an equation $y=mx+c$ that is what is listed here, given a y-intercept and given an x intercept the equation has a form $y=m(x-d)$.

Let us come to two point form we have also seen during the course that this two point form is closely related to slope point form. We also know that given any two points on a line we can determine the slope of a line so in this particular expression m will be replaced by the ratio of the change in y upon change in x. Therefore the two point form will be just replica of this instead of m you will have the difference between y-axis difference between the coordinates of y-axis and difference between the coordinates of x-axis that will be given in this form.

$$(y-y_1)=\frac{y_2-y_1}{x_2-x_1}(x-x_1)$$

Now remember here the points given are not (x_0, y_0) so the points given are (x_1, y_1) and (x_2, y_2) therefore x_0 is replaced by x_1 and y_0 is replaced by y_1 and this thing is nothing but a replacement of m that is how these two forms are also closely related. In an intercept form you will get two intercepts x intercept let us say x intercept is a and y intercept is b then how will these forms change?

If x intercept is a that means I have a point $(a,0)$ so my (x_1, y_1) will be nothing but $x_1=a, y_1=0$ and (y_2-y_1) will be b, x_2-x_1 will be minus a, so $\frac{b}{-a}$ is equal to $\frac{-b}{a}$ x, and if you simplify that you will come up with a very simple expression of the form $\frac{x}{a}+\frac{y}{b}=1$. So, here there is a no-brainer nothing to remember below x you write x-intercept below y you write y-intercept and equate it with 1.

Now if you look at all these forms there is one common feature, let us take the slope point form given a point (x_0, y_0) this $x_0 \wedge y_0$is fixed. The slope of a line is fixed. So, now what we are identifying is we are identifying in a condition in the form of (x, y) what these coordinates should satisfy. So, the variables are x and y.

If you look at all these forms the same feature is visible, the variables are x and y and I have an expression of the form some constant times y some constant times x and added with another constant. Let us take this feature for example $y-y_0=m(x-x_0)$now I want to differentiate between variables and constant. So, I can simply write this as $y-mx=y_0-m x_0$. $y_0-m x_0$ will be the constant associated with this particular equation and y and one variable y is associated with real coefficient 1 and variable x is associated with real coefficient -m.

So in particular I can have a general form of the equation and similar story is true for all this. For example, if you come here, with variable x, $\frac{1}{a}$is a real coefficient that is associated, with variable y, $\frac{1}{b}$that is a real coefficient that is associated and the constant is c. So, I can discuss same things about all these features but one thing is common that I can have a general form of equation which will be of the form $Ax+By+C=0$.

Now let us identify this particular general form with our various expressions like slope point form, the way I discussed the slope point form we already know. In this case we have assumed that b is equal to 1 but I can as well multiply by a constant term throughout the equation and we will have the same equation. So, assuming this holds true let us discuss about this particular expression. So, in this case you can easily see if I relate this equation with this equation that is you rewrite this as $y-mx=y_0-m x_0$.

In that case you can have this expression which will give the value of m when you compare with respect to this expression as $\frac{-A}{B}$ and value of $y_0 - m x_0$, now remember this is a constant term because all these are constants. So, $y_0 - m x_0 = \frac{-C}{B}$. If you are able to understand this then you can easily understand the slope intercept form. Because in the slope-intercept form, $y = mx + c$you have y-intercept which is c therefore your y_0 will be replaced by c and x_0 will be replaced by 0 so if you look at this expression m will still remain $\frac{-A}{B}$, when I am identifying this equation m will still remain by minus $\frac{-A}{B}$, y_0 is identified with $C - \left(\frac{-A}{B}\right) x_0$is 0 so this becomes irrelevant so y_0 is c so $c = \frac{-C}{B}$. In a similar manner you can do for x-intercept and you will get these expressions.

So m as I mentioned $c = \frac{-C}{B}$ and for getting d you just put $x_0 = d$and $y_0 = 0$, you will get this expression. So, same exercise can be done for two point form and intercept form remember this m will be replaced by a ratio of these two differences. So, m is replaced by a ratio of these two differences there is no (x_0, y_0) there will be (x_1, y_1) therefore you will have an expression of this form.

But remember this $\frac{-C}{B}$is common everywhere the slope is $\frac{-A}{B}$ everywhere so essentially, we have got one simple general equation. Similar things you can do for $\frac{x}{a} + \frac{y}{b} = 1$ that is intercept form and you will get $a = \frac{-C}{A}, b = \frac{-C}{B}$. So, what we have seen here is an exact matching one-to-one correspondence of a general equation with respect to this equation.

Now why should I consider general equation? Remember when we figured out this representation our assumption was these are non vertical lines. For vertical lines our slope do not exist but in this case if you; and those lines are where the slope do not exist those lines are vertical lines. They are of the form x is equal to some constant. If you look at this equation which is a general form of this equation you just put B to be equal to 0 you will get $Ax+c=0$ that means x is equal to some constant x is equal to $\frac{-C}{A}$, you will get that is what our intercept form also reveals.

So all these lines are actually vertical lines, so this general equation is capable of handling vertical lines also, horizontal lines are anyway handled here because if you put m is equal to 0 the horizontal line is handled. While we were deriving these forms we were always assuming non-vertical lines. So, non-vertical lines are covered as well as vertical lines are covered therefore this equation is a general form of equation of a line.

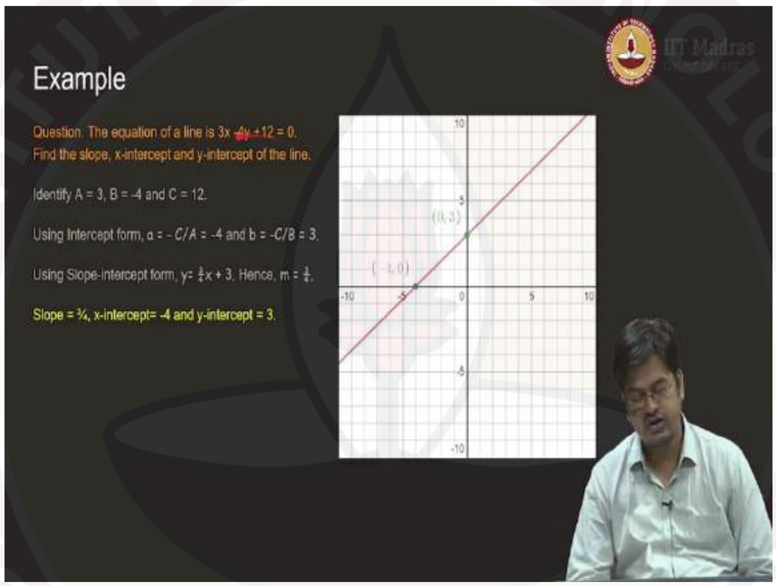
Also, in your earlier classes you might have studied this as a polynomial in without this equal to 0 Ax + By + C is a polynomial in two variables and it is a linear polynomial in two variables. Therefore you will hear a term called linear equation in two variables. So, in particular if this has to represent general form of a equation of line then A and B cannot be simultaneously equal to 0.

If A and B are simultaneously equal to 0 then I am actually equating constant with a zero which is invalid therefore the assumption will always be A and B cannot be simultaneously equal to 0. Though individually they can be equal to 0 for example you can put A is equal to 0 then you will get y is equal to some constant which is a line parallel to x axis. You can put B is equal to 0 then you will get a line x is equal to constant, x is equal to constant is parallel to y axis.

So now we will bring up a definition that any equation of the form Ax + B y + C = 0 where A and B are not equal to 0 simultaneously individually they can be 0 or they can be nonzero as well is called general linear equation because we are handling a linear polynomial which is equated to 0 so it is an equation, general linear equation or general equation of a line. So, what we are summarizing here is a polynomial in two variables or and general linear equation in two variables gives you line.

So this is the identification of a geometric object called straight line with an algebraic representation of general linear equation. So, this will give us both the strength in our analysis because now you do not have to discuss about the line. But you can as well discuss about its algebraic representation or you can start with an algebraic representation of a line and then discuss about the geometric properties of the line. How let us see in the next slide.

(Refer Slide Time: 14:30)



So, here is an example, the example gives you a question that the equation of a line is $3x-4y+12=0$. Now I do not know how this line behaves now I want to see how this linear equation represents a line. So, when I talk about a line what is the natural question we will talk about what are the two points that uniquely determine this line or you can ask what is the slope of a line and give me one point on a line because we have slope-intercept form or we have two point form any of them should be usable.

So in order to discuss about the geometric aspects we can ask a question that find the slope or x intercept or y intercept of a line. So, how will you find this the job is pretty simple let us go back and revisit the previous slide which will make the job very simple. Suppose I want to determine the x-intercept and y-intercept then I have this intercept form right which says that a is the x-intercept and b is the y-intercept.

Now you I have been given an equation in this form which is $Ax+By+C=0$ so I can image lately consider this equation and consider the values of a and b which is $-C/A$ and B is equal to $\frac{-C}{B}$. So, let us go and do the same thing on the on the our; now our problem so we have identified $Ax+By+C=0$. So, what is A, A is 3, B is minus4, C is positive 12. So, what should be my x intercept A as you have seen in the previous slide is $-C/A$.

So what is C? $\frac{12}{A}$ which is 3 so my a is 4, and a minus sign associated with it so $a=-4$. In a similar manner you can talk about y intercept which is $\frac{12}{-4}=-3$ but a minus sign because it is $\frac{-C}{B}$ so it will be 3. So, now we can readily answer the question what is on a x-intercept and y-intercept.

Now the question comes what is the slope of a line. So, for slope of a line you can use the slope intercept form $y=mx+c$. So, identify this equation in the form of $y=mx+c$ so if you look at this equation, I should push this 4y to the right hand side that gives me $y=\frac{3}{4}x+\frac{12}{4}$. So, my m should be $\frac{3}{4}$this is the answer. So, slope intercept form you have y is equal to 3 by 4 x plus 3 so the slope is naturally $\frac{3}{4}$, this easy is our calculation.

Now we have identified an algebraic object as a geometric object. Now let us see what we can do further and we can actually verify this graphically you know although it may be correct it is always better to verify it graphically. So, slope is $\frac{3}{4}$ x intercepts should be - 4 and y intercept should be 3 if you want to satisfy the equation of this line this should happen right.

So this is how we have drawn so the x-intercept is -4, y intercept is 3 and the line passes through this. Now you pick for verification purposes you can pick any point on this line and you can put the values of the coordinates into the equation of a line and verify that it will give you the value 0 that will be the identification that your answer is correct.

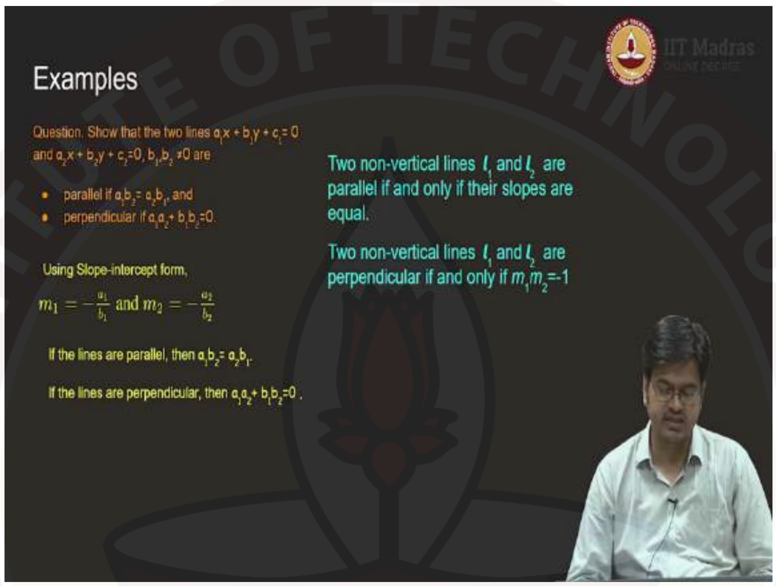
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology Madras – Chennai

Lecture-21
Equation of Parallel and Perpendicular Lines in General Form

(Refer Slide Time: 00:19)



Let us look at next example which is another application of a general form of equation of a line. The example is stated in the form of a question that is if I have been given two lines $a_1x + b_1y + c_1 = 0$, $a_2x + b_2y + c_2 = 0$, $b_1, b_2 \neq 0$. What does this mean? That means the lines are non-vertical. $b_1, b_2 \neq 0$ means the lines are non-vertical you can verify for yourself.

Now two such lines are parallel if $a_1b_2 = a_2b_1$and perpendicular if $a_1a_2 + b_1b_2 = 0$. This is an interesting application of general form of equation of line. And if you recollect, we have derived some characterization of line in terms of slope. So, let us try to see this problem so let me first identify if I want to characterize parallel and perpendicular lines what should I do?

What is a parallel line, how will I identify a parallel line when I will have their slopes to be equal and how will I identify a perpendicular line, when the product of the slopes of the two lines is -1? So, if you remember this then the job reduces to finding the slopes of the two lines. Can I find a slope of these lines? Let us first consider this line $a_1 x+b_1 y+c_1=0$. You should be immediately able to identify this with slope point form which is $y=mx+c$.

So if I want to adjust this equation in the form of $y=mx+c$ then what should I do? Because b_1 is nonzero I can divide throughout by b_1 and shift this coordinate of y to their right-hand side of the equation. So, I will get $y=\frac{-a_1}{b_1}-\frac{c_1}{b_1}$. So what is the slope $\frac{-a_1}{b_1}$. A similar trick you can apply here and therefore you will get $m_2=\frac{-a_2}{b_2}$. So using slope intercept form you have got $m_1=\frac{-a_1}{b_1}\wedge m_2=\frac{-a_2}{b_2}$.

Now let us recollect the famous fact because $b_1 \wedge b_2$ are not equal to 0 we are not considering vertical lines. So, two non-vertical lines are parallel if and only if their slopes are equal. So, what you will do you will just put $m_1=m_2$because you have been given that the lines are parallel. So if you put $m_1=m_2$, minus sign will cancel each other $\frac{a_1}{b_1}=\frac{a_2}{b_2}$. Multiply both sides by $b_1 b_2$, b_1, b_2 are nonzero.

So multiply both sides by $b_1 b_2$, you will get $a_1 b_2=a_2 b_1$. Therefore, the lines are parallel then $a_1 b_2=a_2 b_1$. In a similar manner we also know something about perpendicular lines that the product of their slopes is -1, if the lines are perpendicular. So, just multiply m_1, m_2 and equated to -1. Minus sign will cancel each other so you will get $\frac{a_1}{b_1}\times\frac{a_2}{b_2}=-1$.

So take the denominator on the right hand side that is $b_1 b_2$, so $a_1 a_2=-b_1 b_2$ which essentially means $a_1 a_2+b_1 b_2=0$. Therefore, we have proved the result. So, now what we have done right now is we have related our result about the characterization of perpendicular and parallel line via slope to a general form of equation and this is the new condition that we are coming up with if the lines are parallel and you have been given to two non -vertical lines and their general forms then you just need to check that $a_1 b_2 = a_2 b_1$ for the lines to be parallel and $a_1 a_2 + b_1 b_2 = 0$ for the lines to be perpendicular. This you can consider as another characterization of parallel and perpendicular lines using a general form of the equation of lines.

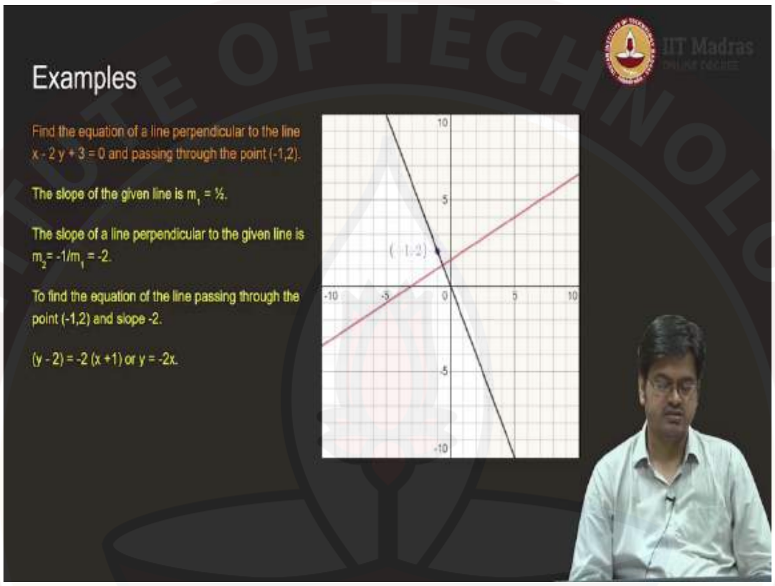
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology Madras – Chennai

Lecture-22
Equation of a Perpendicular Line Passing Through a Point

(Refer Slide Time: 00:15)



So, now you have been presented with an equation of line and a point and the question is you find the equation of a line perpendicular to the line $x-2y+3=0$ and passing through the point $(-1,2)$. So, in this case let us identify the general form of the equation that is $Ax+By+C=0$ and you can easily see that $A=1, B=-2, C=3$.

Therefore, the slope of the given line the line that is given to you will be $\frac{-A}{B}$ which is $\frac{-1}{-2}$ that will be $\frac{1}{2}$. So, the slope of the given line $m_1=\frac{1}{2}$, now if at all a line is perpendicular to it then you already know that the product of the slopes is -1. So if the product of the slopes is -1 then $m_1 m_2=-1$ that is $m_2=\frac{-1}{m_1}$. So m_1 is $\frac{1}{2}$ which will give me $m_2=-2$.

So now the problem reduces to the slope of a given line is -2 and it passes through point $(-1,2)$ and I want to find the equation of a line that is passing through point $(-1,2)$ and has slope -2. So, use the slope point form $y-y_0=m(x-x_0)$, y_0 is 2 so you can easily see $y-2=-2(x+1)$, ¿, so $(x+1)$, rearrange the terms so this 2 will get cancelled constant therefore I will get the equation of a line to be $y=-2x$ or in a general form you can write this as $-2x+y=0$.

So, let us try to figure out whether the line which we have actually found is perpendicular or not. So, the orange line is the line for which the equation is given $x-2y+3=0$ the point $(-1,2)$ is displayed in the graph and the line passing through it is also displayed and you can clearly see the angle that is made is 90 degrees therefore the lines are perpendicular and our answer is correct. So, our verification test has passed.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

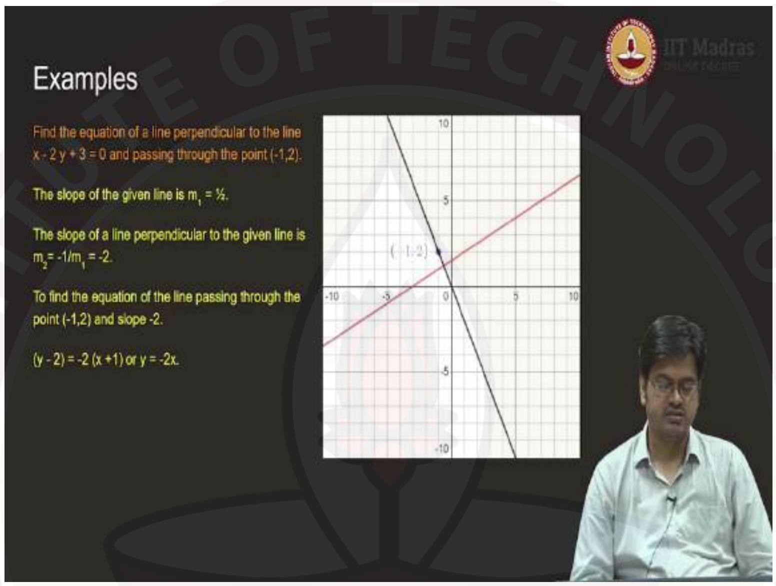
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology Madras – Chennai

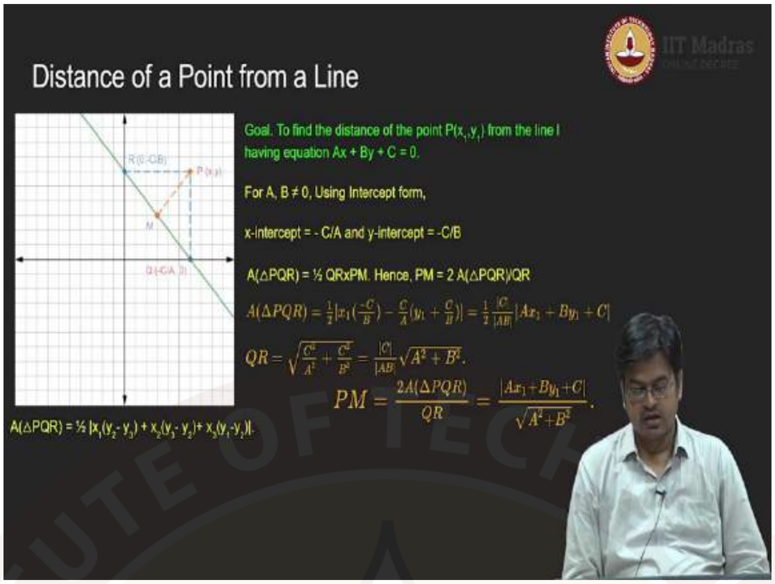
Lecture-23
Distance of a point from the given line

(Refer Slide Time: 00:16)



So, we have verified the answer now what you can see here is (-1, 2) is a point which is lying on the line which is perpendicular to the given line. An interesting question can be asked that what is the distance of this; point from the given line. Let us try to answer that question in the next slide.

(Refer Slide Time: 00:40)



So, the question is given any point I have another line the point is not collinear to the line then what is the distance of that point from a line. So, let us take that as our goal. To be precise we are interested in finding the distance of the point P which has coordinates (x_1, y_1) of this point from the line l which has equation $Ax+By+C=0$ and this is a general form of equation.

Now how will we proceed? So, if I want to understand the location of P, I need to do some analysis because $Ax+By+C=0$ is a completely geometric object. So, I need to understand this line in terms of its geometric concepts. So, what are the geometric concepts associated with this line they are slope x intercept y intercept or points on this particular line. So, let us identify those things first.

So if we assume that A and B both are not equal to 0 then I can rewrite this equation in the form of intercept form that is $\frac{x}{a}+\frac{y}{b}=1$ so in that case my a is actually $\frac{-C}{A}$ that is x intercept and my small b will be $\frac{-C}{B}$ which is y intercept. So, I have identified this line as I have identified the 2 points and these 2 points uniquely determine the line. So, I know how the line is located.

Let us try to visualize this line in terms of the graph of a function. So, as you can see I have mentioned that x intercept is $\frac{-C}{A}$ so it is mentioned as a point Q which is $\left(\frac{-C}{A},0\right)$, y intercept is $\frac{-C}{B}$ which is identified here so $\left(0,-\frac{C}{A}\right)$, the point P is located here it may be located anywhere but right now the point P is located here it has coordinates (x_1, y_1). So, now I want to identify a distance of this point from this line, the line joining the points Q and R.

So, what is the distance? It should be the shortest distance from the line, so the shortest distance in this case if you move along this line the shortest distance in this case is a point where the point is actually perpendicular to the line. So, what I want to say is the shortest distance is the one which is the perpendicular distance. So, the entire question reduces to how to find this perpendicular distance PM.

So, let us try to see what are the geometric objects associated with this. So, you can see from the dotted lines the geometric object that I can associate with this particular distance is a triangle PQR. Now if I want to find the distance PM, I can take help of this triangle PQR so that I will be able to find the distance PM. So, how will I do that? First you see if I want to compute the area of triangle PQR what do I need to know?

I need to know the base and the height and the area of a triangle is half base into height. So, half base into height means $\frac{1}{2}\times QR\times PM$. I do not know what is PM. But we have already seen in this course how to find area of a triangle where its coordinates are given. So, even though I do not know what is PM I know how to compute the area of a triangle. The next question is do I know how to compute the length QR?

Yes of course because this is x intercept this is y intercept and these are the lines which are the distances on x and y axis and all of them form a right-angled triangle. So, by Pythagorean theorem I will be able to find the length of QR. So, I can reformulate the question as

$PM = 2 \times \frac{A(\Delta PQR)}{QR}$. So, now I know how to compute the length PM if I know how to compute area of triangle and how to compute the length of line segment QR both of which I know.

So, let us go ahead and try to compute area of triangle PQR. So, here is our formula for area of triangle PQR which has coordinates $(x_1, y_1), (x_2, y_2), \wedge (x_3, y_3)$. So, let us start with (x_1, y_1), the (x_1, y_1) is the first coordinate remember you will always take this in anti-clockwise direction. So, I will start with this coordinate then I will go to R and then I will go to Q. So, this is (x_1, y_1) this is (x_2, y_2) and this is (x_3, y_3) according to the notation that is given in the formula.

So you will see $x_1(y_2 - y_3)$ so x_1 is first coordinate it will remain x_1 because P has coordinate (x_1, y_1), y_2 is $\frac{-C}{B}$, y_3 is zero. So, you will get $x_1\left(\frac{-C}{B} - 0\right)$ then the next term that is x_2, x_2 here is 0 so this entire thing vanishes then you go to x_3, what is x_3? x_3 is $\frac{-C}{A}$, into y_1, which is y_1 as it is $-\left(\frac{-C}{B}\right)$ so $\left(y_1 + \frac{C}{B}\right)$, this is how I got the formula.

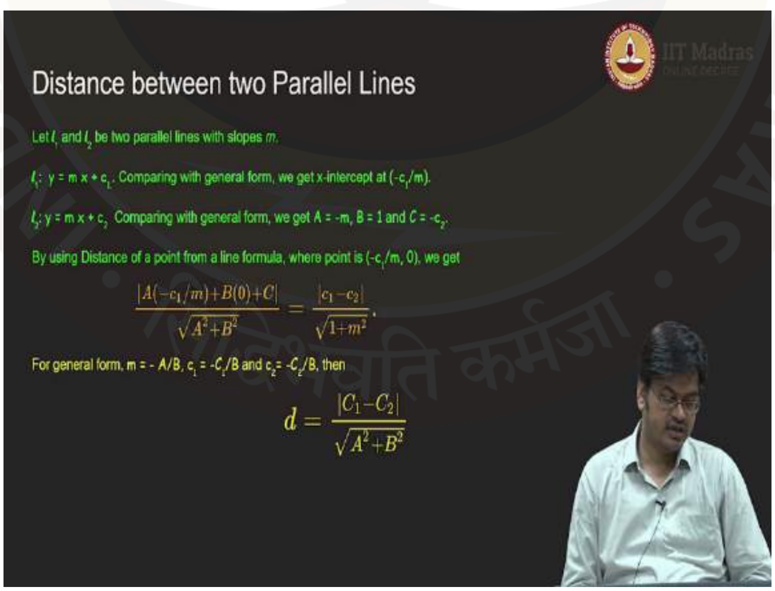
So, if you look at this formula closely you can actually take C common from all within the mod sign so you can take $\left|\frac{C}{B}\right|$, denominator has terms containing B and AB. So, if you want to take those terms out you multiply throughout by AB or you find the LCM is AB and you take AB out so you will get $\frac{1}{2} \times \frac{|C|}{|B|} \times |A x_1 + B y_1 + C|$, remember this is the term corresponding to general form of the equation.

Now we have seen how to compute area of triangle PQR. Next, we will see how to compute the length QR. But length QR is actually very easy because I have a point Q which has only x-coordinate and I have a point R which has only y-coordinate. So, it will be as if computing the distance of length QR is $\sqrt{\frac{C^2}{A^2}+\frac{C^2}{B^2}}$ these are the 2 sides of the triangle and this QR is the hypotenuse of that right-angle triangle.

So, again you can simplify this to amend to this form so you can take out C common so you will get $|C|$, you take A and B common you will get $|AB|$ and then you will get $\sqrt{A^2+B^2}$ which is which is in the numerator and now if you look at this form PM which is the length of the line segment PM is $2\times\frac{A(\Delta PQR)}{QR}$, so just now it is just a matter of feeding the values this half will get cancelled with this 2 and area of triangle PQR is this and QR is this therefore this constants also will vanish because they are same.

And you will get the formula to be equal to $\frac{|Ax_1+By_1+C|}{\sqrt{A^2+B^2}}$ this is how you will calculate a perpendicular distance of a point from a line. Now this idea can be helpful in finding one more thing that is a distance between two parallel lines.

(Refer Slide Time: 10:22)



So, the question can be asked is I have two parallel lines what is the distance between two parallel lines. So, let us take the set up because the lines are parallel l1 and l2 they have common slope or the same slope, so their slope is m. Then you can use the slope point form which is

$y = mx + c_1$. Now I want to use the previous concept that I have introduced distance of a point from a line. So, I will first identify this with x intercept. So, what will be the x intercept in this case?

If you identify this line it is very easy to see go back to the general form and figure out that x-intercept is $\frac{-c_1}{m}$, because $B = 1$ here $B = -m$ and $C = -c_1$, so the intercept is this $\frac{-c_1}{m}$. Let us take another line that is l2 it has same slope identify it with our standard form $A = -m, B = 1, \wedge C = -c_2$. So, now given x-intercept what are the coordinates of this x-intercept $(\frac{-c_1}{m}, 0)$.

So now the problem reduces to finding the distance of this point from this line ok. So, by using the distance of a point from a line formula where the point is $(\frac{-c_1}{m}, 0)$, you just need to substitute this point (x_1, y_1) into this formula for the distance of a line which is given as $\frac{|Ax_1 + By_1 + C|}{\sqrt{A^2 + B^2}}$.

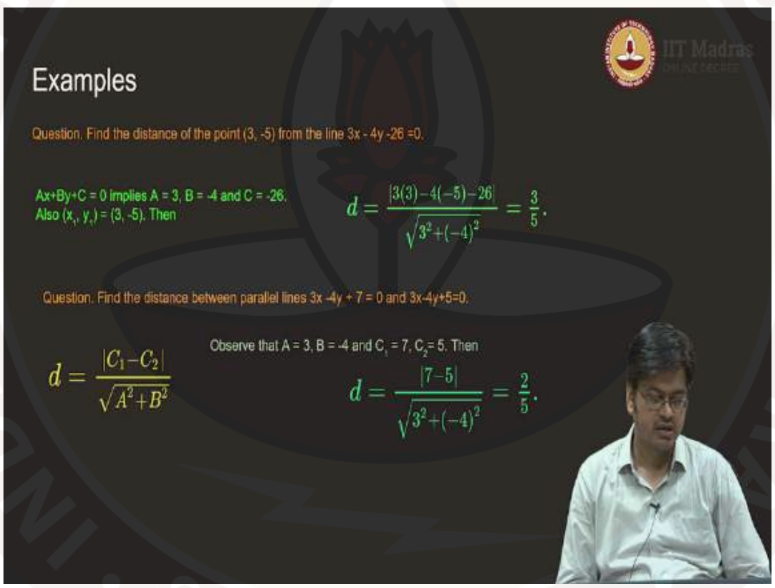
So, my point $x_1 is - \frac{c_1}{m}$ substituted here, y_1 is 0 substituted here you will get the formula to be equal to $\frac{|C_1 - C_2|}{\sqrt{A^2 + B^2}}$ so in this case $\sqrt{A^2 + B^2}$, B was 1, A is -m so it is $\sqrt{1 + m^2}$.

Now you can actually identify this formula in the general equation form also. So, in the general form instead of B = 1 we have slope which is equal to -A/B and $c_1 = \frac{-C_1}{B}$ and $c_2 = \frac{-C_2}{B}$. So, this I am matching with both equations in general form these are slope point forms but now if you match these equations with a general form you will get this description of the line where you have $Ax + By + C1$ as one line is equal to 0 as one equation of line.

$Ax+By+C_2=0$ as equation of the second line. So, in that case this is the form and therefore now you just substitute these values into this expression. So, this m will be replaced by - A/B so you will get $\sqrt{A^2+B^2}$ here and some $|AB|$ will come out common and therefore finally that will cancel off with this B and you will get the expression of the form $\frac{|C_1-C_2|}{\sqrt{A^2+B^2}}, C_1 \wedge C_2$ belong to general form of equation.

So, this gives us a clear-cut understanding of the interconnection between the slope point form and general form of equation and we have figured out what is a distance between two parallel lines using distance of a point from a line formula.

(Refer Slide Time: 14:42)



So, now we will solve some examples to concretize the concepts so here are the examples in line. So, you have been asked to find a distance of a point $(3,-5)$ from the line $3x-4y-26=0$. So, in this case you just need to apply the formula, what is a formula, $\frac{|Ax_1+By_1+C|}{\sqrt{A^2+B^2}}$. So, what is A, B, and C here is the key question? (x_1, y_1) is known to be $(3,-5)$. So, A is 3, B is -4, C is -26 then you just need to apply that formula which will give the denominator square root of 25 it will give me 5 the numerator will be 3.

In a similar manner you can ask a question what is the distance between two parallel lines $3x-4y+7=0$, $3x-4y+5=0$. So, what is a formula that we have derived it is $\frac{|C_1-C_2|}{\sqrt{A^2+B^2}}$ it is very straightforward. So, what is C_1 here? C_1 the first line the constant term is 7 the second line the constant term is 5. So, it will be modulus of $\frac{7-5}{\sqrt{3^2+(-4)^2}}$. So, you will get the answer to be 2 / 5.

IIT Madras

ONLINE DEGREE

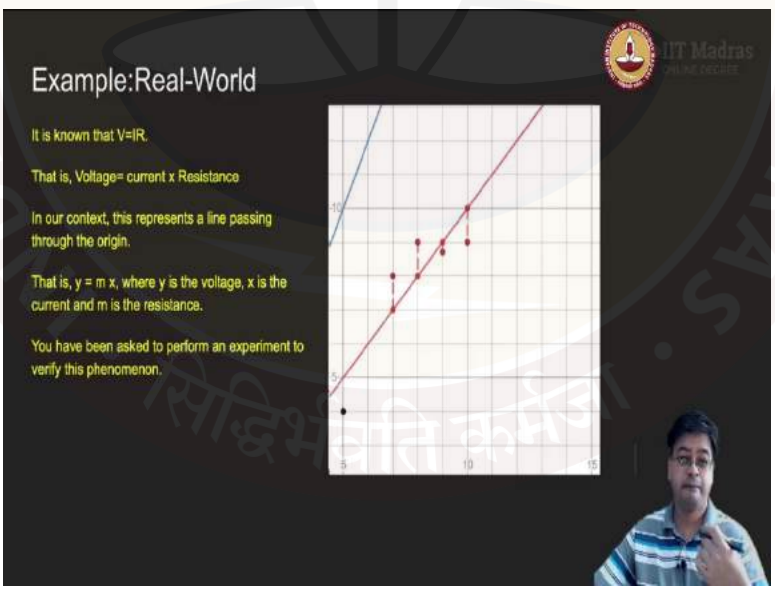
Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology Madras – Chennai

Lecture-24
Straight Line Fit

Welcome friends welcome back so far what we have seen is a distance of a line from a point distance between two parallel lines. But the question now we can ask is, is that the only distance that we can seek as a distance of a point from the line. To demonstrate this let me give you one example where the paradigm will change as we will compare several points set of points and we will compare the distance from those set of points to the line and the paradigm change that I want to say is you will think differently how the distance will change from a line.

So, let us take one simple example this example is related to a small experiment that you might have conducted in your lab.

(Refer Slide Time: 01:16)



It is a physics experiment which says $V = IR$ that is voltage is equal to current times the resistance. Voltage is equal to current times the resistance you all know this is a law this is the law of physics where voltage is measured in volts current is measured in amperes and resistance in ohms. Now the experiment that a physics teacher asked you to conduct is you have to verify this law or using this law can you compute the resistance of a particular equipment.

So now what you will do is you will actually relate this with our equation of a straight line. So, if I want to relate this with the equation of a straight line then what will happen you see V is voltage so on the right hand side you can replace this voltage by y then the current that is delivered to the circuit or the equipment you can denote it by say x and you want to determine the resistance which is an unknown so you can put it as m.

And what is the constant? The constant is 0, so you can relate this with the equation $y = mx$, where y is the voltage, x is the current and m is the resistance and the whole purpose is to determine this resistance over here m. So, the setup is ready the lab technician has arranged a set up and you just have to go and perform the experiment and verify this phenomenon. So, the catch over here is you want to determine what is a resistance.

So, the lab technician was very kind he has given you a priori information that there are only two kinds of resistors our lab has one has a resistance of 1 ohm another one has a resistance of 2 ohms. This is the information that is given to you. Also notice the fact that this line is passing through the origin that means (0, 0) is one point why (0, 0) should be a one point because there is no current then there is no voltage this is our assumption.

So (0, 0) is one point and this line is passing through the origin so if I look at a mathematical theory that I have studied so far I can safely assume if I get one reading if I get one reading from that circuit that will help me in understanding the behavior and I can safely go and tell what is the resistance of this particular equipment. Let us try to see how this assumption works out over here.

Now this is the data you have conducted some experiments you have observed some data so it is like you have passed a current of 1 ampere and you received the output of 2 volts here you can say you have passed the current of 5 amperes and you have received output which is 4 volts and so on and so forth. So, this is how it is working on. Now we want to identify what is the correct line that will fit because I know from theory that this is a line passing through the origin.

So in particular if I tell you this line which is (1, 2) and (0, 0) then I will get the equation of line using a slope point form or point - point form we also know that the intercept is (0, 0) so slope intercept form $y=mx+c$, where c is 0 you can easily see the line that passes through this point is $y=2x$. But with the same register you also got these readings. So, let us see based on the lab technicians' knowledge if we draw two lines, they will be seen they will be visible like this, interesting.

So, if I take only one observation and stop my experiment, I will get the line $y=2x$. But if I go for more experimentation then I am getting a line which seems to be similar to $y=x$. Now what is it that is happening here, which line is a better fit. So, I need to answer this question because this line actually passes through the point (0, 0) and $y=2x$, this line is not passing through any of the points.

So which line is better that is a natural question that comes to our mind? So, we will try to answer this question mathematically. So, how will I answer this question mathematically? Let us zoom in and consider our notion of a perpendicular distance. What is a perpendicular distance? You will actually drop a perpendicular from this point to this point and you will compute the distance of a line. Is that distance a correct distance? Geometrically it is a correct distance that is a distance of a line.

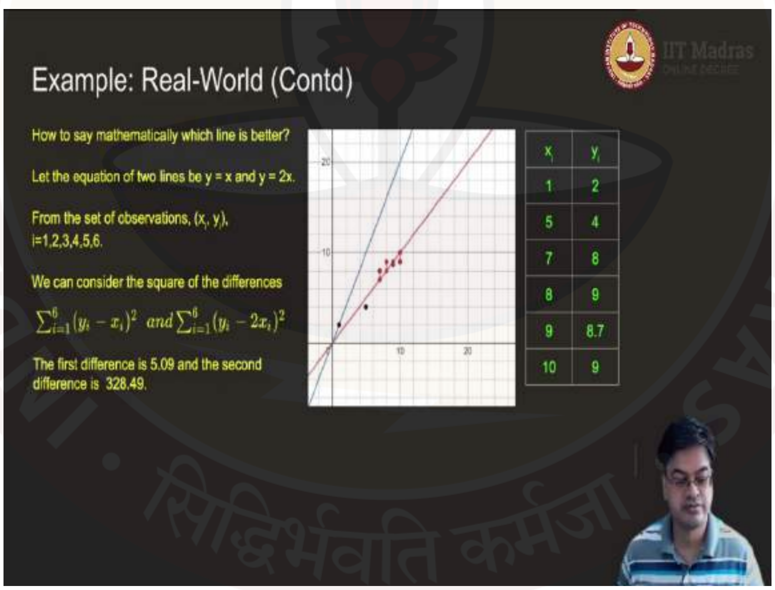
But in this context that we are taking real-world context what is happening here is if I pass a current of let us say 7 amperes this particular line is saying I should get a voltage of 7 volts but actually I got a voltage of 8 volts. So, now I may not be interested if I drop a perpendicular from this point to this point because this line is $y=x$ it may cross this line at point 7.5. I am not interested what is the value of y at point 7.5.

I should be interested in what is the value of y at point 7 because I have passed the current of 7 amperes not 7.5 amperes. So, the perspective of distance changes here because I want to find the distance for this particular value of x from the line and the point how to go about then we will not consider a perpendicular distance. This is a paradigm shift that I was talking about at the beginning of the video.

So now I will not consider this thing but I will consider this distance that is a vertical distance the distance that is parallel to y axis that is what I will consider. So, once I consider the distance that is parallel to y axis, I have to consider these distances. So, again coming back to the question which line is the best-fit line I can consider similar distances over here. And I can consider similar distances over the blue line.

So which line is the best fit? We will try to answer this question mathematically. So, mathematically we have seen that perpendicular distance will not fetch me any result directly. So, I need to consider the distances that are parallel to y axis.

(Refer Slide Time: 08:57)



So, let us formalize this in a real term, this is the data that was shown in the picture. So, for 1 ampere you have got 2 volts current. For 5 ampere you got 4 volts current, for 7 you got 8, 8 you got 9, 9 you got 8.7 and 10 you got 9. So, there is no direct relation between y and x; you cannot figure out the $y = x$ is visible over here but something is there which is making that line pass very close to all these points.

This is the this is the demonstration, so $y=2x$ is way apart and we are assuming that the hypothesis given by the lab technician is correct. So, I want to mathematically formulate this problem. There are two lines $y=x \wedge y=2x$ both pass through the origin, so current 0 voltage 0 hypothesis is correct. Now you have the set of observations xi's and yi's. I want to compute which line is better.

So, let us try to see if I consider the sum of the differences, what do I mean by some of the differences? If I consider $y=x$ is a valid equation of line then I will consider $y_i - x_i$, that is the distance between the line y and x because here y is equal to x if I input x_i my point that I will get is also x_i because $y=x_i$ and the actual output that I have got is y_i so I will consider $y_i - x_i$ as one coordinate and $y_i - 2x_i$ as another difference that will be a point over here $y_i - 2x_i$ it will be a point over here.

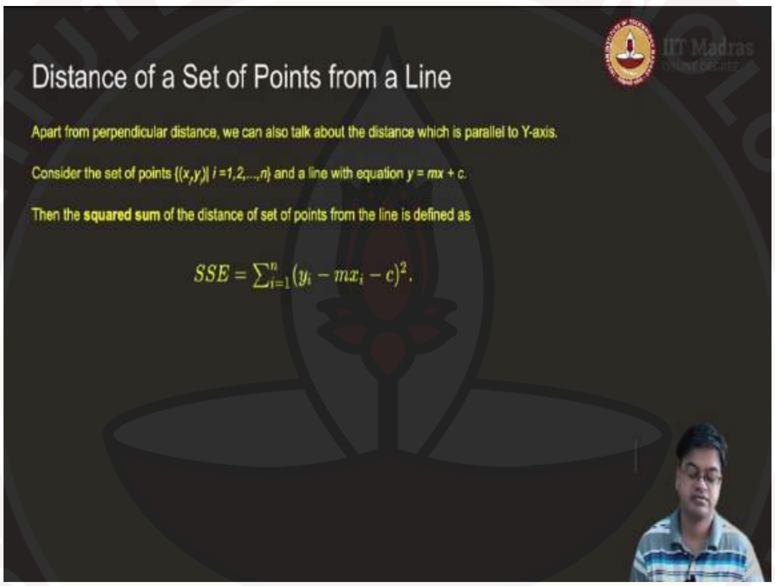
But if I just consider the differences the problem is the differences may cancel each other some differences may be positive some differences may be negative. so, I do not want those differences to cancel out each other so what I will do is I will take square of them. So, in particular we can define the sum square difference that is $\sum_{i=1}^{6} (y_i - x_i)^2 \wedge \sum_{i=1}^{6} (y_i - 2x_i)^2$.

Now what this difference is calculating? It is calculating the difference between $y_i \wedge x_i$ in the first case and $y_i \wedge 2x_i$in the second case that is the error that we have made when we actually saw the output on the error the equipment has made or in our recording the error which is made in whatever way that is the error made. So, $(y_i - x_i)^2$ and $(y_i - 2x_i)^2$ right. Now what do you think which one will be better the one that will be better which will have a least difference.

So, you can actually put in these values and compute these differences and square them sum over them you will get the first difference is 5.09 and the second difference is 328.49. In this situation what should be our conclusion? Our conclusion should be that the difference where the difference is least that is 5.09 this must be a better line as compared to this line that essentially reduces to a conclusion that y is equal to x is a better line as compared to y is equal to 2x which is pretty evident intuitive from the figure as well.

So, you can see this figure you can see this chunk of points that are located around $y=x$ and therefore the resistance of the equipment that is given to us must be 1 ohm that should be our conclusion. So, I want to introduce a notion of this kind to handle the real-world problems. So, let us see what is that notion? In this case you were very lucky the lab technician has given you the set of points or the resistance values there are only two resistance values.

(Refer Slide Time: 13:49)



But real life is not that lucky, so there they may not give you the set of values, and you want to find out what is the best line that is passing through these set of points. In that case this notion of a distance of a set of points from a line may help. So, what is this notion? First of all, we know one notion is perpendicular distance but that perpendicular distance may not be of much use when we are coming to the real-world perspective.

In that case we talk about the distances that are parallel to y axis from the distance of a points that are parallel to y axis. So, in particular if you have been given n points $\{(x_i, y_i) | i=1,2,\ldots,n\}$. You just plot this equation $y=mx+c$. Now remember here this equation is valid when it is not a vertical line. If it is a vertical line this equation is not valid. And if it is a vertical line you do not need such a complicated procedure to estimate it.

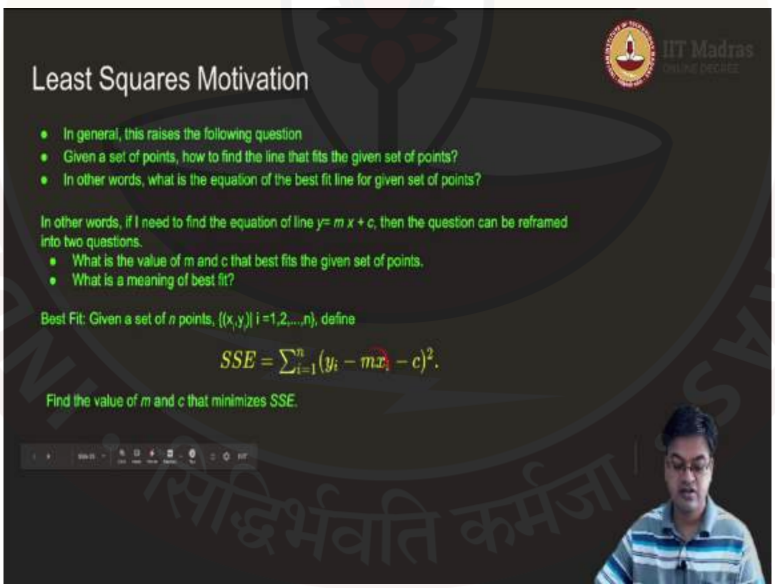
So $y=mx+c$ is our standard equation of line which is a slope point form or slope-intercept form to be precise and then as in the previous case we have defined the squared sum of the distance of the set of points from the line. So, in the previous case $y_i - x_i$ but in this case what should it be $(y_i - m x_i - c)^2$ and you have to sum over all of them. So, $\sum_{i=1}^{n} (y_i - m x_i - c)^2$.

So, we call this as sum squared error or some squared distance, sum squared error so the abbreviation is SSE.

$$SSE = \sum_{i=1}^{n} (y_i - m x_i - c)^2$$

Now the fact is when we are handling a general problem, we do not know what will be m and what will be c.

(Refer Slide Time: 16:08)



So, our goal should be if I want to find the best line, I want to find the best line passing through this point what should be my goal. So, these raises two questions if I have some square, I want to know the value of m I want to know the value of c? So, given the set of points how to find a line that fits the given set of points remember now I am not uniquely determining the line I am saying but that fits the given set of points.

The line may not pass through any of the points in this particular case in other words $y=mx+c$ so what is the equation of the line that best fits the given set of points. This will mean I need to find an equation of a line $y=mx+c$ and then the question can be reframed into two questions that is what do I mean by the value of m and c that best fits the line and then I have to define what is the best fit according to me.

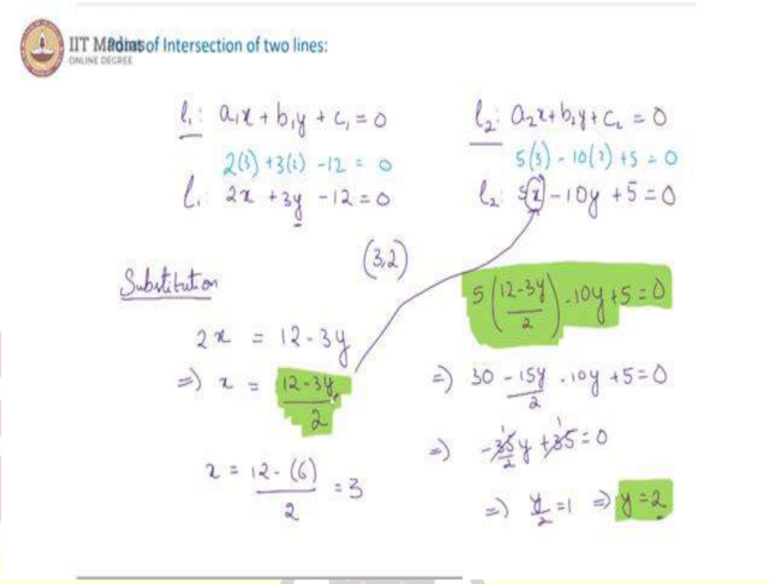
Obviously, the best fit according to me will be the sum squared error minimization. And so, if I define SSE in this manner then I want to find the values of m and c that minimize SSE but this is right now beyond our scope as so far, we have handled only linear terms. But if you look at these terms, they appear to be in the form of squares of something. So, we need to divide some strategies in order to find this minimization for m and c so with that we will see in few upcoming videos of the course, thank you.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week-03 Tutorial - Point of Intersection of two lines

(Refer Slide Time: 0:14)



Hello mathematics students. In this tutorial, we are going to learn to find the point of intersection of two given lines. So, you have two-line equations given to you. Let us call one $a_1x + b_1y + c_1 = 0$, let this be line l1, and line l2 is a $a_2x + b_2y + c_2 = 0$. And we try to find out the point at which these two lines intersect. And that would basically be the solution the (x, y)which satisfies l1 and l2 as well. It is easier to observe this process with example. So, let us take 2 example lines and find out where they intersect.

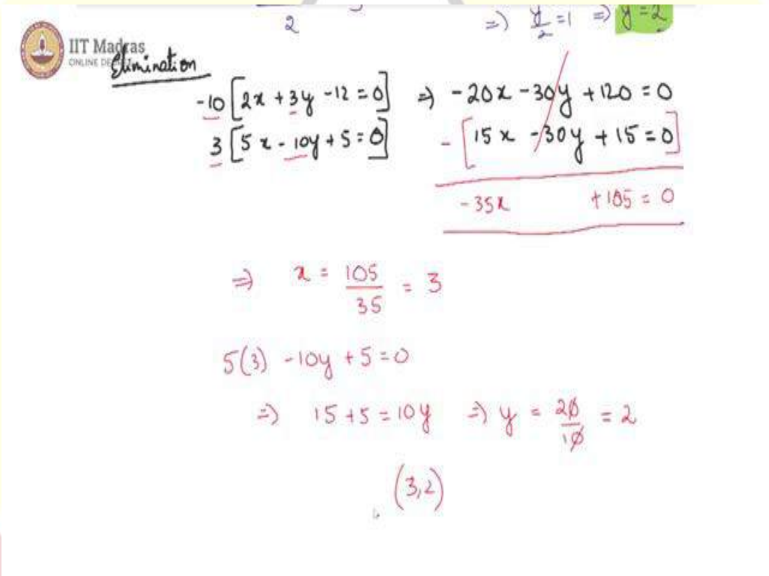
So, for our examples, let us take l1 is $2x + 3y - 12 = 0$, whereas $5x - 10y + 5 = 0$. So, when we have these 2 line equations, how do we solve for x and y. So, the best thing to do is to eliminate one variable, either x or y and get a single equation in the other variable. So, what I mean by that, and this could be done in 2 ways. One way is called substitution. In substitution, in order to remove one variable, we basically express the other in terms of it.

For example, if I wanted to eliminate the y variable, what I do is I express x in terms of y. So, I get all externs on 1 side, so $2x$ is on one side, and the other terms non x terms on the other side, which will give me $12 - 3y$. This would then indicate that x is $\frac{12-3y}{2}$, and then I take this representation of x in terms of y, and substitute it into this equation. What that gives us is, suppose I substituted it, now I will get $5\left(\frac{12-3y}{2}\right) - 10y + 5 = 0$.

So we get $30 - \frac{15y}{2} - 10y + 5 = 0$. That is essentially taking the y common I am going to get $-\frac{15y}{2} + 35 = 0$, canceling off the 35, so I get 1 here, 1 here, that would indicate $\frac{y}{2} = 1$, this implies $y = 2$. So because we eliminated the x here, we got an equation which is entirely in y, which lets us solve for y, and we get the value of y.

Now, to obtain x, we simply have to substitute this value of y in this representation of x, so we will $x = \frac{12-(6)}{2} = 3$. Which means the solution for these 2 line equations is $(3, 2)$, $x = 3$ and $x = 2$. And we can verify this quite immediately by substituting these values into the equations, I will get $2(2) + 3(1) - 12 = 0$. Likewise, $5(3) - 10(2) + 5 = 0$. So it is fairly clear that $(3, 2)$ is the solution which satisfies both linear equations.

(Refer Slide Time: 4:43)

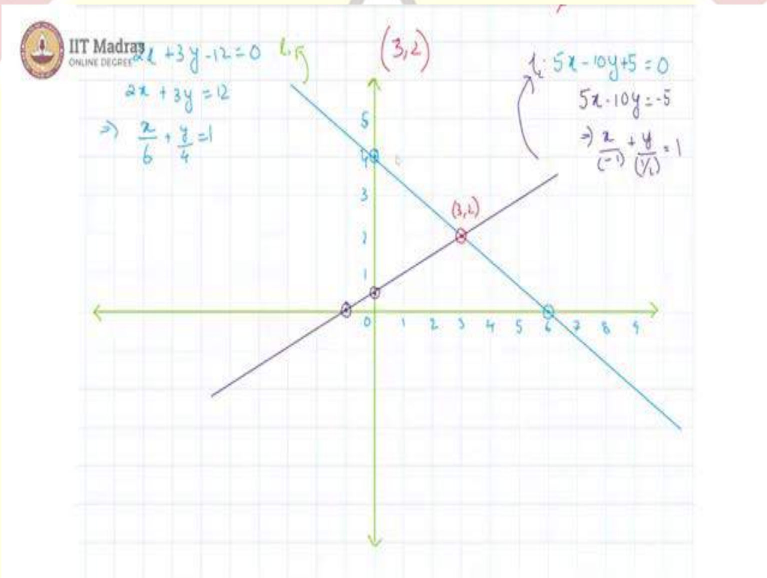


Another method of doing the same thing, which is to solve these 2 equations, we call it elimination. And in elimination, what we do is we again, take these 2 equations, which is$2x + 3y - 12 = 0$, and $5x - 10y + 5 = 0$. We again choose to eliminate either of these variables, because we earlier eliminated x and got an equation in y, now I am going to eliminate y and get an equation x. And for that, what we do is, we multiply this entire equation by the y coefficient in this equation, which is minus 10.

So, I am going to multiply this whole thing with minus 10. And we multiply this entire equation with the y coefficient here in the other equation, that is 3. What that will give us is this would give us minus $-20x - 30y + 120 = 0$. And this gives us $15x - 30y + 15 = 0$. And now what is to be observed is this is $-30y$ and this is also $-30y$, because here we multiply 3 with -10 and here we multiplied -10 with 3.

And that lets us cancel these off, if I subtracted this whole equation from the previous one now. So that will result in $-30y$ by $-30y$ getting canceled, and here, I will get $-35x + 108 = 0$. And this would indicate that $x = \frac{105}{35} = 3$. And now I can substitute $x = 3$ in either of those equations. If I substituted in the second one, I would get $5(3) - 10y + 5 = 0$, this indicates $15 + 5 = 10y$, which gives us $y = \frac{20}{10} = 2$. So, we got our value back, the point back, which is $(3, 2)$. This is the point of intersection of these 2 lines.

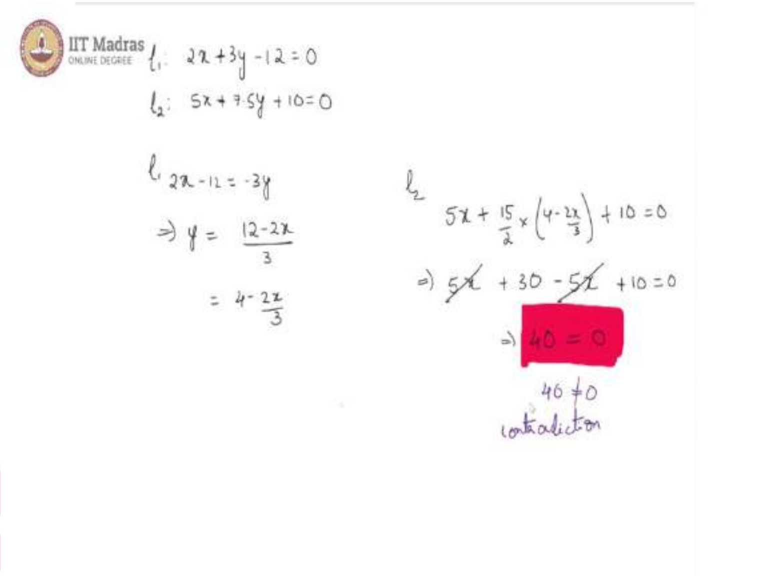
(Refer Slide Time: 7:41)



So, if we plotted these, these are our line equations, let us take the first one, I will reduce this to intercept form, which will have to be to $2x + 3y = 12$ is going to give us $\frac{x}{6} + \frac{y}{4} = 1$. So, x intercept is going to be 6, this and the y intercept is going to be 4, which is this and so our line is this is our l1. Now, if we try to plot the other equation, here, again, I will get $5x - 10y + 5 = 0, \frac{x}{-1} + \frac{y}{1/2} = 1$.

So, here we have this is the x intercept, whereas this is the y intercept 0.5 here. So, this is our line equation 2. And clearly the intersection is happening here at this point, which is you can see this is (3, 2). So, in this way, you can try to find the point of intersection of any 2 given lines. However, you are likely to run into a bit of trouble in 2 cases, and let us see those 2 cases.

(Refer Slide Time: 9:21)

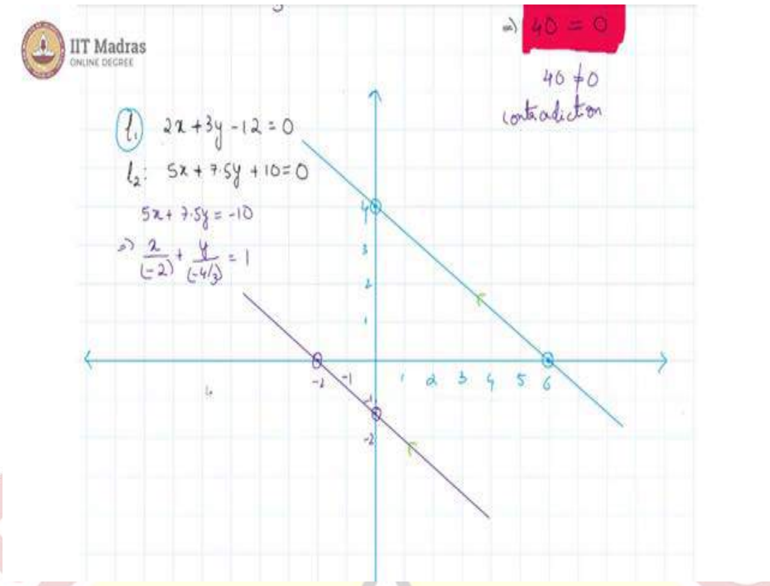


Consider these 2 line equations, 11 is still $2x + 3y - 12 = 0$, whereas $5x - 7.5y + 10 = 0$. If we try to solve this using the substitution method, for example, we would get, I would, let us say I try to eliminate the variable x in which case I should be doing to $2x - 12 = -3y$, which would indicate $y = \frac{12-2x}{3} = 4 - \frac{2x}{3}$. And substituting this in l2, I will get from l2, this is from l1.

And now in l2, if I substituted this, I would get $5x + \frac{15}{2}\left(4 - \frac{2x}{3}\right) + 10 = 0$. This gives us $5x + 30 - 5x + 10 = 0$. And you see that $5x$ and $-5x$ cancels and we come at the strange contradiction where $40 = 0$. And this is not okay right. We know that $40 \neq 0$. So, there is some contradiction we are arriving at.

And what does this contradiction indicate? It indicates that there is no point for which these 2 lines meet. So, you cannot find a point of intersection for these 2 lines. So why is that? That is because they are parallel. If we plotted these lines,

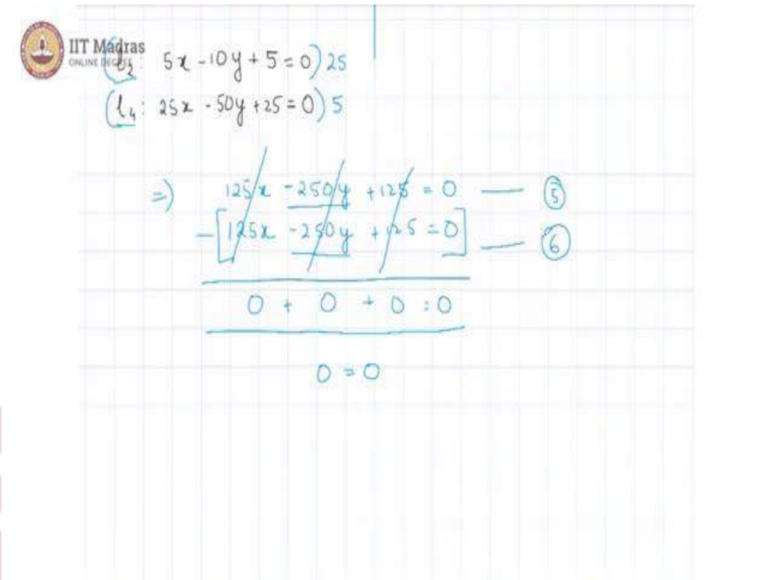
(Refer Slide Time: 11:51)



We know that for l1, the intercepts are 6 and 4, respectively. So this is 1, 2, 3, 4, 5, 6. So this is our intercept for l1, x intercept for l1 and y intercept for l1 is 4 1 ,2, 3 and 4. For l2, we have to see now for l2, we get $5x - 7.5y + 10$, which indicates $\frac{x}{(-2)} + \frac{y}{(-4/3)} = 1$.

So, in $x = -2$, so this would be our point and in $x = -\frac{4}{3}$is a little below -1, which is about one third the way from -1 and -2. So, this would be it. If we plotted these lines now we see that these are, in fact, parallel lines. They just do not meet anywhere, which is why when you try to solve for a point of intersection, you get a contradiction. So here, we can say that there is no solution for this system of linear equations.

(Refer Slide Time: 13:30)



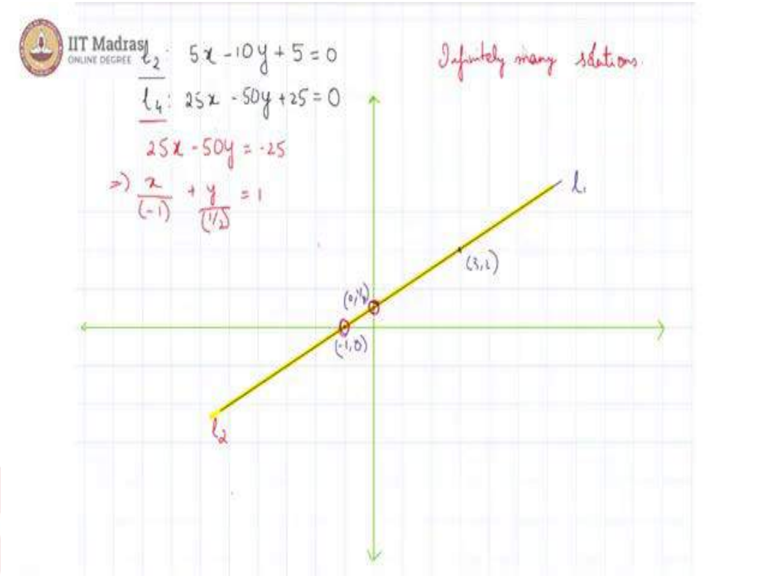
Now, in the third case, let us look at a line equation which is our l2 earlier that was $5x - 10y + 5 = 0$. And there is some other equation l4 let us call it, which is $25x - 50y + 25 = 0$. So, when we solve for these 2 equations, now let me try the elimination method. So, I am going to get 2 equations, then one is $125x - 250y + 125 = 0$. And here I am going to get another one, $125x - 250y + 125 = 0$.

We have the same coefficient for y. So if I attempted to subtract this equation entirely, I will get 0. So, I have this statement, which is always true. Unlike the previous case where it was never true, 40 was never going to be equal to 0, here I get a statement, which is always true, which is $0 = 0$, independent of the coordinates of x and y.

And this means something similar to the previous case, but not exactly the same. What is happening here is since this is always true, it means there are infinite solutions for these 2 equations. If you observe what is actually happening is l2 and l4 are the same line, which is why we got this entirely identical equations, both of these, let us call this equation 5 and let us call this equation 6. And we see that equation 5 and equation 6 are the same, there is no difference, which means our 2 original lines are coinciding.

If they are the same line, then we will get infinitely many points which satisfy both of them. So we have infinitely many solutions for these 2 lines. So whatever x you take, you are going to get a solution for that x. So in the graph, this is what is going to look like.

(Refer Slide Time: 16:28)



We know the intercepts of our l2, which is -1 and y intercept was half, so this would be our l1, it is passing through $(-1, 0)$, and also $(0,\ 1/2)$. And as we had found earlier, it is passing through $(3, 2)$ as well. Now let us consider the other equation. Now let us consider the other equation which is l4, and we will have $25x - 50y = -25$. This gives us $\frac{x}{(-1)} + \frac{y}{(1/2)} = 1$.

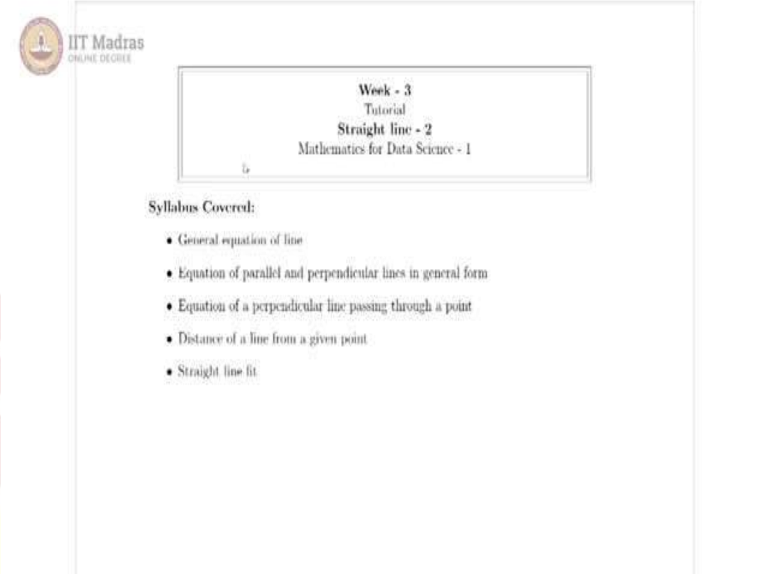
So, again we get the same intercepts. Thus, l2 will have to coincide entirely with l1. And that is what is happening, they are the same line. So, we get infinitely many solutions when we get a true statement, an always true statement independent of x and y in case of the same line, that is both line equations are representing the same line.

IIT Madras

ONLINE DEGREE

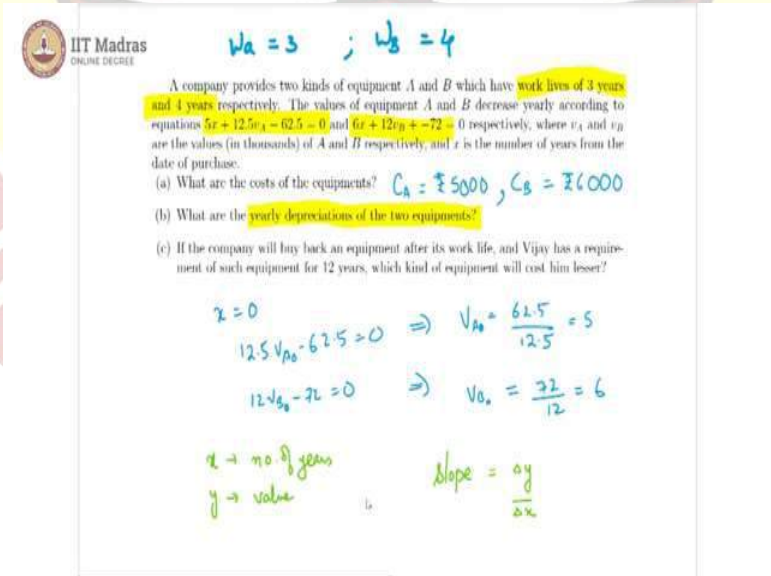
Mathematics for Data Science 1
Week-03
Tutorial-01

(Refer Slide Time: 0:16)



Hello, mathematics students. This is a tutorial for week 3, where we will be doing more straight line concepts problems. Primarily, this is the syllabus that has been covered here. Let us begin with our first question.

(Refer Slide Time: 0:31)



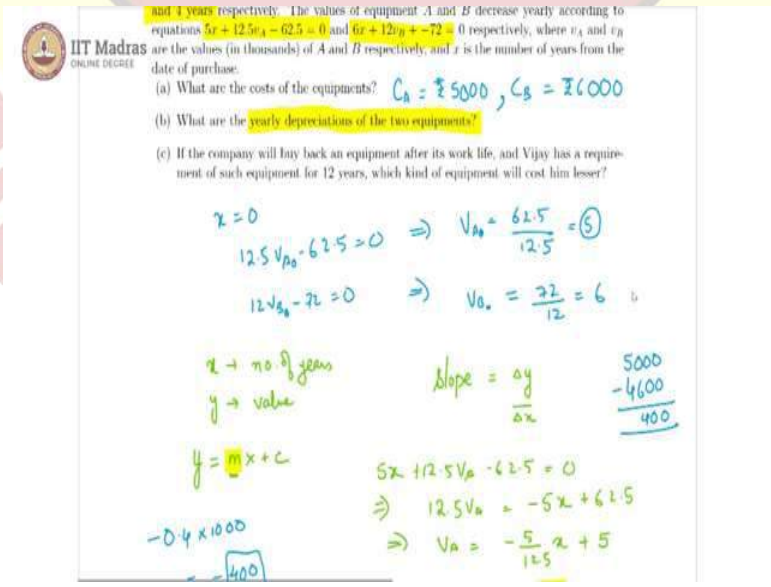
There is a company with two kinds of equipment, A and B. And they have work lives of 3 years and 4 years respectively. So, work life of A is, let us call it W_a is 3, W_B is 4 years. Further, the values of equipment A and B decrease yearly according to these equations. These are our equations, where v_A is supposed to be the value of A and v_B is supposed to be the value of B in thousands, respectively, and x is the number of years for which that value is applicable.

So, what are the costs of the equipments? So, the cost of the equipments would be v_A and v_B values when x is equal to 0, that is, when you just bought it, what is the value of the equipment. So, we just take x is equal to 0 and from this we get $0.5v_A$ - 62.5 = 0, this would give us v_A is equal to, to indicate that this is the initial time I am going to make it A_0, v_{A0} so yes, this is v_{A0} and that is 62.5 /12.5, which is equal to 5.

Therefore, the cost of A, I will call it C_A is rupees 5000. Now, let us work with B. Same thing again, we take x is equal to 0. So, we have $12v_B$ - 72 = 0, this will imply v_B again we are calling v_{B0} to indicate the initial cost that would be 72 /12 which is equal to 6. So, C_B, the cost of B is rupees 6000. Going further, we are asked what are the yearly depreciations of the two equipments.

So, yearly depreciation basically means how much value is decreasing each year. So, let us look at that. Here, in this case, x is number of years, whereas y is the value. So, what is being asked in a yearly depreciation is the change in y for a unit change in x, which is basically just a slope. Because slope is changing y, Δy by changing x. So, when Δx is equal to 1, Δy is equal to the slope.

(Refer Slide Time: 3:56)



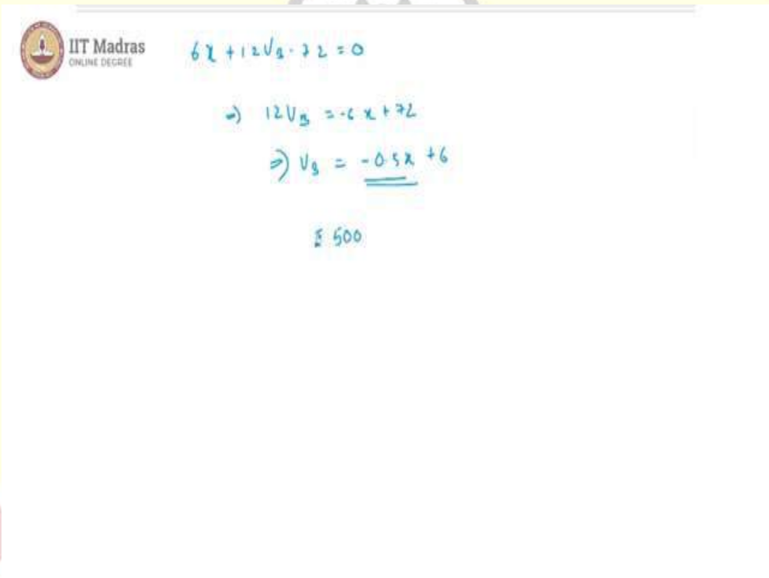
So, we can find this by just finding the slope for each of those two linear equations. And for the slope, we convert our equations to the y = mx + c form, then the m is going to be the slope.

So, one equation is 5x + 12.5v_A - 62.5 = 0. This would indicate that 12.5v_A = -5 x + 62.5. Going further then, it will have v_A = -5x / 12.5 + 62.5 /12.5, we had already seen it to be equal to 5.

So, that is equal to - 0.4 x + 5 = v_A. So here, we are, our m in the equation is basically -0.4. So, this is the reduction in one year, -0.4 ×1000 because we are taking everything in thousands, so, that is basically -400. So, this is the depreciation, 400 is the depreciation every year for the company one, we can also verify this by looking at the values of v_A for year one.

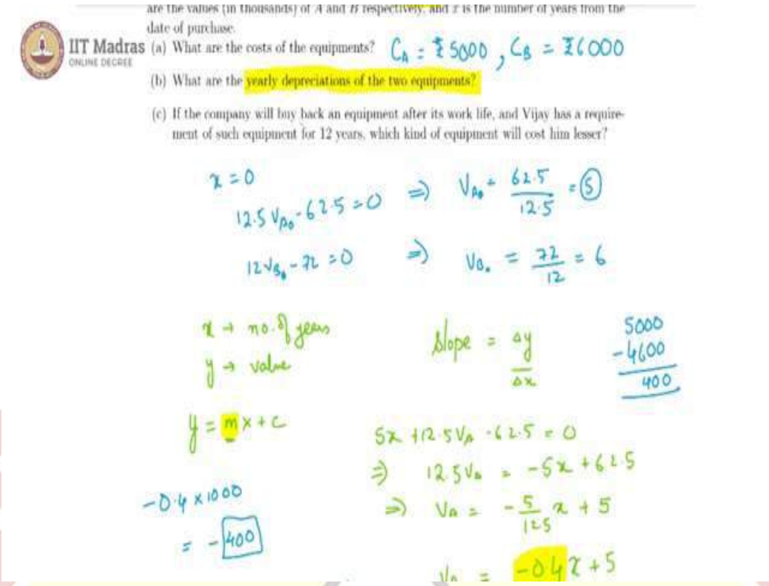
So, when x = 1 we have 5 +12.5v_A - 62.5 = 0, this gives us v_A = 57.5 / 12.5 which is equal to 4.6. So, v_A was originally 5, that means it was originally 5000 rupees and after 1 year it became 4.6 which is 4600 rupees. So, the difference is 400 rupees. So, that is the yearly depreciation for the first equipment.

(Refer Slide Time: 6:48)



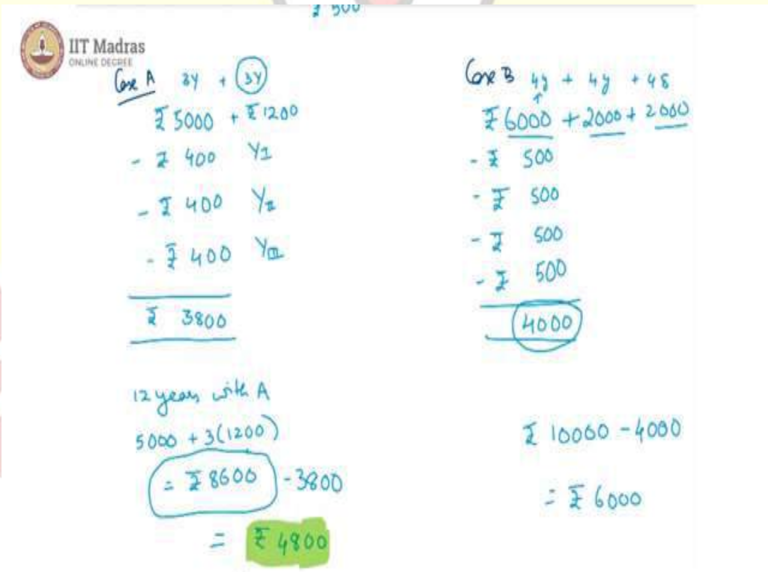
Now, let us look at the second equipment now second equipment the equation was 6x + 12v_B - 72 = 0. Again, if we put this to the y = mx+ c form, the slope intercept form we will be getting first we have to do 12v_B = -6 x + 72. This indicates v_B = -0.5 x + 6, thus -0.5 is the slope here. Which means 500 rupees is the yearly depreciation.

(Refer Slide Time: 7:42)



In the last part, they said that the company will buy back the equipment after its work life. And Vijay has a requirement of such equipment for 12 years. Which kind of equipment will cost him lesser.

(Refer Slide Time: 7:58)



So, in the case of the first equipment, let us call it case A, and here let us have case B to consider. And in case A the initial cost was 5000 rupees and each year there is a decrease of 400 rupees. So, in first year we lose this much, in the second year we lose another 400 rupees and at the end of the third years, there is a loss of another 400 rupees. And we are aware that 3 years is a worklife for A, whereas for B it is 4 years. This is to say that at the end of 3 years, the value of the machine is 3800 rupees.

So, if now, Vijay buys the equipment afresh, then and the company is buying back this 3800. All that Vijay needs to spend now is rupees 1200 and this way he gets an additional 3 years. So, with 5000 he got 3 years and now another 3 years this way. So, in order to get 12 years with equipment A, the total money that Vijay will require to spend is 5000, which is the initial first 3 years and from then on 3 years plus 3 years plus 3 years because it is totally 12 years.

So, 3 times 1200, that is rupees 8600 in case of A, whereas in B, B is more expensive. So, we have 6000 and every year there is a loss of 500 rupees in value and this is required to be done 4 times because the work life for B is 4 times. So, we are effectively subtracting 2000 rupees from the original value. So, we have 4000 at the end of it, which means for the first 4 years there is an expenditure of 6000 but then, for the remaining 8 years, there has been only 2000 each.

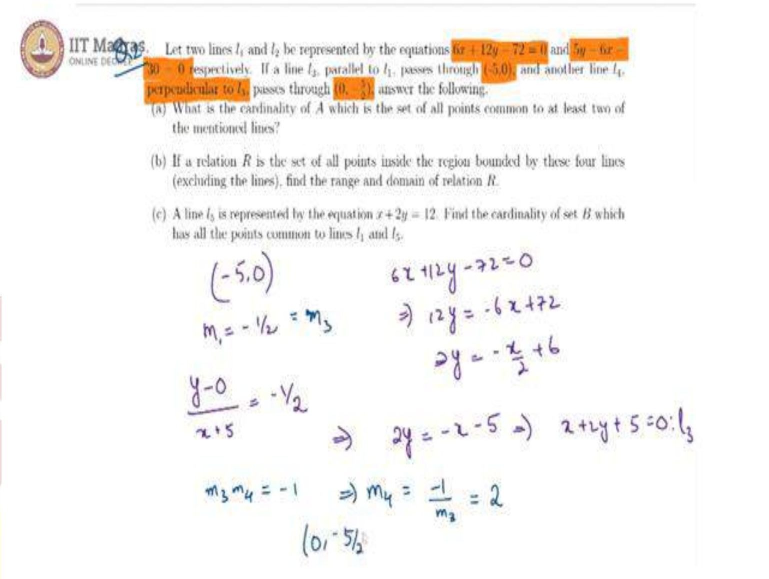
This is so because the product's value is already 4000 rupees and in order to get a new version of equipment B, Vijay only has to spend 2000 rupees. So, total expenditure in this case is going to be 10,000 rupees because 6000 + 2000 + 2000. Here, we are not supposed to forget one thing though, that is the end of, after these 3 years, 3 years pass, at the end of 12 years, he can sell it off for 3800. So, we are supposed to further subtract 3800 here and likewise here, we can sell it off for 4000. So, here we get rupees 4800 whereas, here we get rupees 6000. So, the expenditure is clearly lesser for A. So, A would be the good choice for Vijay.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week-03
Tutorial-02

(Refer Slide Time: 0:16)

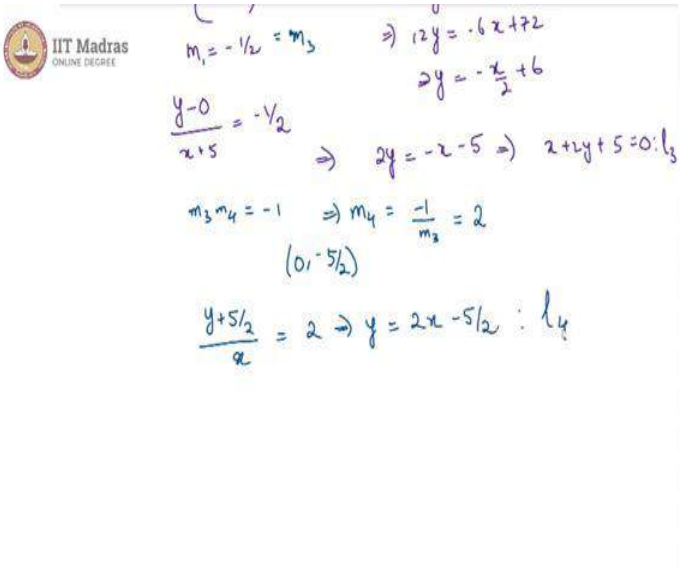


And for our second question, there are 2 lines, and these are the equations, which represent our lines. And a line l_3 is parallel to l_1, and passes through (-5, 0). Now we can find l_3, by using the point slope form, we already have the point, which is (-5, 0). And we can also find the slope from l_1 slope, we already have l_1. And we can write, so l_1 is this, $6x + 12y - 72 = 0$, which tells us that $12y = -6x + 72$.

And that gives us $y = -\frac{x}{2} + 6$ so, the slope here is $-\frac{1}{2}$, because $y = mx + c$. So, slope is $-\frac{1}{2}$. Now if we did point slope form on this, we would get $\frac{y-0}{x+5} = -\frac{1}{2}$ which indicates $2y = -x - 5$. So therefore, $x + 2y + 5 = 0$ is basically our line l_3. And now if we look further, we have line l_4 which is passing through this point, and it is perpendicular to l_3.

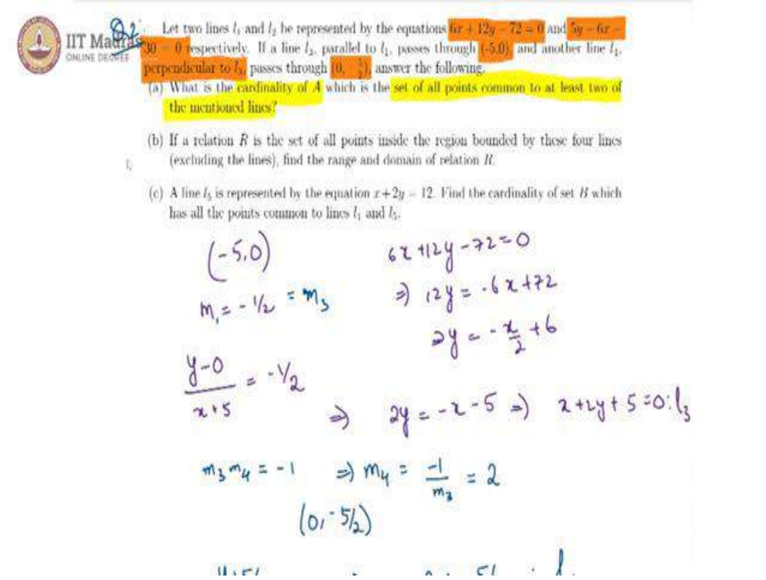
So, if we took this to be $m_1 = -\frac{1}{2} = m_3$ because m_1 and m_3 are the same slope. And let us consider the slope of l_4 to be m_4, so we can say $m3 \times m4 = -1$, because they are perpendicular, that would indicate $m_4 = -\frac{1}{m_3}$, which is basically 2. So we now have the slope of l_4. And it also goes through this point.

(Refer Slide Time: 2:57)



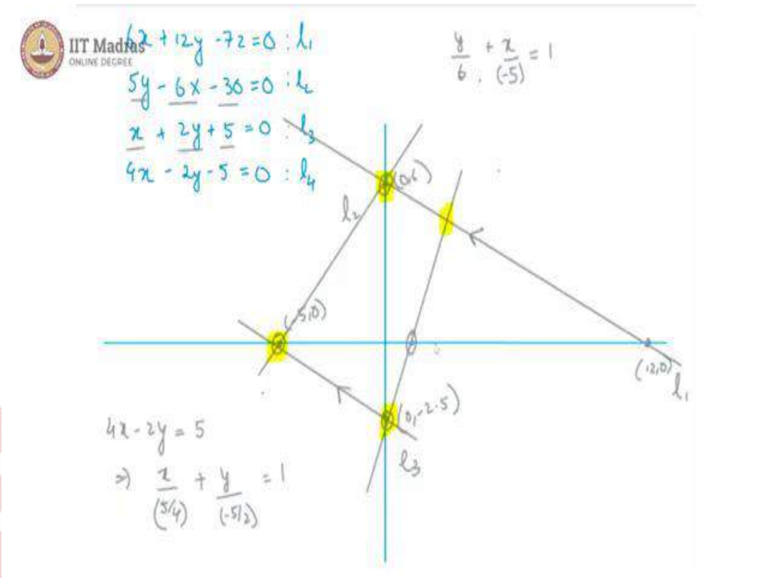
So again, using point slope form, we have $\frac{y+\frac{5}{2}}{x} = 2$, that would indicate $y = 2x - \frac{5}{2}.$ So this is our l_4.

(Refer Slide Time: 3:14)



Now, the question is being asked is, what is the cardinality of A, which is a set of all points common to at least 2 of the mentioned lines.

(Refer Slide Time: 3:28)



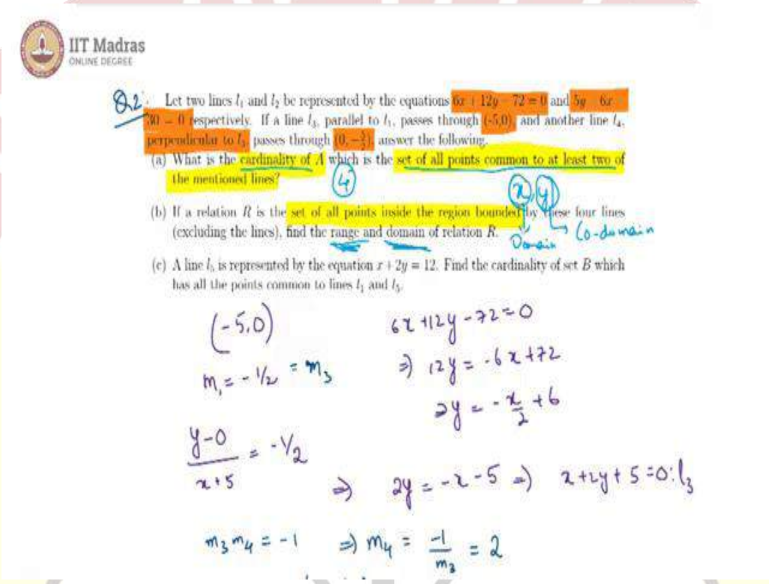
For that, let us try to draw our lines on the graph. $6x+12y-72=0$ would give us if $x=0$, it gives us $y=6$ which means some point let us call this here is (0,6), it goes through this point. And if $y=0$, you get $x=12$. So that would be some point here. So, our l_1 is this line. And now we know l_3 is parallel to this line. So l_3, if we, again did the same thing of putting $y=0$, x becomes -5, which is somewhere here.

So, as you can see, I am doing this on a rough estimate. I am not trying to be accurate, but even a rough estimate can work out here, because you might not always find graph paper when you require it. So often developing an intuition for the rough estimates is a good idea to solve problems. Now, this is one point and when $x=0$, y becomes -2.5, which is somewhere like this. So we have (0, -2.5). As you can probably see from our last rough estimate itself that these do appear to be parallel, they seem to be in the same direction.

Now, l_2 if we look into it with a similar logic, we can see that l_2 can be reduced to $\frac{y}{6}-\frac{x}{5}=1$. So in our intercept form, we can now tell that if I made this plus, this becomes -5, so the x intercept is -5, which is this point, again, and y intercept is 6, so that is this point. So, l_2, in fact, passes through these 2 points. So, this is our l_2. So, this was l_1 now, this is l_3 and this is l_2.

Lastly, let us reduce our l_4 into the intercept form, we get $4x - 2y = 5$, therefore, $\frac{x}{5/4} + \frac{y}{-5/2} = 1$. So, when we look at this then 5/4 is a quantity just a little greater than 1, so it is probably somewhere here and 5/2 is a 2 and a half basically. So, -2.5, so this and this plus we have something like this happening. So, overall there are four points, which are common to any pair of these four lines.

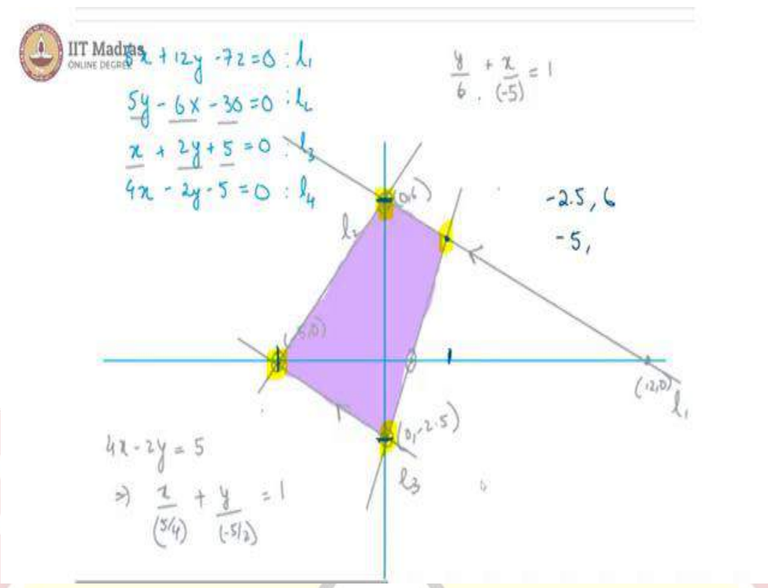
(Refer Slide Time: 7:02)



So, our question, the cardinality of A, where A is a set of all points common to at least 2 of the mentioned lines. So, that would be 4, there are 4 points of intersection here. Now, if R is a relation, and it is the set of all points inside the region bounded by these 4 lines. So, here we are, when we say relation, we are basically saying every point in the set when is taken as a ordered pair like this (x, y), then x would be from the domain of the relationship and y would be from the co-domain.

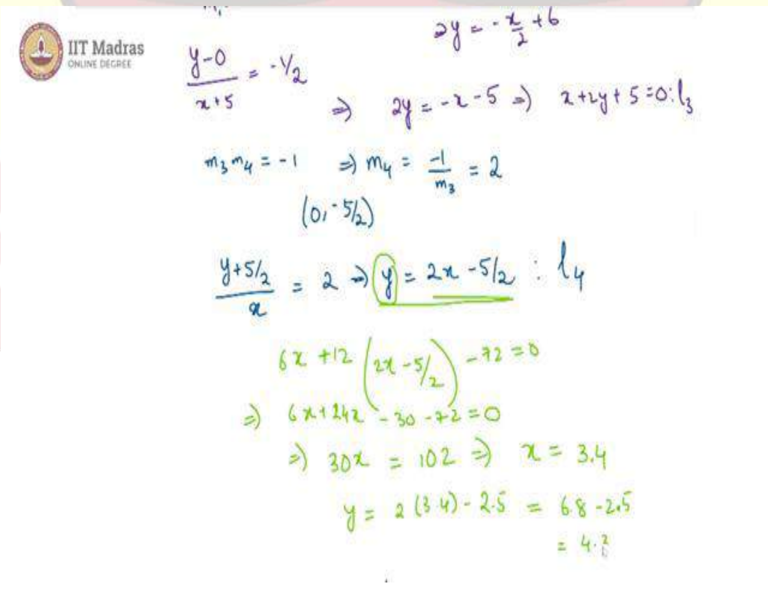
So, this is seen as a relation from the set of x values and to the set of y values. And now, we are asked to find the range and domain of relation R, which is to basically find when we say range, all the possible y values and the domain is all the possible x values.

(Refer Slide Time: 8:16)



So, here in this region that we are looking at, the possible y values would be between this value and this value. So, all possible y values are between -2.5 and 6, whereas the possible x values are between this point and this point, that is between -5 to some particular quantity, which is the x coordinate of this point. And that point is the intersection of l_1 and l_4. So, let us try to solve l_1 and l_4 to find that point of intersection.

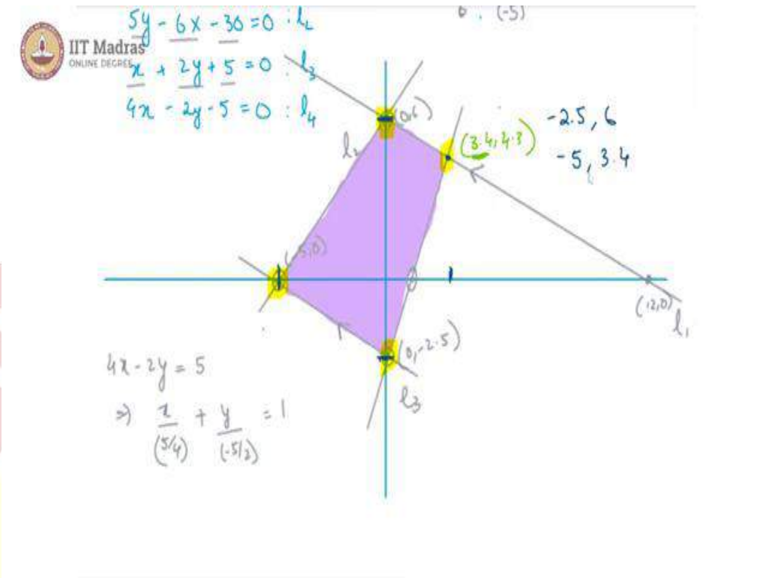
(Refer Slide Time: 9:03)



We know that this as l_1, and this is l_4 and from l_4, we know that y is basically $2x - \frac{5}{2}$. If we substituted this into l_1 we would get $6x + 12(2x - 5/2) - 72 = 0$. This would give us

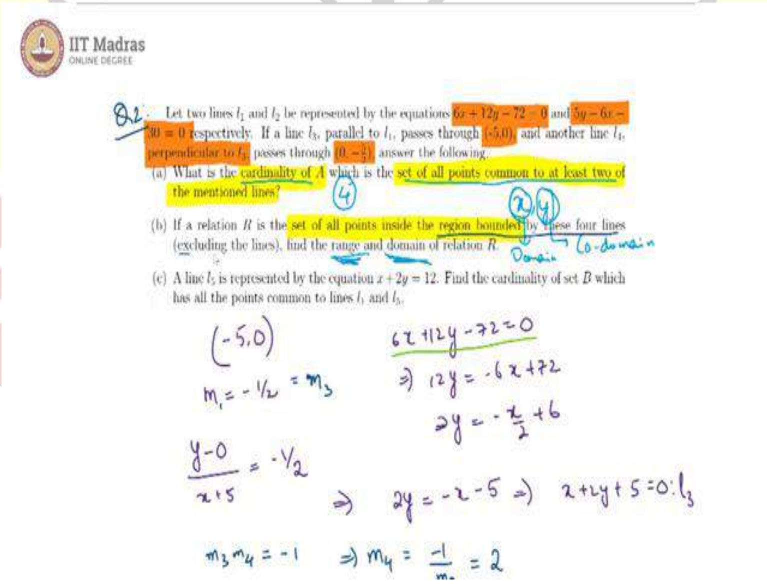
$6x + 24x - 30 - 72 = 0$. That indicates $30x = 102$ which indicates $x = 3.4$. Correspondingly, y would then be $2 \times 3.4 - 2.5$, because 5/2 is 2.5, which gives us 6.8 - 2.5, which is equal to 4.3.

(Refer Slide Time: 10:24)



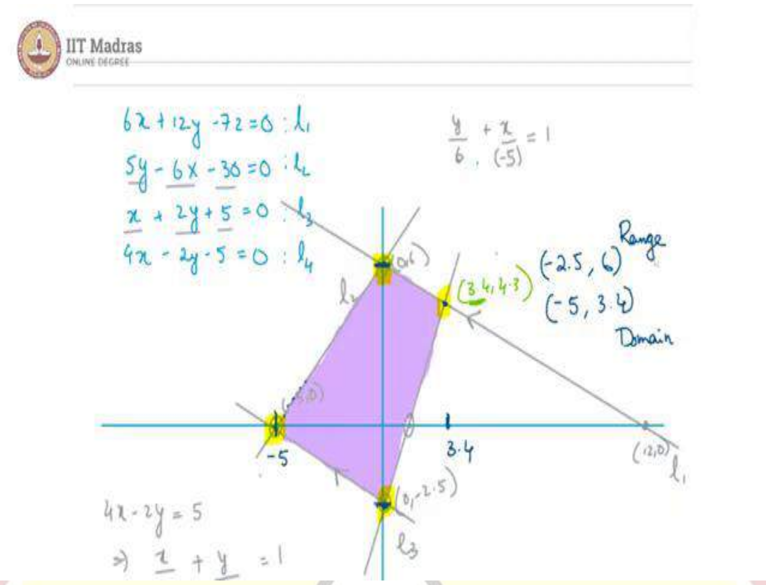
So this point here is (3.4, 4.3) and we only require the x value. So the x values range from -5 to 3.4.

(Refer Slide Time: 10:42)



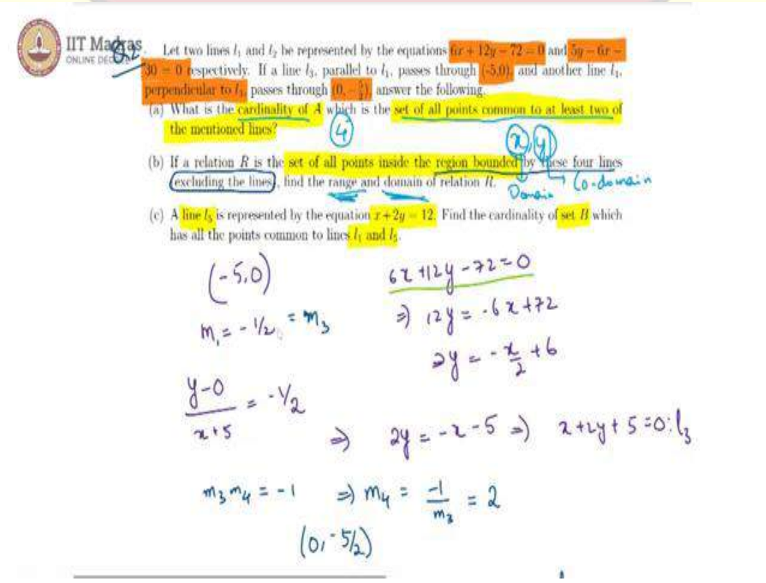
However, one important thing we need to look for here now is the region bounded by these 4 lines, but excluding the lines themselves.

(Refer Slide Time: 10:52)



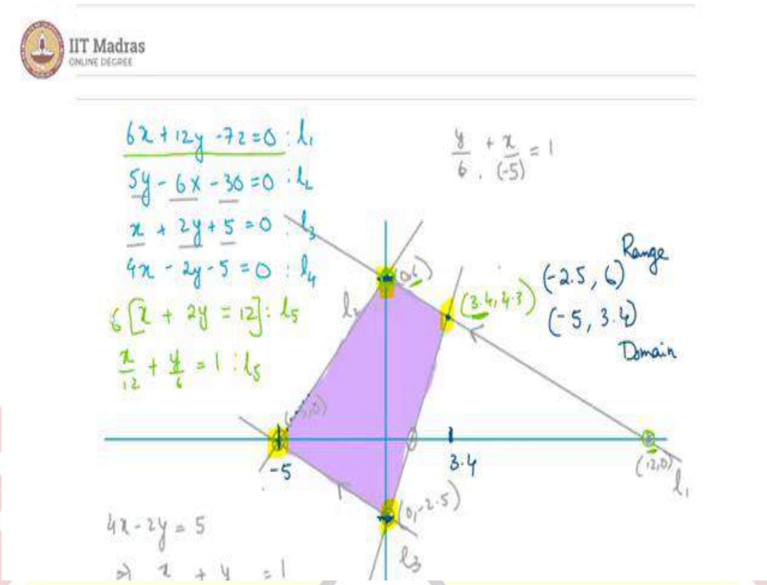
Which means -2.5 and 6 themselves do not fall into our domain because we are not interested in the points on the curve. So this point is on the curve, this point is on the curve, but it is not inside, similarly, for each of these, because they are the border points. So, -5 is not an x value inside the domain. Similarly, 3.4 is not a value inside the domain. So, our domain is the $(5,3.4)$. Likewise, -2.5 is not a y value inside the range and 6 is also not a y value inside the range, so our range is (-2.5, 6).

(Refer Slide Time: 11:50)



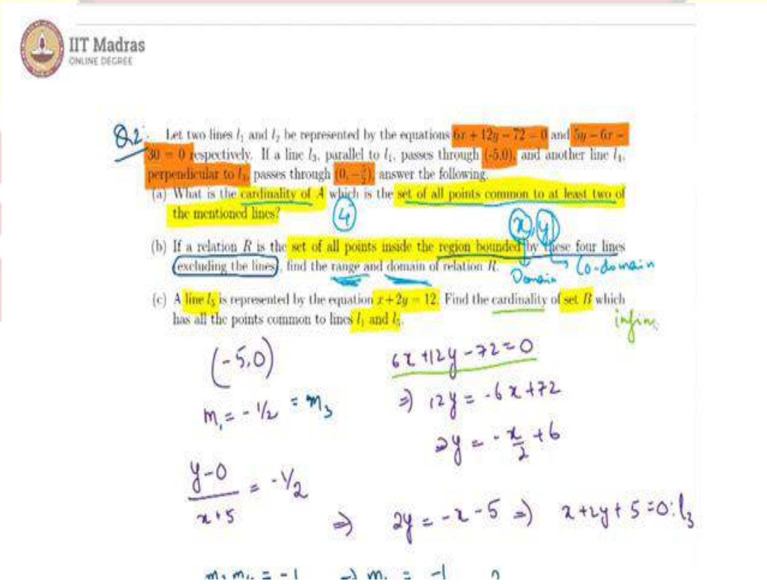
Lastly, there is a line l_5 represented by this equation given to us find the cardinality of set B, which has all the points common to l_1 and l_5.

(Refer Slide Time: 12:04)



Let us look at l_1 and l_5. l_5 is given as $x+2y=12$. Now, if we applied our intercept form again, we would get $x/12+y/6=1$. Let us look at that $x/12$ indicates x intercept of l_2, $y/6$ indicates y intercept of 6. So, we see that l_5 is basically the same line as l_1, indeed if you multiply this whole equation with 6, you will just get the form of $l_{1.}$ Therefore, l_1 and l_5 are the same lines.

(Refer Slide Time: 12:54)



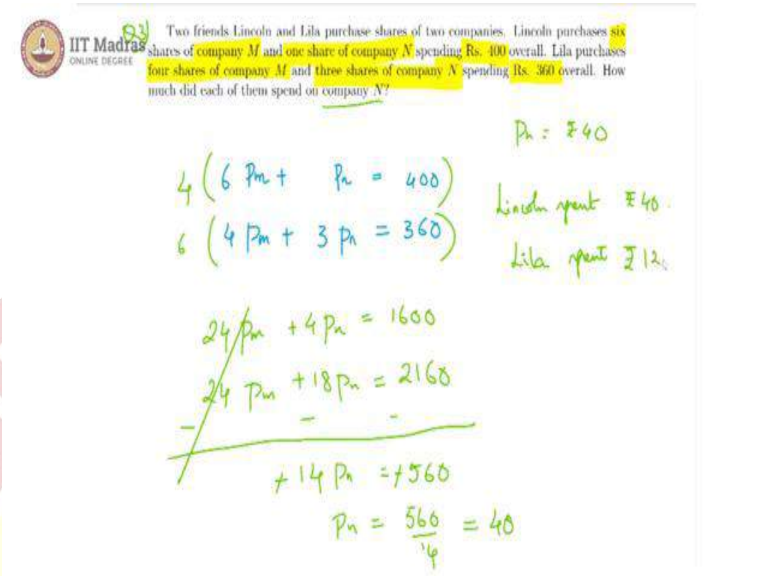
Then, the question is asking, find the cardinality of set B, which has all the points common to the lines l_1 and l_5. There are infinite points because they are the same line. So, the cardinality of set B is infinite.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week-03
Tutorial-03

(Refer Slide Time: 0:14)



Now, third question, you have two friends Lincoln and Lila who purchase shares of two companies. Lincoln purchases six shares of a company M and one share of company N and overall spends 400. This can be encapsulated as if the company M's share price is P_m and for n that is P_n we can say that $6P_m + P_n = 400.$ Then for Lila there is four shares of Company M and three shares of Company N coming to 360.

So, for Lila we have $4P_m + 3P_n = 360.$ How much did each of them spend on n? So, we need to know what is P_n and $3P_n$, that is what we are interested in. To find the values of P_m and P_n we will require to solve these two linear equations. However, we only required to find P_n because the question is only pertaining to the company N's shares. So, we can work towards eliminating the P_m variable from these two equations.

So, we can multiply this equation by 4 and this one by 6 because $4 \times 6 = 24,\ 6 \times 4 = 24$ and that way we should be able to subtract $24P_m$. So, we are going to get from the first equation $24P_m + 4P_n = 1600,$ whereas, from the second equation we get $24P_m + 18P_n = 2160.$ Now, if we subtract second equation from the first we get these two canceling off and here we get $-14P_n = 560.$

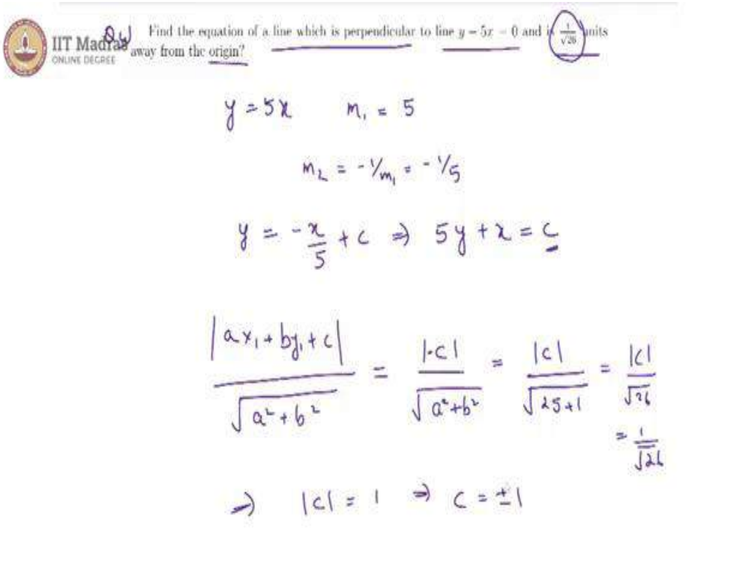
And this indicates that $P_n = 560/14$ because we can cancel out the plus and the plus and that is equal to 40. So, P_n is 40 rupees per share. And now since Lincoln has purchased only one share, Lincoln spent only 40 rupees on company N, whereas, Lila spent three times that which is rupees 120.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week-03
Tutorial-04

(Refer Slide Time: 0:15)



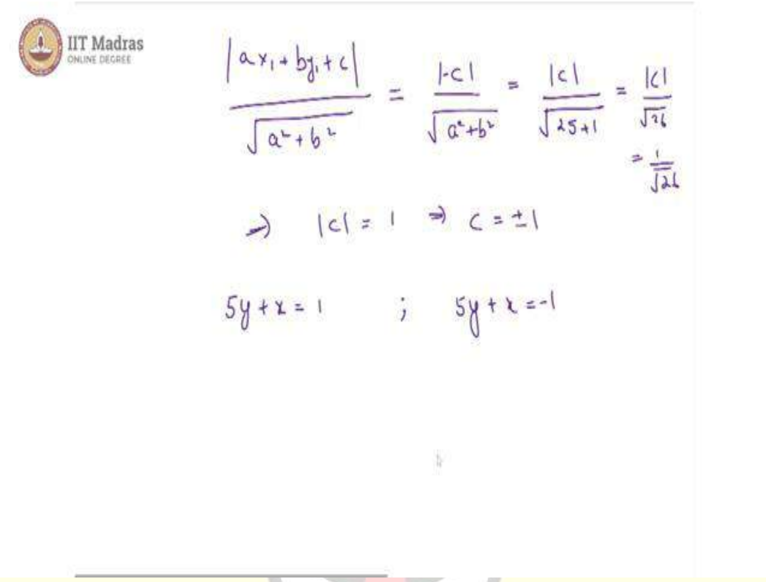
For our fourth question, we want the equation of a line which is perpendicular to this line, and is at this distance from the origin. So, from $y-5x=0,$we get $y=5x,$ so therefore, the slope m_1 is 5. And if our line is perpendicular to it, then our line m_2 must be $-1/m_1$,which is equal to -1/5. So, we know that our line is some $y=-\frac{x}{5}+C.$

If we kind of simplify it, we are going to get $5y+x=C,$ this C is not the same thing as the previous C, I have just used that as C because it is an arbitrary constant, which is yet to be determined, otherwise it should have been 5C. Anyway, now we have to find the value of this C in this equation. For that, we are going to use the next bit of information that is given to us, which is the distance from the origin.

So, this line has this distance from the origin. So the distance from a point formula is $\frac{|ax_1+by_1+c|}{\sqrt{a^2+b^2}}$, where (x_1, y_1) is the point from which we are measuring the distance for this line. So, in our case, (x_1, y_1) is (0, 0) because we are doing from the origin. So in our case, we get modulus of $\frac{|-c|}{\sqrt{a^2+b^2}}$, So, modulus of $|-c|$is just the same thing as $|c|$.

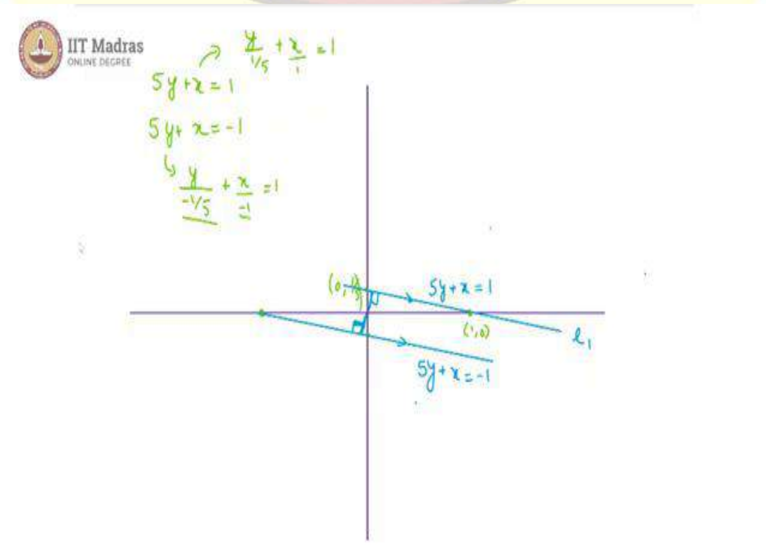
And root of $\sqrt{a^2+b^2}$, in our case comes out to be $\sqrt{25+1}$,, that is $\sqrt{26}$. So we have $\frac{|c|}{\sqrt{26}}$, this is given out to be $\frac{1}{\sqrt{26}}$, which would imply $|c|=1$, and that would imply $c \pm 1$. So, we get two answers.

(Refer Slide Time: 2:49)



What are the two answers? One is for c being +1, we have $5y+x=1$. And in the other case, we get $5y+x=-1$. for the other choice. So, how does this happen, what is actually happening here to try to plot our lines?

(Refer Slide Time: 3:08)



So, we have two lines, which we were looking at, which is $5y + x = 1$ and $5y + x = -1$ and from this we get the intercepts to be, the intercept form of this would be $\frac{y}{1/5} + \frac{x}{1} = 1$. And in this case, we get $\frac{y}{-1/5} + \frac{x}{-1} = 1$. So, in one case, we have a y intercept of 1/ 5. So, let us assume this is 1/5, then x intercept is 1 which is 5 times of that, so that so it must be somewhere here, so this would be (1, 0) and this is (0, 1/5).

And our line is going through these two points, giving us something like this. Let us call this l_1 and where do we get the $\frac{1}{\sqrt{26}}$ distance from the origin, we get it when we measure it perpendicularly from the origin. Now, let us look at the other equation. So -1/5, so, this should be exactly below this this way and this is -1, so this would be exactly opposite in this way at the same distance.

So now we have these two points, so we can also construct this line, which goes this way. And as you can see, they are both parallel and exactly opposite to that $\frac{1}{\sqrt{26}}$ you get this distance which is again perpendicular distance and it is also at $\frac{1}{\sqrt{26}}$. So, we have two lines which satisfy our requirements, one is $5y + x = 1$, the other is $5y + x = -1$.

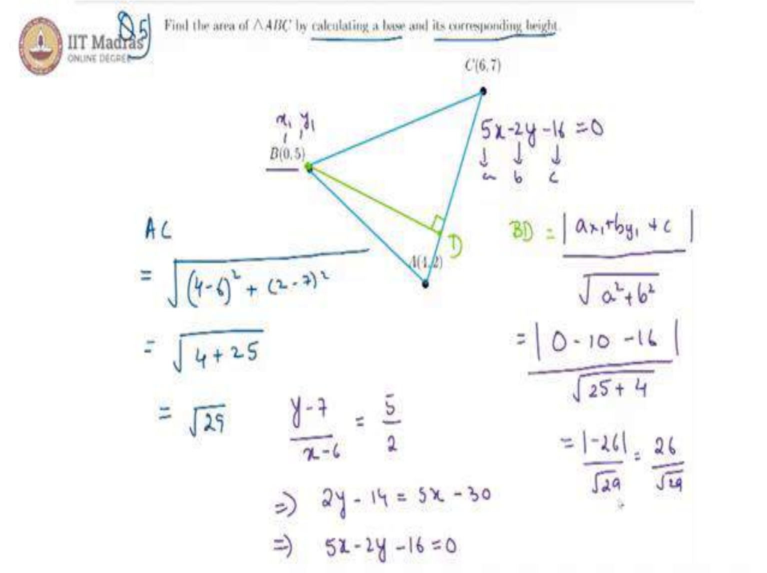
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week-03
Tutorial-05

(Refer Slide Time: 0:14)



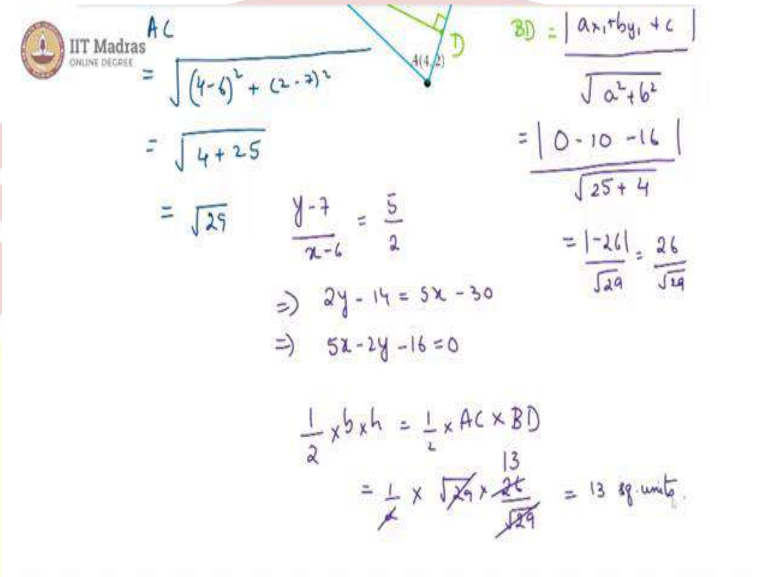
In this problem, should be our fifth question. Suppose to find the area of ABC, there are three points here. So, we need to make that triangle, our triangle would look something like this. But to find the area, we are supposed to calculate a base and its corresponding height. So, we are not supposed to use the formula which involves the three coordinates, instead, we will take any of these sides to be the base. So, let me take AC to be the base. So, we need to find the base length, which is AC, that would be by Euclidean distance formula, $\sqrt{(4-6)^2+(2-7)^2}$.

So, this comes out to be $\sqrt{4+25}$. So, that gives us $\sqrt{29}$ is the base. Now the altitude, the height from B would be something like this, let us call this point D and this is 90 degrees, so B to D that length would be the height. So, BD is going to be the distance of the point B from the line AC, the shortest distance of point B, the line AC. So, for this we can use the distance formula of a point from a straight line. However, we first need to find out the equation of AC.

For that, let us use the 2 point form because we have 2 points, we will get $\frac{y-7}{x-6}=\frac{7-2}{6-4}=\frac{5}{2}=2.5$. Anyway, if we cross multiply, we get 2y-14=5x-30, which gives us the equation to be 5x-2y-16=0 and the distance of (0,5) which is our B from this particular line. So, this line is our 5x-2y-16=0 .

So, that distance can be calculated from the formula, which is the $\frac{|ax_1+by_1+c|}{\sqrt{a^2+b^2}}$. So here a is our 5, b is -2 and c is -16. So, substituting and (x_1,y_1)is our coordinates of B, this is x_1 and this is y_1, so the coordinates of B. So here we get $\frac{|0-10-16|}{25+4}$, which then gives us $\frac{|-26|}{\sqrt{29}}$, $|-26|$ is then 26. So, this would be the height.

(Refer Slide Time: 3:48)



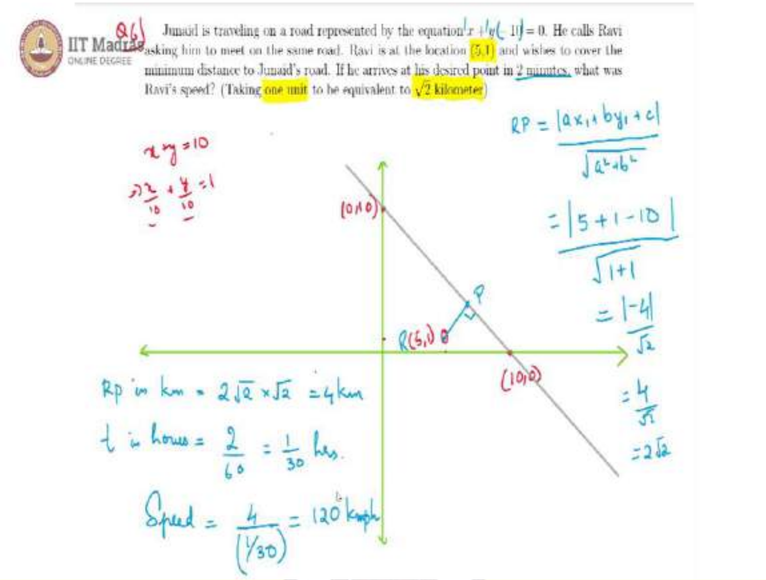
Combining these two quantities, we get our area as half into base into height, which will then be $\frac{1}{2} \times \mathrm{AC} \times \mathrm{BD}$, which then gives us $\frac{1}{2} \times \sqrt{29} \times \frac{26}{\sqrt{29}}$, $\sqrt{29}$ and $\sqrt{29}$ cancels off, 2 cancels with 26 giving us 13. So, we get 13 square units as the area of our triangle ABC.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 03 – Tutorial 06

(Refer Slide Time: 00:14)



Sixth question. We have Junaid who is traveling on a road represented by the equation x+y-10=0. So, in the graph if we plot that, we can see that x+y=10, which gives us $\frac{x}{10}+\frac{y}{10}=1$, which means the x intercept and y intercept are both equal to 10. So, if this is 10 and this is also 10, so this would be (10,0). Whereas this is (0,10) and the line that passes through them is the road that Junaid is traveling on.

So, this is the line that Junaid is traveling on and he calls Ravi, asking him to meet on the same road. But Ravi is at this point. So, that would be 5, it would be somewhere here halfway and 1 would be somewhere here. So, this is 1 and this would become our location of Ravi, that is (5,1) and Ravi wishes to cover the minimum distance to Junaid's road. So, we know that minimum distance is achieved when you go perpendicular that is normal to the other line.

So, we can see that Ravi goes along this path and intersects, that path intersects somewhere over there, let us call this point P and he arrives at this point P in 2 minutes and we are being asked, what is Ravi's speed? So, we need to first find out, assuming Ravi's original location (5,1) is R, if we find out what RP is, then we should be able to find out the speed. So, RP is basically the shortest distance of R from this particular line. So, we can calculate it from that formula, which is $\frac{|ax_1+by_1+c|}{\sqrt{a^2+b^2}}$.

And now, here a is 1, b is also 1, c is -10. So, we have this is equal to $\frac{|5+1-10|}{\sqrt{1+1}}$, because a is 1 and b is 1, the squares are also 1. So, we get $\frac{|-4|}{\sqrt{2}} = \frac{4}{\sqrt{2}}$, which is then equal to $2\sqrt{2}$ units and now it is given that 1 unit is equivalent to $\sqrt{2}$ kilometres.

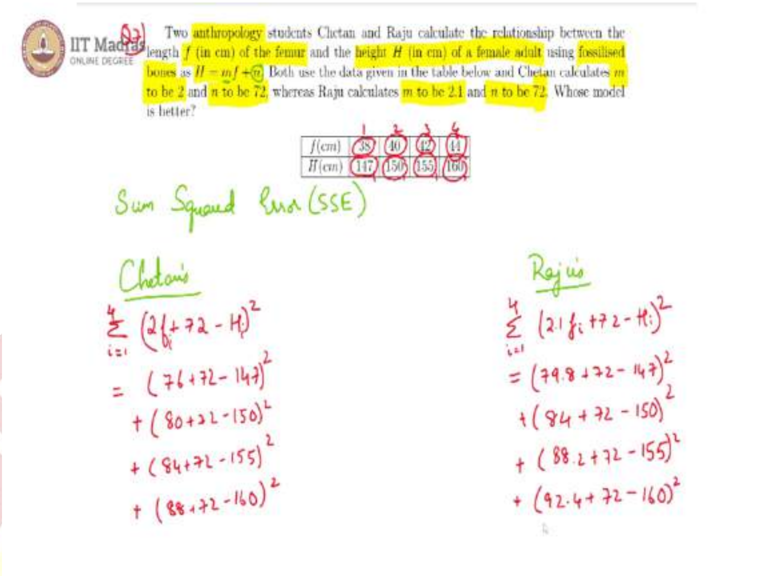
So, that means the distance in kilometre RP in km is equal to $2\sqrt{2} \times \sqrt{2}$, that is 4 km and Ravi has taken 2 minutes. If we write it in hours, t in hours is then $\frac{2}{60}$ that is $\frac{1}{30}$hours. So, the speed would be distance by time, which is $\frac{4}{\frac{1}{30}}$ which gives us 120 kmph, kilometre per hour. This is the speed that Ravi travels at.

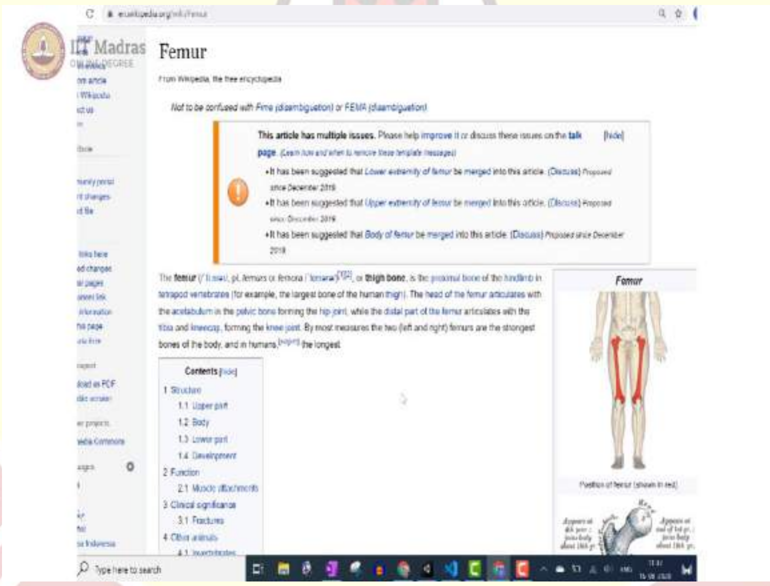
IIT Madras

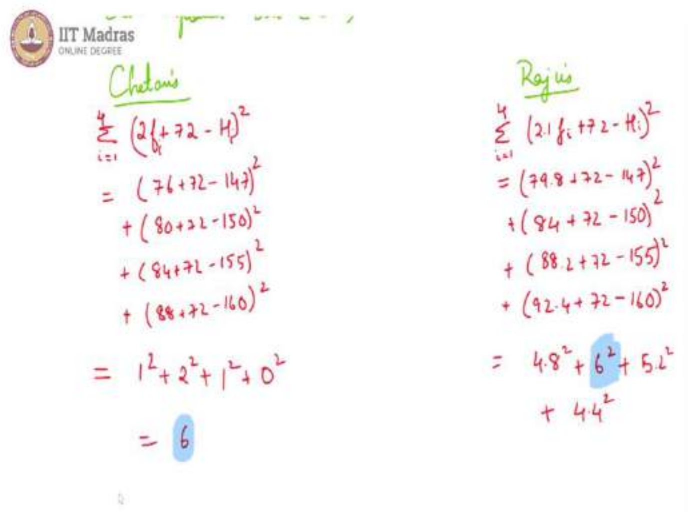
ONLINE DEGREE

Mathematics for Data Science 1
Week 03 – Tutorial 07

(Refer Slide Time: 0:14)







In our seventh question, we have this interesting thing, where there are two anthropology students and they are calculating the relationship between the length f of the femur and the height H of a female adult using fossilised bones. So, what is exactly happening here? What is femur?

From Wikipedia, we can see that the femur is the thigh bone, which is this particular bone. So, what is happening is, in our question, there are fossilised bones and these anthropology students, anthropologists try to study the nature of humans and their societies as they were evolving.

So, here we have fossilised bones and suppose we have the femur of what we know to be a female adult, then we are estimating the height of that female adult from the length of the femur bone, from the thigh bone. So, it is given that this relationship is linear, we have $H = mf + n$. Both use the data given below, so this is the data that is available. We have the femur length and the height of the adult female.

So, from this we are trying to develop this model and Chetan has found $m = 2, n = 72$, whereas Raju has calculated m to be 2.1 and n to be 72. So, both of them agree on n, this parameter is already fixed. It is the m that we are trying to see, whose m is better. So, in terms of linear equation, m is basically the slope of the line. So, how do we do this? We want to use the concept of Sum Squared Error, which we call SSE.

So, in both cases, we are going to look at what is being predicted in terms of height and what is the actual data. So, let s look at case one, let us look at Chetan’s case here and Raju’s case here. In terms of Chetan's case, we would have the $H = mf + n$, where m is 2, so we have $\sum_{i=0}^{4}(2f_i + 72 - H_i)^2$ and we sum it over how many items 1, 2, 3, 4.

So, let us call this f_i, H_i and i goes from 1 to 4 and in case of Raju's measurements, this error would be again, i goes from 1 to 4 and we have $\sum_{i=1}^{4}(2.1f_i + 72 - H_i)^2$. So, I think we just need to do the calculations now. So, let us look at this here, so this is case 1. So, this is case 2, this is case 3, this is case 4, f_1 is 38 and H_1 is 147.

So, when we put in 38 here we get 2 times 38 is 76 + 72 - 147 the whole square and then in case 2, we have 40 and 150 as f_2 and H_2. So, we will get 80 + 72 - 150 the whole square and then we have, f_3 is 42 and H_3 is 155. We have $(84 + 72 – 155)^2$ and lastly, we have f_4 is 44 and H_4 is 160. So, we have $(88 + 72 – 160)^2$.

$$\sum_{i=0}^{4}(2f_i + 72 - H_i)^2 = (76 + 72 - 147)^2 + (80 + 72 - 147)^2 + (88 + 72 - 160)^2$$

So, this is the total sum squared error for Chetan. Whereas in case of Raju, we would get 2.1 times the same thing. So, 2.1 times 38 is 79.8 + 72 - 147 the whole square and in case 2 we get 84 + 72 - 150 the whole square + in 3 we get 88.2 + 72 - 155 the whole square and lastly, in case 4, we have 92.4 + 72 - 160 the whole square. We calculate these values then we get $1^2 + 2^2 + 1^2 + 0^2$.

$$\sum_{i=1}^{4}(2.1f_i + 72 - H_i)^2 = (79.8 + 72 - 147)^2 + (84 + 72 - 150)^2 + (92.4 + 72 - 160)^2$$

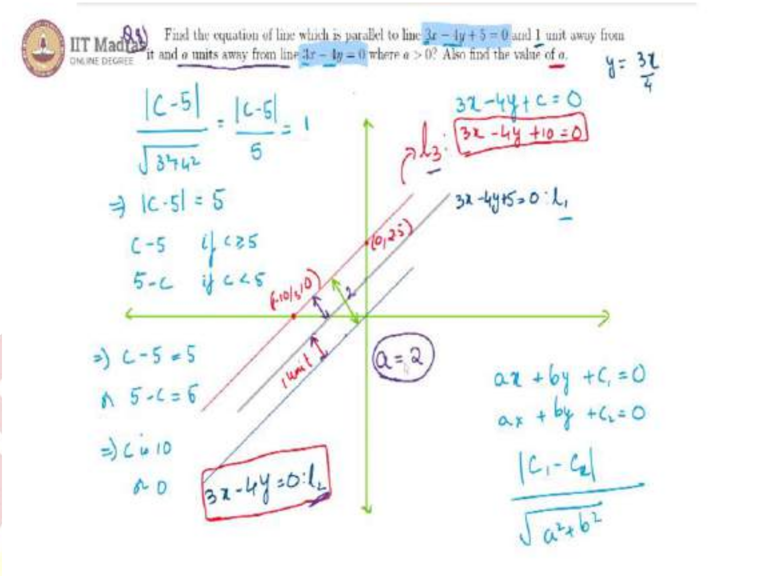
So, this is 6. So, sum square error for Chetan is 6. Whereas in Raju's case we would have $4.8^2 + 6^2 + 5.2^2 + 4.4^2$. Now clearly, in this error, there is a 6^2 which has to be greater than 6, which means Raju's error is much more than Chetan's error. Therefore, Chetan's line fit is better.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 03 – Tutorial 08

(Refer Slide Time: 00:14)



In our eighth question, we want the equation of a line which is parallel to this given line. So, let us first plot the line that is given, lets plot this line. If you look at that line, it is $3x - 4y = -5$, which gives us $\frac{x}{\frac{-5}{3}} + \frac{y}{\frac{5}{4}} = 1$. So, the x intercept and y intercepts can be marked out as 1.6s. So, if we take this to be 1 and this to be 2, so this is - 1 roughly and this is - 2, roughly. Yeah, this might be our intercept, which is $(\frac{-5}{3}, 0)$ and our y intercept, again, if we take this to be 1 and this to be 2, 1.25 is somewhere likely here.

So, this is probably our y intercept, $(0,\frac{5}{4})$. As you can see, we are doing a thoroughly rough plotting, we do not always have to be very accurate with our plotting. This is only for an indication. So, this would be our line, $3x - 4y + 5 = 0$ and now we have another line given to us, which is $3x - 4y = 0$.

Clearly, these two lines are parallel to each other because they have the same slope and that slope would be, we write it as y is equal to, we will get 3x / 4, so the slope is $\frac{3}{4}$ and it is passing through the origin, because if I put x = 0 and y = 0, the line equation is satisfied, that is there is no constant term. So, this line is our $3x - 4y = 0$ and we are trying to find a line that is parallel to these 2 and it is at a distance of 1 unit from the $3x - 4y + 5 = 0$ line.

Let us name these lines as well. Let us call this l_1 and this is l_2. I am going to erase the intercepts to make it look a little clear and now, we can find our equation and for that, we will use the formula of separation between 2 parallel lines. So, that two parallel lines and we write them with the same coefficients.

So, $ax + by + c_1 = 0$ and the other one would be $ax + by + c_2 = 0$. This is how two parallel lines would look like. You can reduce them to have the same coefficients for x and y, like in this case so this is 3 and this is - 4, this is also 3 and this is also - 4. In this case, the separation between these two parallel lines would be $c_1 - c_2$ the modulus divided by $(\sqrt{a^2 + b^2})$. So, the equation we are looking for, the line we are looking for also is going to be some 3x - 4y + c = 0.

So, its separation from our l1 is going to be applying the formula modulus of $\frac{c-5}{\sqrt{3^2+4^2}}$. $3^2 + 4^2$ is 25. Therefore, you have modulus of $\frac{c-5}{5}$, which is what is expected to be 1 unit. That gives us modulus of c - 5 = 5.

Now the model is, indicates that there are two possible values here, one could be c - 5 if $c \geq 5$,. Because then c - 5 would be positive, and the other would be 5 - c if c < 5. So, what we get is two separate solutions, one is c - 5 = 5 or 5 - c = 5, in which case we get c is 10 or 0. So, we have 2 lines, one is $3x - 4y + 10 = 0$, the other is our $3x - 4y = 0$, this is our line.

So now, because of this, we can say that this length between, the separation between these two lines is now one unit and therefore, the other line, which is our 3x - 4y + 10 = 0 is going to come on the other side of l_1, which is going to look like this. So, this line is our l_3 and it will have intercepts equal to, this should be 0,$\frac{10}{4}$ which is 2.5 and this would be (- $\frac{10}{3}$,0), this is our other plan.

And now, we should also find out what the value of a is, because a would be the distance between the lines l_2 and l_3 that is what they are saying, its units away from our l_2 and we know that this is one unit and this is also one unit. So, this total length is going to be 2 units. So, a = 2.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

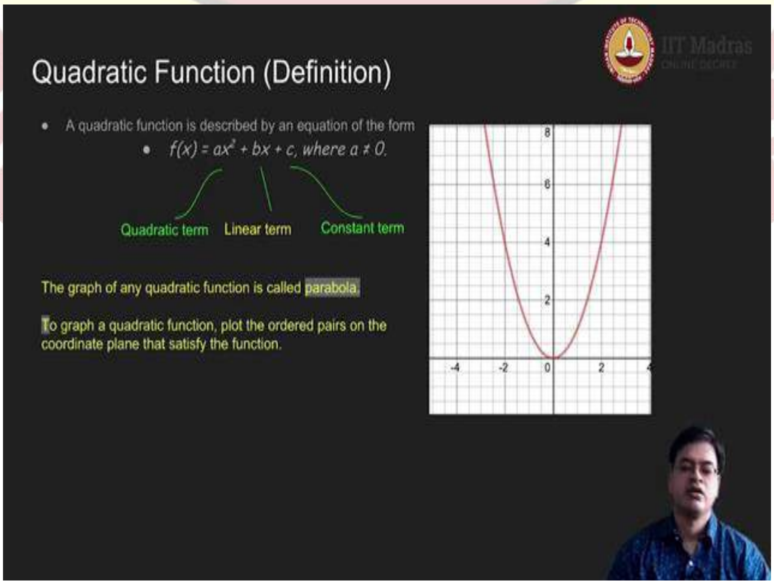
Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 4.1 A
Quadratic functions

Welcome students, today we are going to start the new topic in our syllabus that is Quadratic functions. Before starting this new topic, let us revise what we have studied so far. We started with some simple geometric objects like points and lines, after studying points and lines geometrically we plotted them on coordinate plane and seen how to derive the algebraic equation of a geometric line.

When we have seen the algebraic equation of a geometric line, we got a form of the form $y = mx + c$, it is also known as linear function that we have seen in last few lectures. And if you recall recollect it from the first week where you have studied functions, this is $f(x) = mx + c$is a linear function.

Now, we want to enhance our knowledge further and add 1 more intrication or 1 more complexity in this particular function and that is why we are studying quadratic function. Here we will take an approach where we will first state the algebraic form and then derive its geometric properties as opposed to what we did in straight lines. So, let us start with quadratic functions.

(Refer Slide Time: 01:38)



The first question is, how will I define quadratic functions? The answer to this question is given in this slide. So, a quadratic function is described by an equation of the form $f(x) = ax^2 + bx + c$, where $a \neq 0$ is a crucial condition. Why? If $a = 0$ it simply reduces to a linear function. Let us talk about the name quadratic function. The name quadratic function is derived from 1 foreign language where the quadra term, actual word quadratic term means square and quadratic means related to square.

So, a quadratic function is a function that is related to square of the variable as can be seen from the definition, it has a term containing a x^2. So, if $a = 0$ then it does not have a term containing square so it no longer remains a quadratic function and it is a linear function which is equivalent to a straight line as a geometric object.

So, we will put a condition that $a \neq 0$ that means we are studying a quadratic equation. The next question is how to plot a graph of this function. So, this equation is actually composed of 3 terms, let us describe them 1 by 1 that is a x^2, this term is a quadratic term.

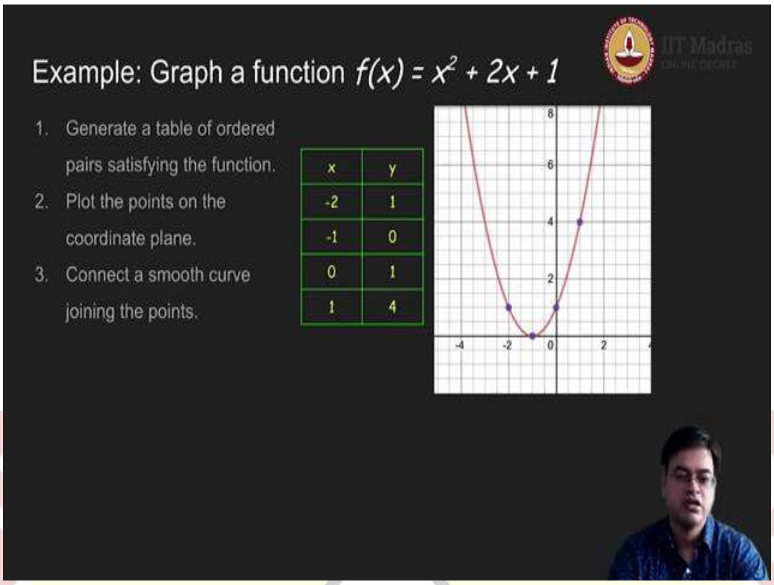
As I mentioned earlier, when $a = 0$, the term $bx + c$ survives and that term bx is a linear term. And finally, if you put $x = 0$, only term that survives is c so that is nothing but a constant term. So, a quadratic equation can be split into 3 parts. If ax^2 is not there, then I know how to handle this term on a coordinate plane, it just simply represents an equation of a non-vertical line.

So, I know how to handle these terms. So, what if the a x^2term remains that is $a \neq 0$? We can graph this particular function and graph of any quadratic function will be called as parabola. Graph of any quadratic function will be known as parabola. So, what are the important features of parabola? In order to do that we first need to plot the parabola.

So, what is the best way to do it? We have already seen to graph any function what we need to do is, we need to take the value of x, put it in the formula $f(x)$ and evaluate it and get the values of y. So, consider all ordered pairs and plot them on the coordinate plane so that they satisfy this function, is the best way to handle it. For example, let us take this let us take for example, when $b = 0$ and $c = 0$ and $a = 1$, let us take that particular function that is $y = x^2$.

In that case what I will do is, I will put $x = 0$, I will get back 0. So, I will plot a point (0,0), I will take $x = 1$ I will get 1, $x = -1$ I will get 1, and then $x = 2$ then I will get 4 and if I take $x = -2$ again I will get 4. So, $y = x^2$can be easily plotted by joining these points smoothly this is the curve $y = x^2$, this is how we plot our quadratic function.

(Refer Slide Time: 5:35)



Let us take 1 example. Let us say I want to graph a function $f(x) = x^2 + 2x + 1$. How will I graph this function? In 3 steps. First, I will generate a table of ordered pairs satisfying the given function. Second, I will plot those points on the coordinate plane. Once I plot those points on the coordinate plane, I will connect a smooth curve joining the 2 points, this is the recipe for drawing a function. Let us draw it here, for that I have computed some points you can verify by yourself, if you put the value of $x = -2$, you will get $(-2)^2 + 2(-2) + 1$ and on solving you will get 1.

You take the value $x = -1$ you will get 1, for x square you will get -2 and 1 in the constant term. So, together they will cancel and you will get 0. Similarly, you can compute for $x = 0$ it is 1 and for $x = 1$ it is 4. Now, our job is to consider a coordinate plane and plot these points so I have plotted these points.

So, these points are plotted and now I need to draw a graph, which is connecting all these points. Now, here you remember I have plotted these 3 points, how will I know the shape of this graph in this zone? That is a major question that you can ask, but this parabola is somewhat symmetric in a sense, suppose I take this point, what is the point here, the point is (1, 4).

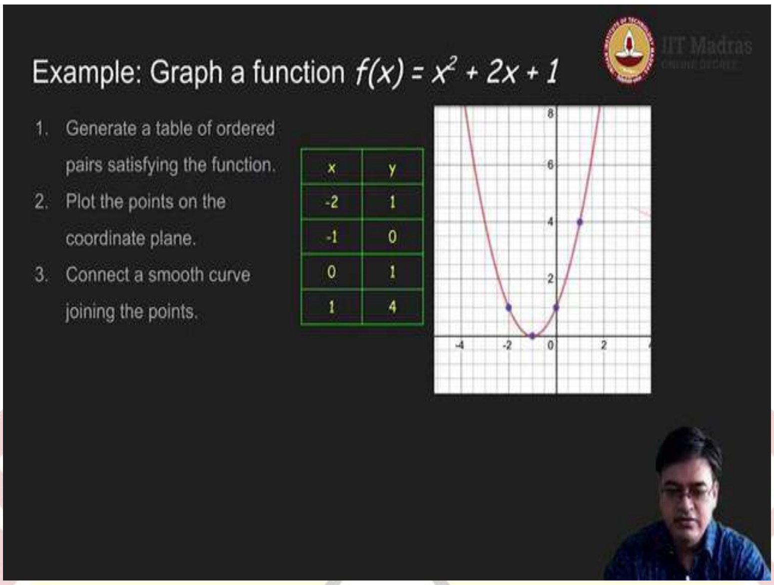
Now, if I consider this point which is -1 where it takes the value 0 and consider the point 1 it is 2 units apart. So, somewhere in this where -3 will come, which is 2 units apart from -1, the value of the parabola will be again 4. I will keep the cursor here, see. So, there is some kind of symmetry underlying this particular function, we need to understand that symmetry in a better way.

So, what essentially is happening is, if I consider this point which is the bottom of the curve, and if I draw a straight line, which is the line $x = -1$, then if you look at all these points for every point there is a similar point on $y - axis$ at the same distance from this particular point. This also can be called as a symmetry of a parabola. We will study this later in the next slide. So, right now, our job is to graph a function which we have plotted and let us explore further properties of this parabola like this symmetry, what is the meaning of the symmetry and all those things.

(Refer Slide Time: 09:00)

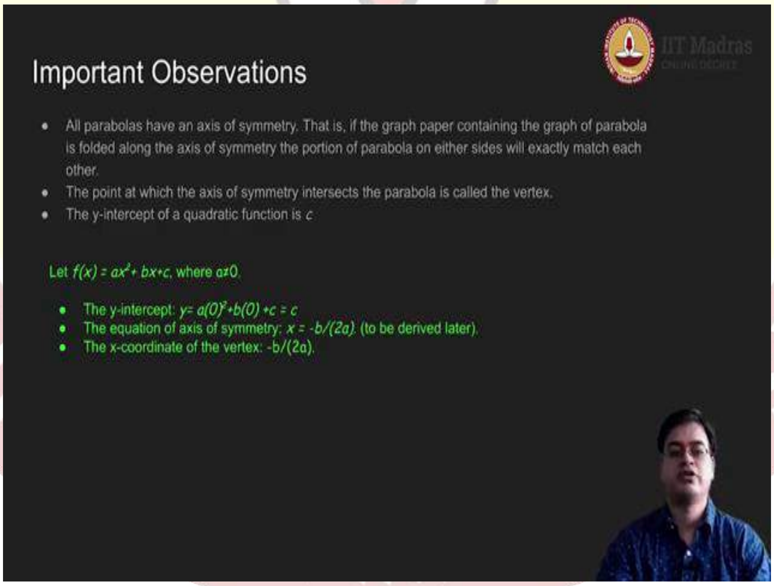
So, there are a few important observations, if you consider equation of $y = ax^2 + bx + c$, these are all parabolas, I have shown you two parabolas $y = x^2$, $y = x^2 + 2x + 1$both parabolas have axis of symmetry. Inevitably all parabolas will have an axis of symmetry that is, what is axis of symmetry. Let us go to the previous slide and see.

(Refer Slide Time: 9:28)



The axis of symmetry over here, as I mentioned was $x = -1$. If I take this graph paper and fold along $x = -1$, then the curves that we have plotted here must exactly match each other that gives us a recipe to draw a parabola.

(Refer Slide Time: 9:50)



So, all parabolas will have axis of symmetries that is if you take a graph paper containing the parabola, and if you fold it along the axis of symmetry, the portions of the parabola on both sides will exactly match with each other, this is the beauty of a parabola. So, now if I know how the parabola appears on one side, I know how the parabola appears on the other side of the axis of symmetry. It is a pure reflection of whatever is happening on one side.

Then, the point this axis of symmetry as we have seen in the previous graph, the point at which this axis of symmetry meets parabola, we will call that point as a vertex of the parabola. This is again a nomenclature we will call that point as a vertex of the parabola and the point at which x, if you put $x = 0$, then the point at which the y coordinate is taken is called the value c or you can simply refer to the equation $ax^2 + bx + c$, put $x = 0$, that will be the y intercept which will be given by c.

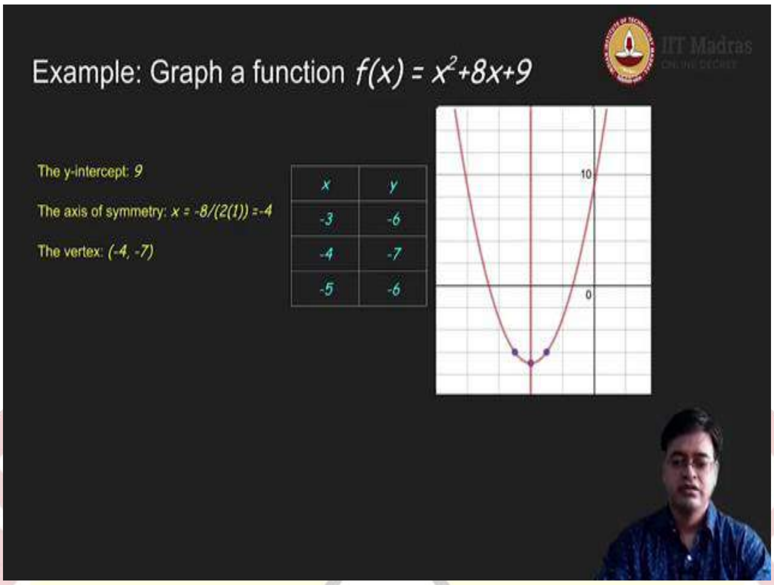
These 3 points play a crucial role in graphing the parabola. How? Let us do it one by one, Let us say, our quadratic function is $ax^2 + bx + c$ where $a \neq 0$, you can easily figure out that the y intercept of this point by putting the value 0 in c.

Now, I want to know the axis of symmetry, this plays a crucial role. So, I will derive the expression for axis of symmetry later but right now you memorize this equation as $x = \frac{-b}{2a}$ as this needs some algebraic skills which we do not have right now. So, I will derive it later. But right now, you understand that $x = \frac{-b}{2a}$.

Remember, the equation of the quadratic function is given by$ax^2 + bx + c$, and c will not play any role in this and b and a will play a role. So, it is $\frac{-b}{2a}$ is the axis of symmetry and where the graph meets this parabola, it is called vertex. So, the x coordinate of the vertex is $\frac{-b}{2a}$ obviously, because the axis of symmetry has $x = \frac{-b}{2a}$. So, the x coordinate of the vertex is $\frac{-b}{2a}$.

Let us see how this knowledge helps us in understanding how to draw a parabola. So, there are 3 steps in drawing the parabola, first you need to generate a table of values, but if you generate a table of values only on one side and you do not have a table of values on vertex, then you may not be able to draw the parabola appropriately, that is way the knowledge of these facts is important, so let us see how to draw a parabola by example.

(Refer Slide Time: 13:07)

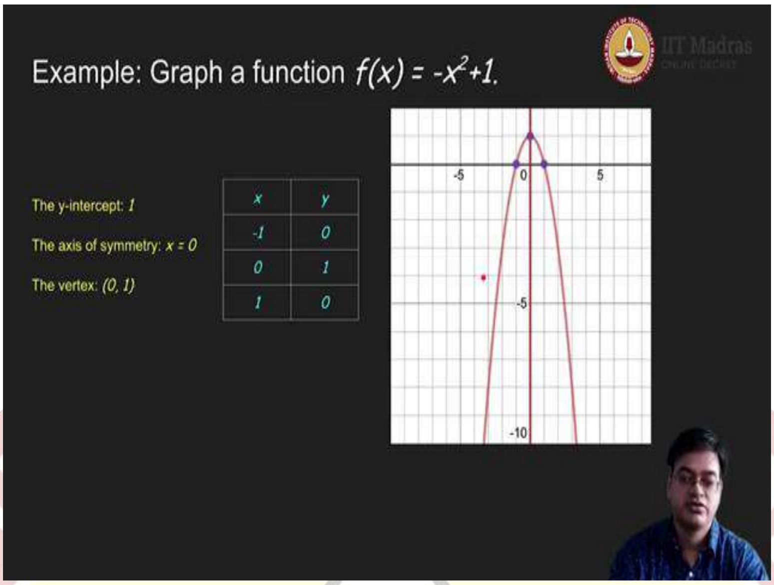


$f(x) = x^2 + 8x + 9$, I want to graph this function. So, I will reiterate on previous points. So, what is the y intercept, y intercept is 9, because if you put $x = 0$, the y intercept is 9. Next, I want to know the equation for axis of symmetry. So in this case, what is b, b is 8, a is 1, so $\frac{-b}{2a}$ is $\frac{-8}{2}$, which will give me axis of symmetry to be $x = -4$, y intercept is 9, axis of symmetry is $x = -4$ so I can evaluate the coordinate that is $(-4, -7)$ will be the vertex. How this -7 comes, you just substitute -4 over here in this expression, you will get the value to be equal to -7.

So, now with these 3 terms, how will I draw the function? So, now I know that around vertex I need to find the points. So, based on this, I will draw a table. So, around -4 I have simply taken three points, fourth point is already with me, (0, 9) is the 4th point. So, around the around the point -4 I have taken the values so -3 which is the value of -6 when you substitute in the function -4, -7 already known and -5, -6. So, I have 3 points and the point (0, 9).

So, I will plot these points on a graph paper, take a graph paper, plot the axis of symmetry because around this the curve should be symmetric, take these 3 points, these 3 points are here and I know (0, 9) is another point. So, it should be somewhere here (0, 9). Now, let us plot a graph. So, now we have plotted a graph with much ease because of the knowledge of axis of symmetry I know where the point where the minimum has occurred or the vertex point is that is the beauty of axis of symmetry. So, this is how you will be able to plot any function any quadratic function given to you, this is about the graphing of a function.

(Refer Slide Time: 15:48)



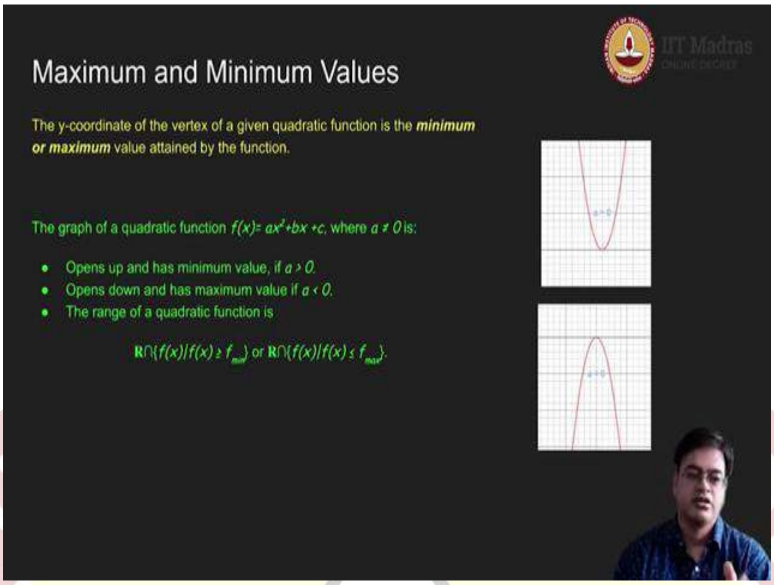
Let us try to see, is this the only shape that is possible that is the upward shape. Let us try to figure out whether is this a quadratic function first of all, $-x^2 + 1$, the answer is yes, $a = -1$, $b = 0$ and $c = 1$. Now, in this case, let us try to figure out the 3 summaries that is what will be the y intercept for this? y intercept will be 1, what will be the axis of symmetry for this because b is 0, it does not matter what is the value of a it will be 0.

So, $x = 0$ is the axis of symmetry that is y axis is the axis of symmetry for this particular function. And the vertex is (0, 1). In this case, we are not really getting much information because what this is saying is (0, 1) is the y intercept that is (0, 1) is the coordinate, axis of symmetry is 0 that means (0, 1) is the vertex as well, right?

But still this information will suffice because I know I have to find the points around 0. So, let Figure out the points around 0; -1, 0 and 1, these are the 3 points, their y coordinates respectively are 0, 1, 0. Now you see there is a change, earlier we were only dealing with positive side of y axis or the y axis where the curve is opening up, here the curve is opening down. For example, if I plot an axis of symmetry over here, which is y axis and if I plot these 3 points, these 3 points look like this that means the curve will go downward. So, the curve is opening down, why this has happened.

In earlier examples, if you look at it closely then this the form, general form of this expression $ax^2 + bx + c$, in all of them a was equal to 1, and in this particular expression $a = -1$ therefore, the curve is actually opening down instead of opening up, this point needs to be noted.

(Refer Slide Time: 18:15)



So, let us know this point and figure out what happens when this a is greater than 0 or is less than 0. So, that leads us to the next question that is maximum and minimum values. So, the y coordinate of the vertex of a given quadratic function is minimum or maximum value attained by the function.

Do you all agree with this, we have seen 3 to 4 graphs of the functions, first we have seen $y = x^2$ where it goes to bottom and 0 is the minimum value, then we have seen $y = x^2 + 2x + 1$which again gave us 0 value, then finally we have seen Third graph that we have seen the last graph that is$-x^2 + 1$, because the value of a was negative, it was going downward, and that will give me the maximum value and all the values are below that value.

So, the y coordinate of the vertex of a given quadratic function gives us the minimum or maximum value attained by the quadratic function. In particular, given any graph given any function $f(x) = ax^2 + bx + c$ where $a \neq 0$. The graph of this quadratic function if a is greater than 0 will open upwards and will have a minimum value.

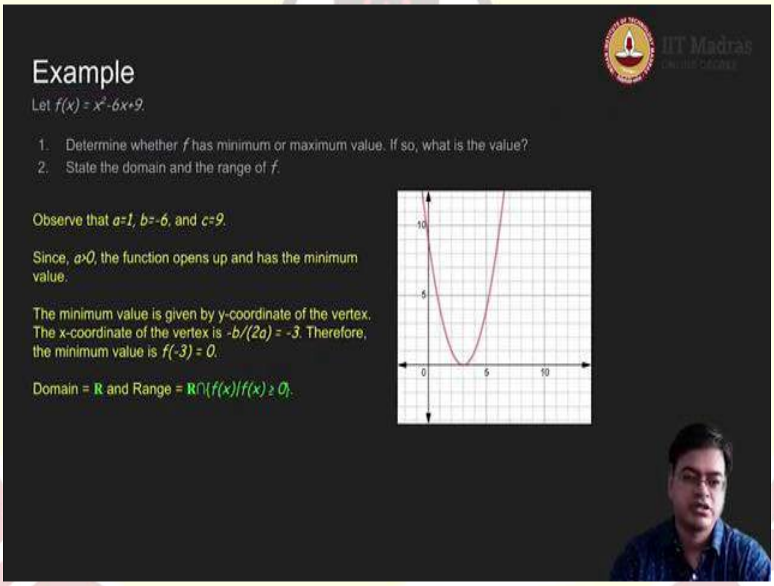
If a is less than 0, the graph will open downwards and will have the maximum value and there will be either maximum or minimum values, not both, this is the beauty of the quadratic function. Another thing that you can see is the range of the quadratic function, if you relate to your weak 1 background, where you are discussing about the domain-codomain range, so the range of this quadratic function will be.

So, let us say a is greater than 0, then it attains the minimum value then it will be the minimum value and all of the real line that is above the minimum value. And if a is less than

0, then it will be the maximum value and an entire real line which is below that particular thing, I can denote this using the set theoretic notation as it is a set of real numbers R intersected with a set of all $f(x)$, these y values such that $f(x)$ is greater than or equal to f min when $a > 0$, or if $a < 0$ it is set of real numbers intersected with $f(x)$ such that $f(x)$ is less than or equal to the maximum value that f has achieved.

So, let us try to visualize this. For example, if $a > 0$ your graph looks like this. So, in particular the range of the value, range of y values is from this point to upward. So, this is the entire real line above this value. Similarly, if $a < 0$, the range of the values that is taken by this function is this, if you relate this to domain codomain terminology, what is the domain of this quadratic function, it is an entire real name and range is restricted to some subset of real life.

(Refer Slide Time: 21:51)



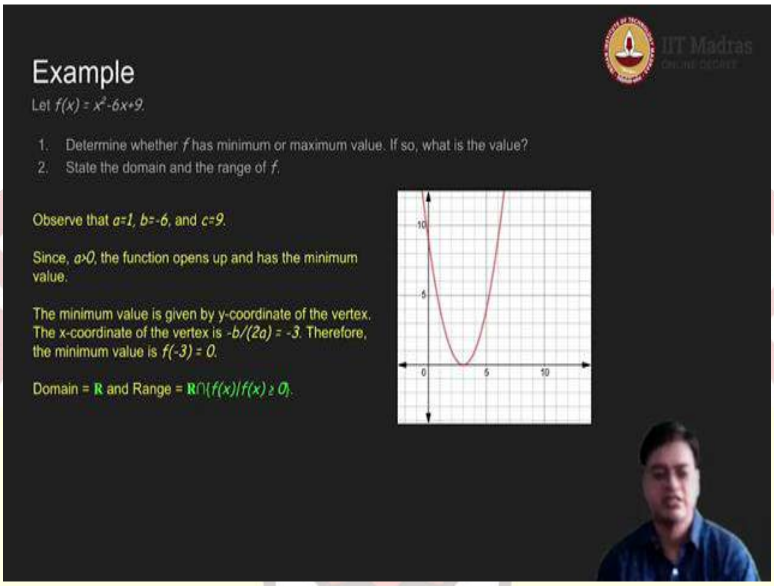
So, we will try to improve upon this concept using this example.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 4.1 B
Examples of Quadratic Functions

(Refer Slide Time: 00:16)



Let us say, this example we have been given a function $f(x) = x^2 - 6x + 9$ and we are asked to determine whether f has minimum or maximum values, if so what is the value and you need to state the domain and range of f. Let us first attempt the second question, what is a domain? Domain of f is enter real line we do not have to worry, what is the range of f?

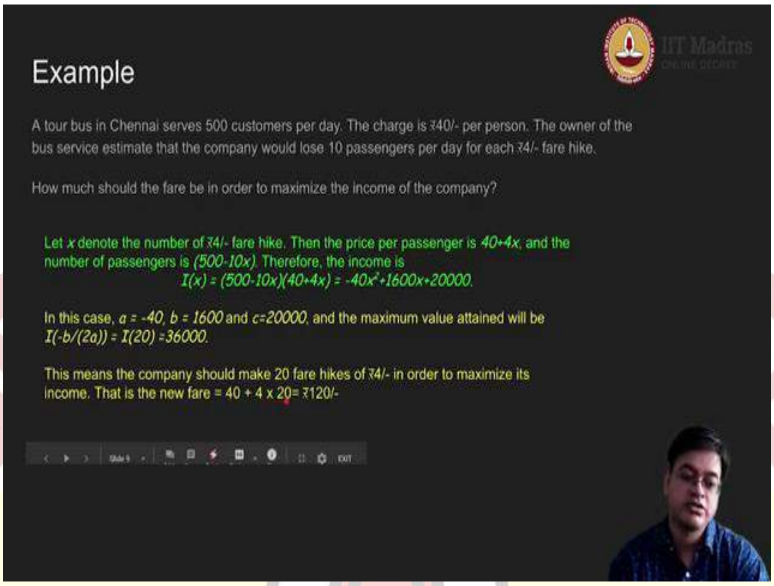
Let us take this function identify a, b, c so $a = 1, b = -6, c = 9$. Since $a > 0$, the function opens up, if the function opens up then it will have a minimum value. So, the answer to first question is whether f has minimum or maximum value, it has a minimum value, once it has a minimum value it cannot have maximum value, if so what is the value?

You need to figure out what is the vertex of this particular parabola. So, what is the formula for vertex of the parabola, $\frac{-b}{2a}$, $b = -6$, $a = 1$ so $\frac{-b}{2a}$ is $\frac{-(-6)}{2}$ which will give me minus 3. Sorry, this is wrong, it should give me $+3$, $\frac{-(-6)}{2}$, it should give me $+3$, which is written wrong here, but the graph is correct here where we are getting $x = 3$ is the vertex. So, if you substitute $f(3)$, what do you get? $f(3) = 3^2 - 6(3) + 9$ and therefore, the value of this is nothing but 0. So, this -3 is wrong it, should be $+3$.

And obviously, the range if it has a minimum value, the range is minimum upwards, this is the entire real line above this minimum. So, that is R intersected with $f(x)$ such that $f(x) \geq$

0. So, we have understood how to find the minimum and maximum values of a function, if a is negative you can similarly find the maximum value.

(Refer Slide Time: 02:58)



So, let us try to make this example more realistic. So, let us take 1 realistic example. Where a tour bus in Chennai serves 500 customers per day, they charge rupees 40 per person. Now, they want to revamp their strategies, so the owner of the bus service estimate that the company would lose 10 passengers per day for each Rs 4 hike in the fare.

So, if they hike the fare by 4 rupees, then they will lose 10 customers per day, this is the estimate. Now, the company wants to maximize the profit, so how much should be the fair in order to maximize the income of the company is the question. So, let us try to answer this question using our knowledge of quadratic equation.

So, let us say, 1 unit of hike is 4 rupees so let x denote the number of Rs 4 fare hikes. So, what will this impact? This will impact the number of passengers because we are losing 10 passengers per fare hike. So, what will be the corresponding fair price for the passenger? It will be $40 + 4x$, 40 rupees is the fees that we are charging per person, the company charging per person and if I hike the fare it will be four times x, this will be charged per person.

Now, the number of passengers with this hike if you increase x units, that means, you will lose 10 passengers every x units increase. That means $500 - 10x$ is the passengers that still remain. So, in this case, the income of the company will be the number of passengers into the fare, they have charged so that is $(500 - 10x)(40 + 4x)$. If you open this, open the bracket and multiply them, then you will get the expression to be$-40x^2 + 1600x + 20000$. This is the income.

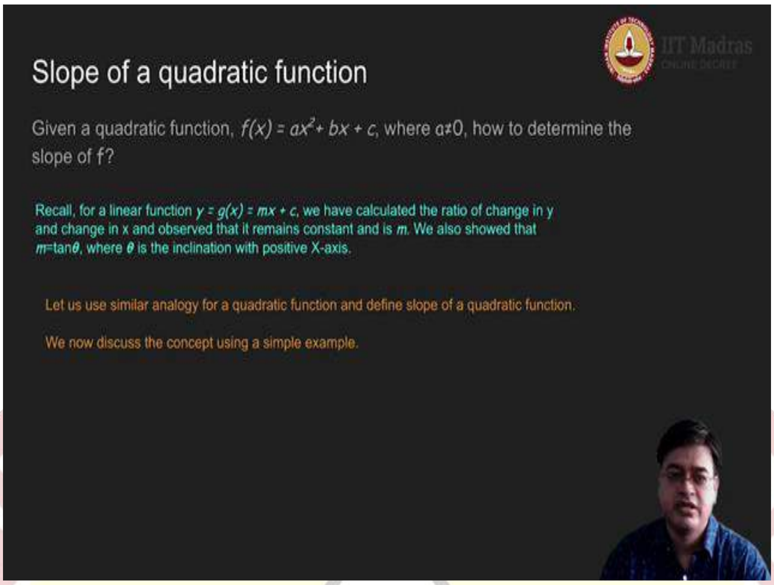
Now, the company wants to maximize the profit, first of all after getting this quadratic equation, can you tell me is the maximum possible? The answer is yes, and why the answer is yes, because it lies in the coefficients $a, b$and c. So, what is $a$here? $a = -40$, $b = 1600$ and $c = 20{,}000$. Because $a = -40$, a < 0 so, the curve will open downwards that means the maximum is possible.

And what will be the maximum value attained then that is what we have to figure out. So now, the next question is okay. So, where this maximum will be attained? The maximum is possible, maximum will be attained on the vertex, y coordinate of the vertex will give me the maximum. So, I will simply figure out what is the x coordinate of the vertex, x coordinate of the vertex is point of intersection of the axis of symmetry, what is the axis of symmetry $x = \frac{-b}{2a}$, what is b? 1600, c is 20,000 and a is -40.

So, $x = \frac{-1600}{2(-40)}$ which will give me 20, so that is what 20 is yes. And of maximize y coordinate the maximum fair that we will get is 36,000. Right now, how much we are earning, how much the company is earning, it is 500 customers they are serving, where everybody is paying 40 rupees so they are simply earning only 20,000 rupees that is when you do not increase any fare $x = 0$, you get 20,000. So, the main question is how much the fare should be?

Now, what we are suggesting here by solving this problem is there should be a 20 units of hike of rupees 4 each that means, what we are suggesting is there should be 80 rupees hike in the fare. So, the new fare for the company should be 40 plus four times x that x is 20. So, $40 + 4 \times 20 = 120$and that is what is the recommended hike by the company. So, now every person should be charged 120 rupees as opposed to 40 rupees and then the company will be profitable and you may have to serve less customers. This is how we are using real life, we are using quadratic equations to solve real life situations.

(Refer Slide Time: 8:37)

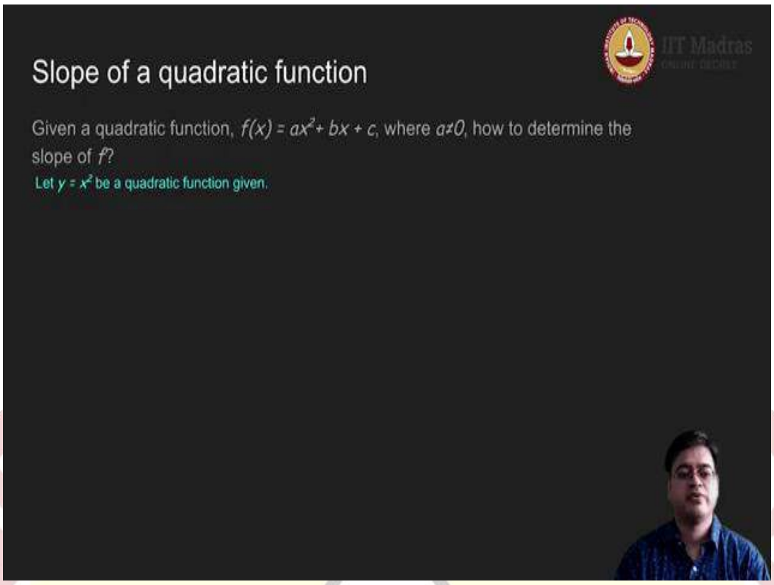


Now, let us go back to our linear functions, where we studied the slopes of the lines. What was the slope of a line? Slope of a line was change in y by change in x. Let us see what the concept of slope has to do with a quadratic function. Let us try to analyse that. So, my goal in this set of slides is given a quadratic function $f(x) = ax^2 + bx + c$ where $a \neq 0$, how to determine the slope of a function f.

So, in order to generalize this notion of slope of a function, we will first recall what we do know about linear function. So, if you look at a linear function which is y which is equal to $g(x) = mx + c$, we know that this m represents the slope and m can be calculated by considering a ratio of change in x upon change in y.

We have spent a lot of time in understanding the slope and when I consider a linear function, I also know that the slope remains constant okay. We also know that the slope is nothing but tan of some inclination and that inclination is with positive x axis. I want to relate all these concepts and try to figure out what is the slope of this quadratic function. Let us go ahead, we will use a similar analogy for a quadratic function and define the slope of a quadratic function. First let us take one example to discuss this concept of slope.

(Refer Slide Time 10:25)



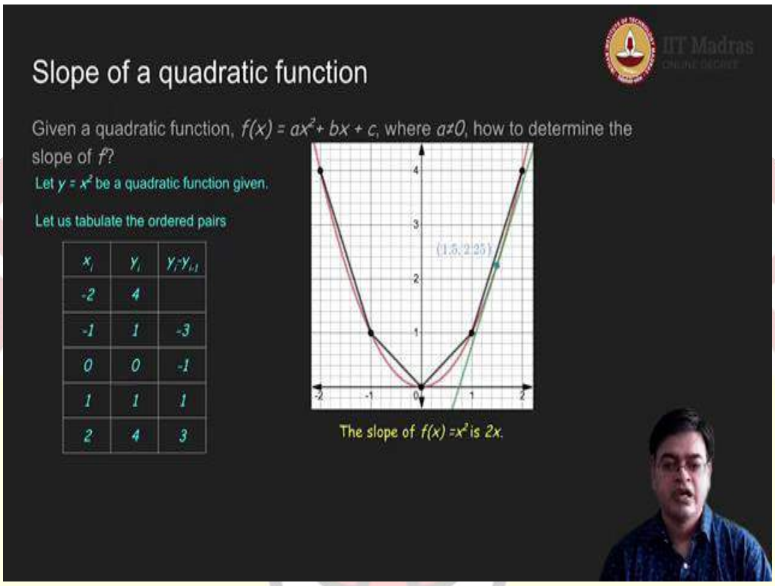
Let us take our standard prototype example. We are trying to answer this question, $y = x^2$.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 4.1 C
Slope of a Quadratic Function

(Refer Slide Time: 00:15)



So, in this situation, I want to know what is the slope of this curve and is it a constant or which variable or what else? So, we want to answer this question. So, first, we need to plot this function, for plotting the function, we know what is the axis of symmetry for this b is 0. So, so $y - axis$ is the axis of symmetry and it will be symmetric about $y - axis$. Minimum value will be 0 as it can be seen.

So, I will take a symmetry about $y = 0$ that is, I have taken -2, -1, 0, 1, 2, these are the points then I have evaluated the ordered pairs, that is 4, 1, 0, 1, 4. The symmetry is clearly visible in these. Now, what is the definition of slope? It is change in y upon change in x.

So, if you look at the left-hand side, the first column, the change in x is constant. It is 1 all the time so I will use this notation, and I will go ahead and figure out what is the difference between y_i values because the denominator is always 1, it suffices to take the difference between y_i values.

So, the first value is -3, 1 -4 is -3, $y_i - y_{i-1}$. $0 - 1 = -1$, $1 - 0 = 1$,$4 - 1 = 3$, so I got the changes in y with respect to 1 unit change in x so this is the slope, but where does this slope lie or at what point is this slope? Because if it is a straight line, I know the slope is constant. So, in order to understand this let us go to a figure and try to understand.

This is a curve, $y = x^2$. Now, when I consider these 2 points $y_i - y_{i-1}$, what I am actually doing is, I am assuming a straight line connecting these 2 points and I am calculating slope for it. So, I have assumed all these straight lines and I have calculated the slope for it. Is this a slope for a curve? No, basically not because it is a slope for that straight line.

So, now how will I identify this slope? So, if at all I want to decide what is the slope of the line if you look at our old definition the change in y by change in x also associated with $tan\ \theta$, the $tan\ \theta$ plays a crucial role, what is θ? θ is the angle of inclination. So, if I consider any point over here, and if I draw the inclination of, if I draw a line passing through that point and if I measure the inclination of that point with this positive $x - axis$ then I will get a slope because the definition of $tan\ \theta$ was not dependent on the line per se, it was dependent on that line on that particular inclination.

So, tan of that is still a slope of a line. So, let us try to use this idea and see what we can get. So now, I have identified 1 point let us say this point is actually (1.5, 2.25) because I am considering a curve, which is $y = x^2$. What will be the slope of a line at this point? We can ask this question but if you look at this line, this vertical line, this vertical line and if I slide this vertical line slightly for this point, then this is nothing but a tangent to this curve, it passes through it only once.

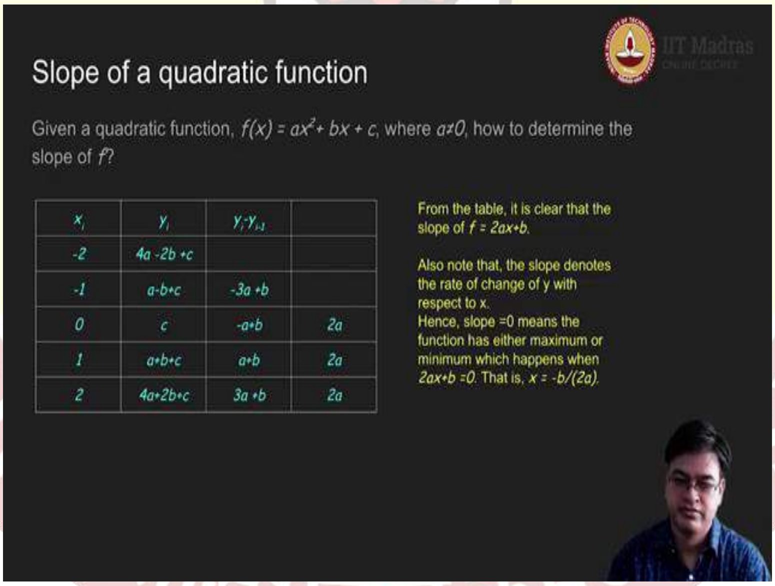
Let us try to actually plot that line. Yes so, once we have plotted this line, this is the tangent to that curve and the line is actually parallel to this line and the point is 1.5. This gives me a hint that this is something like you have -3, the point is, you have a slope between these 2 points as 3, you have a point which is 1.5 and if I divide this point, this particular difference by that point I am getting 2. Then let us look at these differences, what are these differences, the difference is $-1 - (-3) = 2, 1 - (-1) = 2$ so, all these differences are 2.

If you look at the second differences of these points, there are 2 that means there is some relation, 3 and 1.5, 1.5 times 2 is 3. So, I can safely assume that this point 1.5 is actually a midpoint of 1 and 2 on the $x - axis$ and therefore, whatever value is given to it is actually the value of the slope of a curve. And in particular, if I go here, for example, if I go here, and if I talk about the point 0 and 1, then what I will get is a point 0.5, the midpoint of this. Again, I can do a similar exercise, I can draw a line and the line again will be parallel to this line and at a point 0.5, I will get the line with a slope 1.

In a similar manner if I go here, I will get a line with a slope -1, in a similar manner here, I will get a line with a slope -3 and therefore, I can safely conclude that the slope of this particular curve is $2x$. How? I have computed it. So, let us now verify our hypothesis. So, let us take a point 0, consider any 2 points about 0, let us take symmetric points because I need a symmetry.

So, let us take the point (1, -1), what is the slope of this line? It is horizontal line, so the slope should be 0 and that is what this slope is. So, in particular, I can verify for all points if I consider a point, let us say a, a is used here. Let us say if I consider a point z then I will go z +u, z$-$u, I will consider those 2 things and I will assume their values, draw a straight line joining them and whatever is the value of the slope for that straight line will be the value of slope for my point. This is a beautiful idea that can be generalized to a general quadratic curve.

(Refer Slide Time: 8:33)



So, let us answer the general question that is, I want to find a slope of a quadratic function$ax^2 + ax^2 + bx + c$ where $a \neq 0$. So, we will simply take 5 set of points, standard 5 set of points, -2, -1, 0, 1, 2, I will just substitute these values in the function. So, I will get a corresponding values of y_i's, which are here, $4a - 2b + c$, and$a + b + c$, c, $4a - 2b + c$. I will take the first differences of these two, those are given here and then I will take one more difference of these two, all these differences will turn out to be $2a$.

Now, if I look at the points which are here, and if I consider the midpoint of this midpoint of these 2 that is 1.5 so$-2a \times 1.5 + b$ will give me the answer to my question that what is the slope of that particular value, because if you look at this -3, -3 is actually 2 times 1.5. This -1 is actually 2 times -0.5, 1 is 2 times 0.5, 3 is again 2 times 1.5 so I am essentially getting the slope of all these values that means, my answer to the question that the slope of this curve quadratic function is $2ax + b$.

Now, from the table it is very clear the 2 way comes here,$ax + b$ I have derived it because this is a value containing 1.5 in the middle, so this is 2 times 1.5 So, that is ax so $2ax$ that is what this is $2ax + b$. Now, we can do some interesting observations, we have already seen around point 0 for $y = x^2$, the slope was flat it was 0. So, when will that happen? Right.

So, you can equate this $2ax + b = 0$, slope 0 means the function has reached its minimum or maximum, slope is 0. So, when will that happen? That is $x = \frac{-b}{2a}$. This is one of the reasons why $x = \frac{-b}{2a}$ is the value of the minimum or maximum, because the slope reaches the value 0. So, here what actually slope, calculates?

Slope actually calculate the rate of change with respect to $x$and a rate of change of $y$with respect to change in x. So, if the rate of change is becoming 0, that means the function has reached its minimum or maximum. So, this justifies the idea that why a quadratic function should have a minimum or maximum value at the point$x = \frac{-b}{2a}$.

Still, that point is pending where we want to find why the axis of symmetry is $x = \frac{-b}{2a}$ and we will come to it later. But as you can see here, the slope of a quadratic function is significantly different from slope of a line, slope of a line is constant, whereas the slope of a function quadratic function f is no longer a constant. In fact, it is variable that is $2ax + b$. It depends on $a$and b, not on the constant c, which is expected.

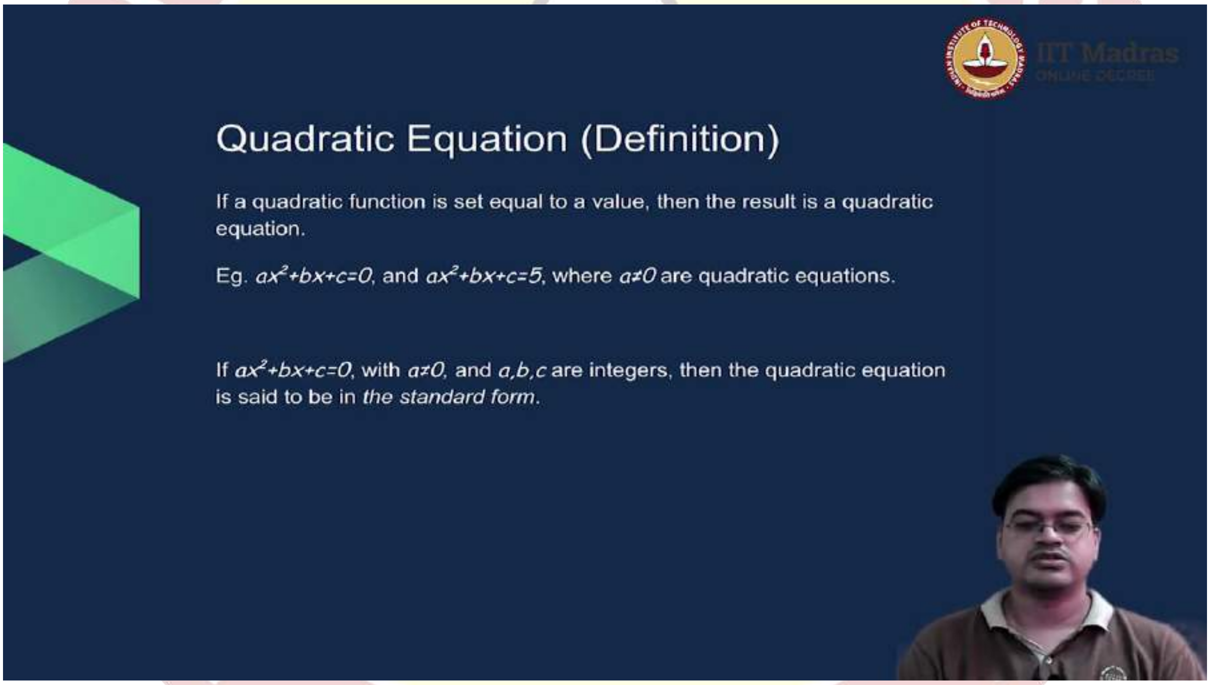
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 4.4
Solution of quadratic equation using graph

In today's video, we are going to learn what are Quadratic Equations. And once we set up the Quadratic Equation, we are going to see, what are the solutions of the Quadratic Equations, that are called roots of the Quadratic Equation and how to solve these Quadratic Equations, using the technique that we have demonstrated, for quadratic functions. That is Graphing technique. So, let us start.

(Refer Slide Time: 00:42)



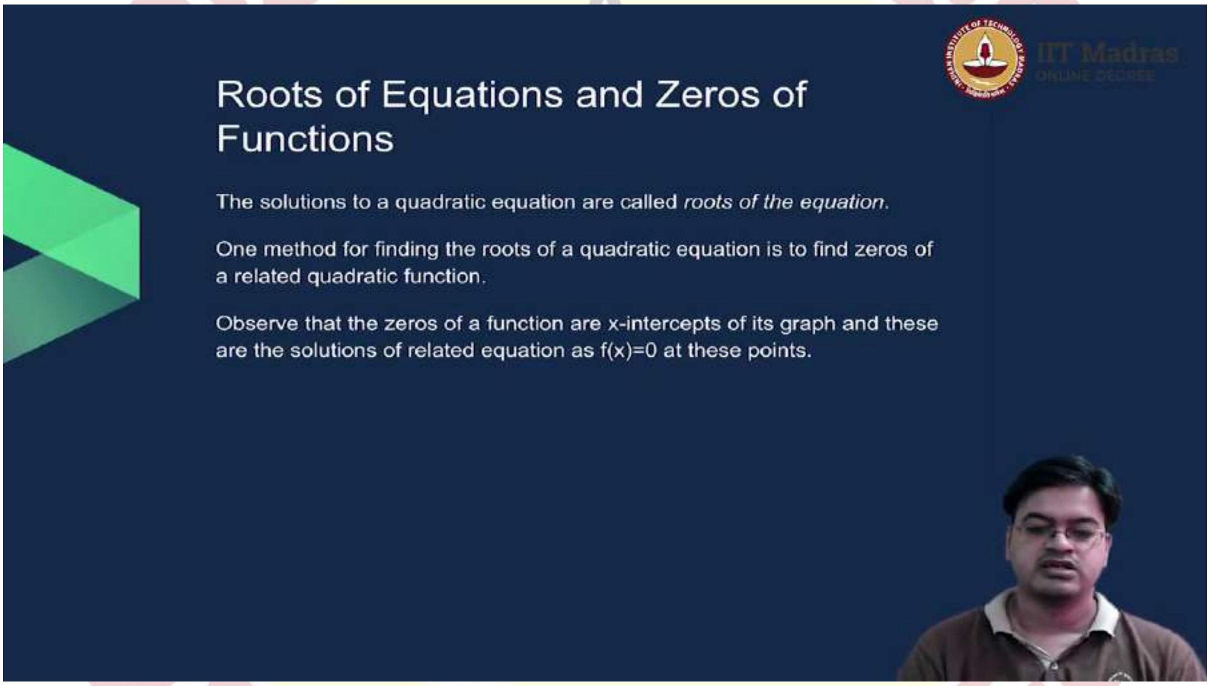
So, first of all, let us understand what is a quadratic equation and how it is related to quadratic function. So, here is a definition. If a quadratic function is set to be equal to a value, then the result is called quadratic equation. So, let us see one example. For example, $ax^2 + bx + c = 0$, is one quadratic equation, where $a \neq 0$.

In the similar manner, $ax^2 + bx + c = 5$, is another quadratic equation. Obviously a should not be equal to 0. Now, once we get the Quadratic Equation, if the coefficients, what are the coefficients. Coefficients are like a, b and c. These are called coefficients of the Quadratic Equation.

If the coefficients are from set of integers, which we have studied in week 1. So, if a, b, c, the coefficients are integers, and on the righthand side, it is equated to 0. That is, you have an equation, $ax^2 + bx + c = 0$, where a is not equal to 0 and a, b, c are integers. Then the Quadratic Equation is said to be in the standard form. So, on this slide, we have seen two definitions. One, what is Quadratic Equation. Quadratic Equation is nothing, but a quadratic function, where it is equated to some value.

And what is a standard form of Quadratic Equation? That is $ax^2 + bx + c = 0$, where $a, b, c \in \mathbb{Z}$ and a is not equal to 0. Then the Quadratic Equation is said to be in a standard form.

(Refer Slide Time: 02:37)



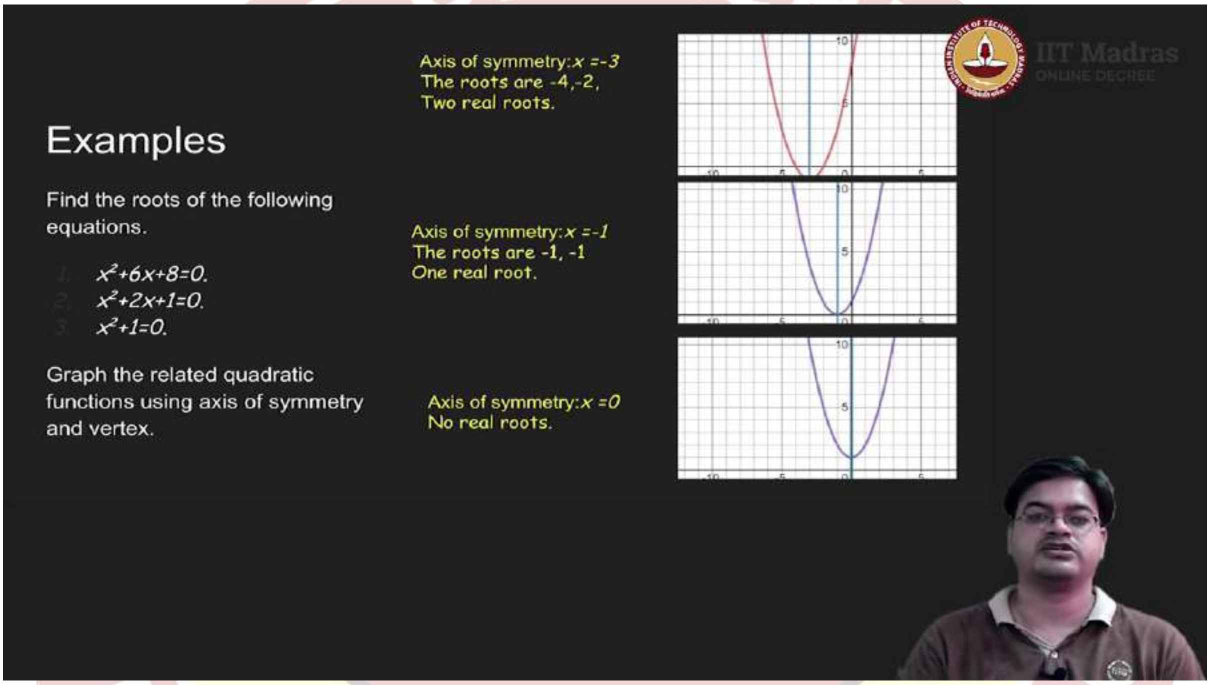
Now, once we have a Quadratic Equation in standard form, we can discuss about roots of the Quadratic Equation or zeroes of the functions. And we will see how the concept of roots of Quadratic Equation and zeroes of quadratic function are related, in this slide. So, the solutions of the Quadratic Equation are called roots of the equation.

What do I mean by that? If $ax^2 + bx + c = 0$, then what is the value of x, that gives me 0, is called the solution to the Quadratic Equation. And also, that value of x will also be known as root of the Quadratic Equation. So, this way we get the root of the Quadratic Equations.

So now, which way you can find the roots of the Quadratic Equations? One method, which is very easy. If you have a quadratic function associated with this Quadratic Equation, then you just plot the quadratic function and find its zeroes. What is a zero of a quadratic function? Zero of a quadratic function is nothing but its x intercept. So, in particular, if you observe that, zeroes of the functions are x intercepts of its graph and these are the solutions to the related equation, $f(x) = 0$, at these points?

So, if you are having a quadratic function, what you need to do is just plot it and see where it intersects x axis. If it intersects x axis, then you got the solution or the root of the Quadratic Equation. So, let us try to see this through some examples.

(Refer Slide Time: 04:32)



So, here are some examples. So, the question is to find the roots of the following equations. First equation is $x^2 + 6\,x + 8 = 0$. Second one is $x^2 + 2\,x + 1 = 0$. And third one is $x^2 + 1 = 0$. Now, we will take these equations one by one. So, essentially what we are proposing is, we want to chart these equations or we want to plot these expressions on a graph paper.

So, if you recollect from our last few videos, in order to plot a quadratic function, we need to understand the axis of symmetry of the quadratic function. So, let us take the first example, where you have $x^2 + 6x + 8$. Now, I want to understand, what is the axis of symmetry of this particular function.

Let us see. So, in this case, for our standard notation, our related quadratic function is $x^2 + 6x + 8$. So, $a = 1, b = 6\ and\ c = 8$. So, y intercept is obviously 8. And axis of symmetry is $x = -\frac{b}{2a}$, which obviously means it is $-\frac{6}{2}$, which is -3. So, axis of symmetry is x = -3.

So, axis of symmetry is x = -3 and a, the value of a is positive. So, what are the things that we can conclude from our previous videos? That is, if a > 0, the curve opens up, the graph of the function opens up. It attends the minimum.

And the axis of symmetry in this particular example is x = -3. So, the simplest thing that we can do here is, put x is equal to minus 3 in this expression. And you will see that, the expression will take a negative value. That means the y value taken is negative. That means if the y value taken is negative, you can easily see, the curve opens up. That means it will intersect x axis in two points.

Now, we want to guess those two points. Without plotting, right now based on our visual interpretation of this curve, can we guess the two points? Okay. So, -3, the value is negative. That means, for -3 it is negative. Then let us check it for x is equal to -2. If you substitute x is equal to - 2, you will get $4 - 12 + 8 = 0$. So, one root I have got, which is -2. If -2 is one root, -3 is one, -3 is axis of symmetry. That means, at a distance one apart from this, there will be another root. That means - 4 will be the second root.

Wow. So, we were able to understand, that -4 and -2 will be the roots of this equation, without even drawing, just on the basis of what we have understood. So, what we have understood here is, -2 and - 4 will take the value 0 and for x is equal to -3, you will get one negative value. And based on that, you have prepared a table. And therefore, you can plot this graph easily. Right?

So, we will graph the related quadratic function, using axis of symmetry and vertex. We have already discussed this. So now, axis of symmetry x = - 3, the roots are - 4 and - 2. And therefore, the Quadratic Equation given here, $x^2 + 6x + 8$ has two real solutions, two real roots. How will the graph look like? It is very easy. We have already imagined the graph. Yes, so this is the graph, where - 4 is a point here and -2 is a point here. -4, - 2 are the roots. And here, it achieves the minimum, which is - 3.

So, you can easily plot this graph. Let us go to the second equation. Now, in this second equation, again we will consider the associated quadratic function. What is the associated quadratic function? $x^2 + 2x + 1$. What will be the axis of symmetry for this? $-\frac{b}{2a}$, that will be - 1. Because b is 2 and a is 1. So $-\frac{b}{2a} = -1$.

So, x = - 1, is the axis of symmetry for this particular quadratic function. Let us substitute the value of x = - 1, in this quadratic function. So, you will get $(-1)^2$, which is 1, 2×-1, which is – 2, + 1. So, you will get 0. Oh! so, -1 itself is a zero. Correct? But that is a point of the vertex, where it achieves the minimum. So, there using axis of symmetry, you can conclude that there cannot be any other point, other than - 1, where it will take the value 0. Because that's the point, where the vertex arises.

That means the axis of symmetry for the second equation is x is equal to -1. a is greater than 0. So, the curve opens up. And therefore, it achieves the minimum. And therefore, the roots are -1 and -1. What is the value at -1? It is 0. So, that is that itself is a root. And therefore, it has only one real root, which is repeated twice. So, in particular, the graph of a function will look like this.

Now, the next problem is very interesting.$x^2 + 1$, where if you compare this with a standard form of the equation, $ax^2 + bx + 1$, then you will get b to be equal to 0. That means this curve or this, the graph of this function will be symmetric about x = 0, that is y axis. And since a is greater than 0, the curve will open upwards. So, the curve is opened up.

Now, a > 0, it will achieve the minimum value. Where it will achieve the minimum value? At the vertex. So, what is the vertex of this particular function? Because x = 0, so, where it, you substitute x is equal to 0 here. So, that value is 1,$x^2 + 1 = 1$. So, (0, 1), so 1 is the minimum value of this function. Can this function be equal to 0 then? It cannot be. So, this will give us the answer, that axis of symmetry x is 0. There are no real roots for this particular function, because it never intersects x axis.

And the function will look like this. So, this in short summarizes, what are possible solutions in any scenario, $ax^2 + bx + c = 0$ is given to you. In particular, if $ax^2 + bx + c$, if you are able to find the vertex and the vertex takes the negative value and a is greater than 0, the curve opens up. So, it will have two roots which are real numbers. If the curve opens up, but the value at the vertex is 0, then it has only one root.

And if the curve opens up and it is above the X-axis, that is it takes a positive value on the vertex, y coordinate of the vertex is positive. Then it will never intersect x axis. In the similar manner, you it is for you to study, that when a is less than 0, what will happen. So, I can give you the rough interpretation. If a is less than 0 and it achieves the maximum on the vertex.

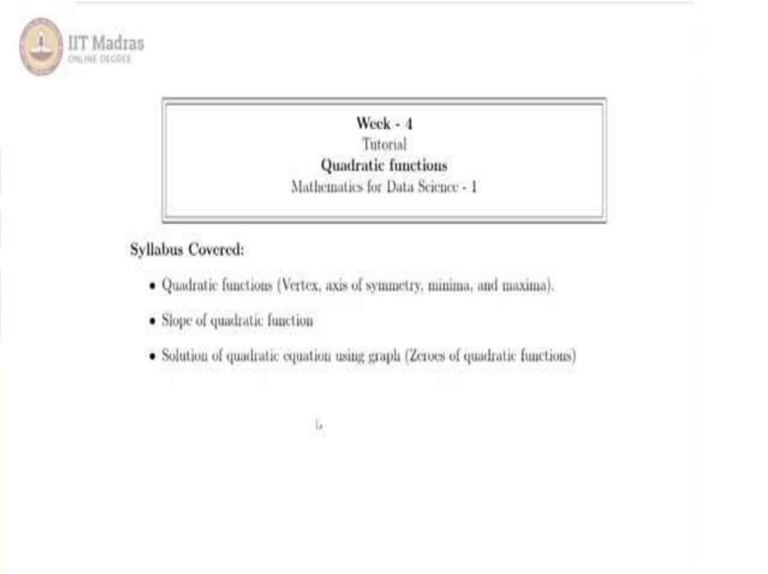
And if that maximum is positive, then it will have two real roots. If a is less than 0 and at the vertex, the value is 0, then it will have a single real root. And if a is less than 0 and it is below

X- axis, then it will have no real roots. So, these are the scenarios, that we can cover using this, this graphing technique.

IIT Madras
ONLINE DEGREE

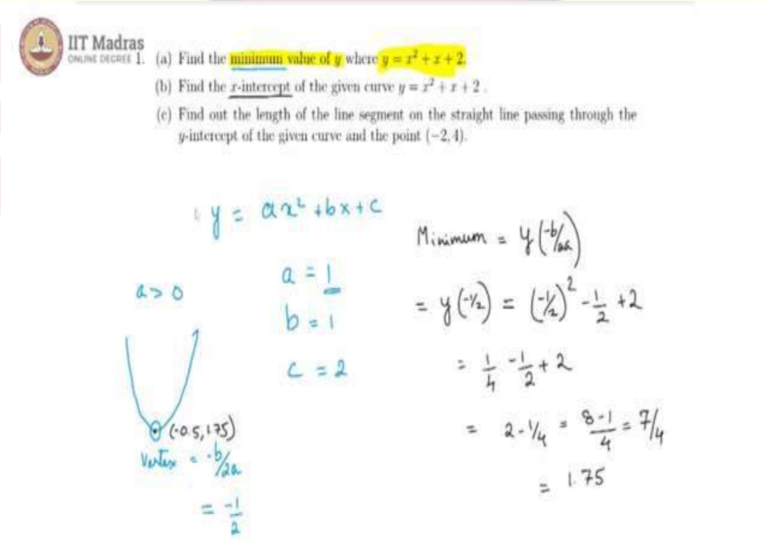
Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial - 01

(Refer Slide Time: 00:15)



Hello mathematics students in this tutorial, we are going to look at problems related to the topics covered in weak four. And so, these are the topics and we will begin with our first question.

(Refer Slide Time: 00:26)



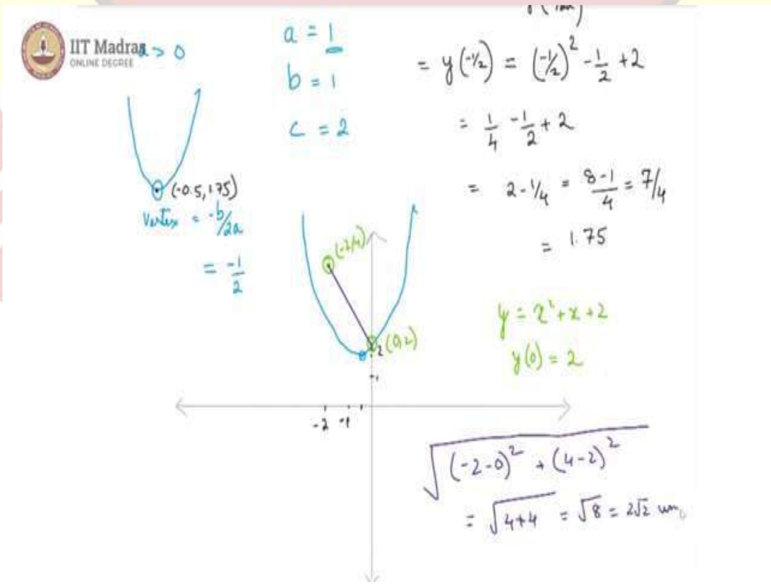
Here, we would like the minimum value of y for this particular quadratic function. And first, let us put down the quadratic function in its standard form, the standard form would be $y = ax^2 + bx + c$. In which case, our particular equation, the one that is given here would give us $a = 1, b = 1$ and $c = 2$. We are looking at the minimum value. Now, because the x square coefficient a is 1 that is a is greater than 0.

So, our parabola will be in this form, if a were lesser than 0, it would be inverted, it would be a downturned parabola, but right now it is in this form, and the minimum value is going to occur at this point, which is the vertex, which we know to be $\frac{-b}{2a}$. And so, we know our vertex for this particular equation is $\frac{-1}{2}$. And the value of y at $\frac{-1}{2}$ would be the minimum.

So, I can write the minimum is equal to $y\,(\frac{-b}{2a})$, which in this case is $y(\frac{-1}{2})$. And if I substitute that, I would get $(-\frac{1}{2})^2 - \frac{1}{2} + 2$, which is essentially $\frac{1}{4}$ - $\frac{1}{2} + 2$, which gives us 2 - $\frac{1}{4}$, which is equal to $\frac{(8-1)}{4}$, which is equal to $\frac{7}{4}$ and that is essentially 1.75. So, this point, here it is now we know it to be (-0.5, 1.75).

Now, they are asking us for the x- intercept and this is what we need to observe about the x −intercept.

(Refer Slide Time: 2:57)



Point (-0.5, 1.75) assuming this is 1 and this is 2, this is -1 of course, so this is negative side and this is -2. So, -0.5 is going to be somewhere here and on the Y- axis, this would be 1 and this would be 2, 1.75 is somewhere here so our vertex point is here.

And from here, we know that this is an upward parabola, which is going to be something like this. And that means it never touches the X- axis at all. There is no x −intercept for this parabola.

And lastly, it is asked to find the length of the line segment on the straight line passing through the y- intercept of the given curve and the point (-2, 4). So, (-2, 4) is somewhere over here, and we need to find this point here the y −intercept. And the y −intercept is easy to obtain, since our curve is y = x^2 + x^2+ 2. $y-$ intercept is obtained when the curve cuts the Y- axis that is when $x = 0$ so $y(0) = 2$, therefore, our intercept is actually (0, 2).

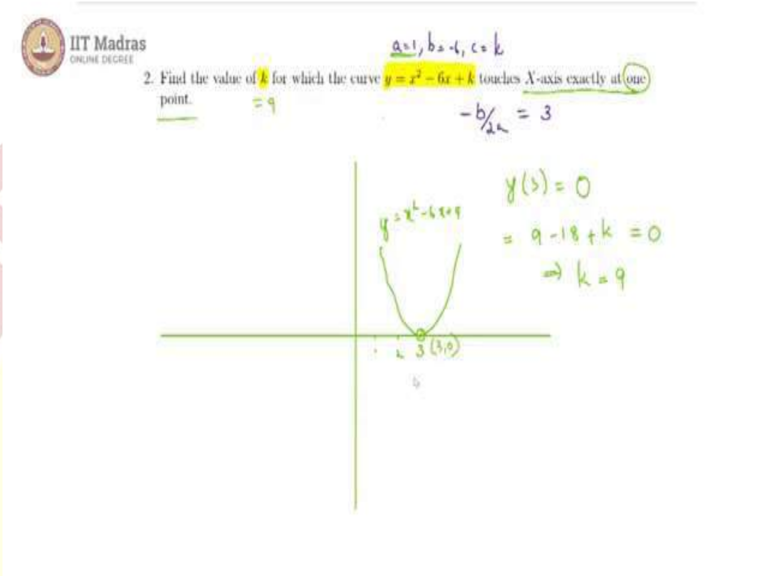
And the point we are looking at is (-2, 4) and it is the length of this line segment that we require. And that line segment we will get by using the Euclidean distance formula, it will be $\sqrt{(-2-0)^2 + (4-2)^2}$ which is essentially $\sqrt{4+4}$, that is root $\sqrt{8}$, which is $2\sqrt{2}$, units.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial - 02

(Refer Slide Time: 00:14)



Now, second question we are going to have, this quadratic functions curve touches the X-axis exactly at 1 point. And for that what is the value of k supposed to be? First observation should be that the vertex is given to us. The vertex, which is $\frac{-b}{2a}$, here $a = 1, b = -6$, and $c = k$, thus the vertex is $\frac{-b}{2a}$, which is $\frac{6}{2}$ that is 3. So, this is 1, this is 2 and this is 3, our vertex is on this particular line that is $x = 3$. And we are told that it touches the X- axis, the parabola touches the X-axis at precisely 1 point.

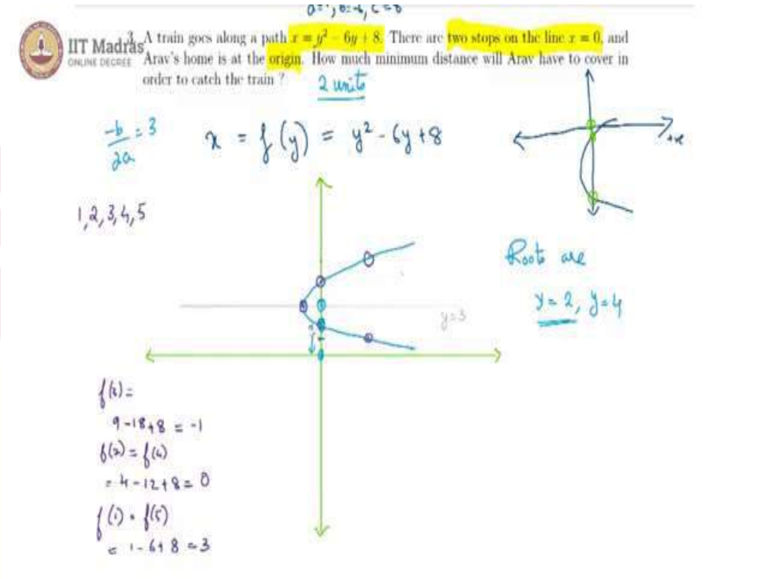
We also can see that a is positive, so this is an upward turn parabola, upturned parabola. And if it touches the X-axis at exactly 1 point that is only possible when the vertex is right here on the X-axis itself, and from here, our parabola looks something like this. That means, for this condition to be satisfied at the vertex, $y = 0$ that is $y\,(3) = 0$. And that is equal to $9 - 18 + k = 0$. This gives us $k = 9$ that is it so $k = 9$. When that happens, our equation is $y = x^2 - 6x + 9$ and it has its vertex at (3, 0).

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial - 03

(Refer Slide Time: 00:14)



In question 3, there is a path $x = y^2 - 6y + 8$. So, if we observe here, we are basically saying x is a function of y. And that function is quadratic, we have $y^2 - 6y + 8$. So, we are now switching the axis and so our parabola is expected to look something like this, or like this, because this is the X-axis and this is the Y-axis. Now, we see that the coefficient of y square, which is a is 1, and b is -6, the coefficient of y and lastly, the constant term c is 8.

Since, a is greater than 0, we expect that this is an upturned parabola. In the case of upturn in X, what we mean is it is towards the positive X-axis. So, our parabola is expected to be something like this. Of course, it could be moving about, we do not know where exactly it cuts the axis or where the point is. And for that, we will have to go further. They are saying this 2 stops on the line $x = 0$ that is on Y-axis and of course, these will be this point and this point, basically, if we looked at it in terms of our standard $y = f(x)$ these are what are the roots of our equation.

And Arav's home is at the origin, Arav lives at the origin so this is where Arav is. How much minimum distance will Arav have to cover in order to catch the train? So, the question is simple, you have two routes for your $x = f(y)$, and these routes will be on the Y-axis now, because we have switched the axis and which route is closer to Arav's home that is which route is closer to the origin. So, let us try to find out now, let us try to plot this particular graph and let us see where the tool train stops are. From the equation, we know that the vertex will be $\frac{-b}{2a}$, which is again $\frac{-(-6)}{2}$, that is 3.

So, here we are basically saying $y = 3$ is the vertex. So, this is 1, this is 2, then this is our $y = 3$ and thus, the vertex will be along the line, $y = 3$, the axis of symmetry is $y = 3$. So, this is our axis of symmetry, $y = 3$. And for plotting the graph, we are now going to look at various points, which will be 3 and 1 to the other side of 3, 2 and 1 to this side of 3, 4 and then 5, and then 1, this should give us a reasonable idea of what the graph looks like.

So, f (3) at the vertex, what is the x-coordinate that would be $f(3) = 9 - 18 + 8 = -1$, so $x = -1$, which is going to be somewhere around here, this is our vertex. And f (2) will be equal to $f(4)$ because of symmetry. So, if I just substitute 2, I will get $4 - 12 + 8 = 0$, ok, that is good so we now have roots, we know that on 2 this point, and at 4 our curve is going to intersect the Y-axis.

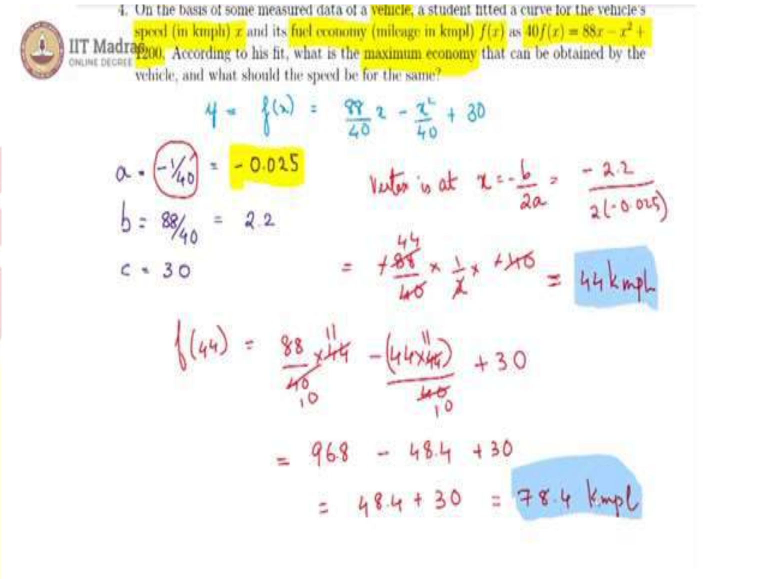
So, if you want, we can further look at what is $f(1)$ which is also equal to $f(5)$, that is going to give us $1 - 6 + 8 = 3$, so we got to be somewhere over here, for these two points we are going to get somewhere here and thus our quadratic parabola looks like this. And we know for a fact that the routes are $y = 2$ and $y = 4$. Clearly $y = 2$ is closer to the origin. So, the minimum distance that Arav will have to cover is 2 units that is from the origin to this particular point, and this is the distance he will have to cover.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial - 04

(Refer Slide Time: 00:15)



In the fourth question, there is some data of a vehicle, and a student fitted a curve for the vehicle's speed x. So, this is our variable x and its fuel economy mileage in kilometre per litre as $f(x)$. So, it is a function of x and this function is given in this way, we are going to use it for y which means if we reduce it to the standard form, we will get $y = f(x) = \frac{88}{40}x - \frac{x^2}{40} + 30$. So, we have the situation where a the coefficient of x square is $\frac{-1}{40}$, which is equal to -0.025. And b is the coefficient of x which is $\frac{88}{40}$ and that is $\frac{22}{10}$, therefore 2.2, and lastly, $c = 30$.

Now, we may observe that the x square coefficient is negative so this is a downturn parabola, which is why they are asking what is the maximum economy. So, at the vertex, you will get the maximum fuel economy so we need to find the vertex. And we know that the vertex is at x is equal to $\frac{-b}{2a}$, which in our case is then $\frac{-2.2}{2\times(-0.025)}$. This is probably better than in fractions.

So, if we write it down in fractions, we have $-b = \frac{88}{40}$ and this will be $\frac{1}{2}$ into $\frac{1}{2}$ and $\frac{1}{a}$ is then -40 itself, because a is $\frac{-1}{40}$. So, we have the 40 and the 40 cancelling off and minus and minus become plus 2, and 88 will give us 44. So, we have the vertex that is we get the maximum fuel economy at a speed of 44 kilometres per hour. And what is the maximum economy at this particular speed that we can calculate from our equation directly we have $f(44) = \frac{88\times44}{40} - \frac{44\times44}{40} + 30$ so this is 4, 10s a 4, 11s.

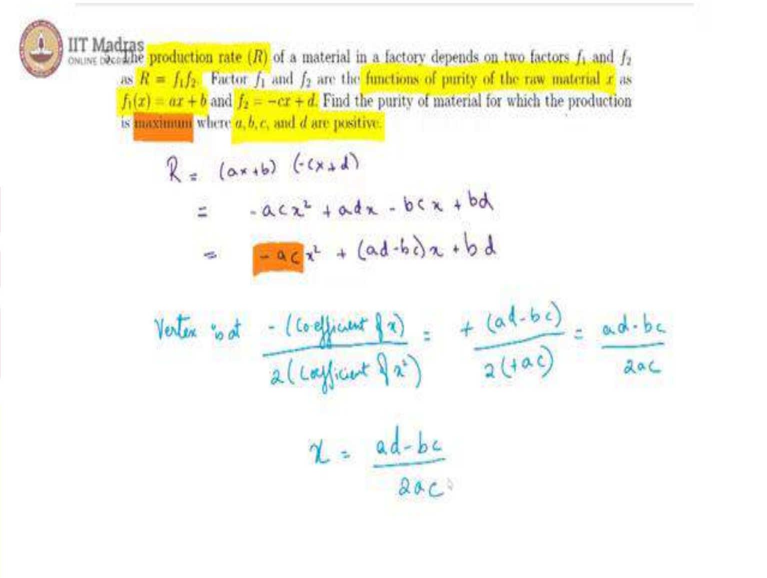
This is also 4, 10s and 4, 11s and we get $96.8 - 48.4 + 30$, which is then further equal to 96.8 is two times 48.4 so you will get $48.4 + 30$ giving us 78.4 kilometre per litre. So, we can say that this is our maximum fuel economy which is achieved at this particular speed.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial - 05

(Refer Slide Time: 0:15)



Our fifth problem looks a little complicated, but let us go one by one. And here we have the production rate of a material which is being made in a factory depends and it two factors f_1 and f_2 as $R = f_1 f_2$. And these two factors, they are the functions of the purity of the raw material. And that variable is x, x is the purity of the raw material. And both these functions are given to be linear $f_1(x) = ax + b$, $f_2(x) = -cx + d$. And it is given that a, b, c, d are all positive.

And it is asked find the purity of material, that is the value of x for which the production is maximum. So, let us understand what is being done here. We have two linear functions and the rate of production $R = f_1 f_2$, which will then $R = (ax + b)(-cx + d) = -acx^2 + adx - bcx + bd = -acx^2 + (ad - bc)x + bd$.

We are told that a, b, c, d are all positive, and that indicates the coefficient of x^2 is negative because the negative of ac and that means this is a quadratic function whose parabola is downturned, therefore, we will be able to get a maximum value at some point and this is going to be at the vertex, we know that this is going to be at the vertex. So, the vertex is at $\frac{-b}{2a}$, that is because here we have a, b, c, d already.

Let us write it down more carefully, that is the $\frac{-(coefficient\ of\ x)}{2(coefficient\ of\ x^2)} = -\frac{(ad-bc)}{2(-ac)} = \frac{ad-bc}{2ac}$, is where we will get the vertex. And since we know that the maximum is going to occur at this particular x, we get the $x = \frac{ad-bc}{2ac}$.

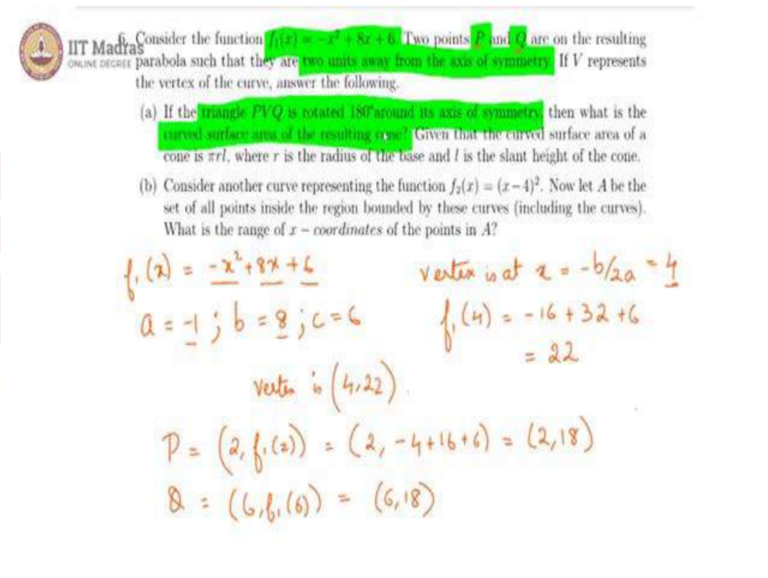
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial – 06

(Refer Slide Time: 0:14)

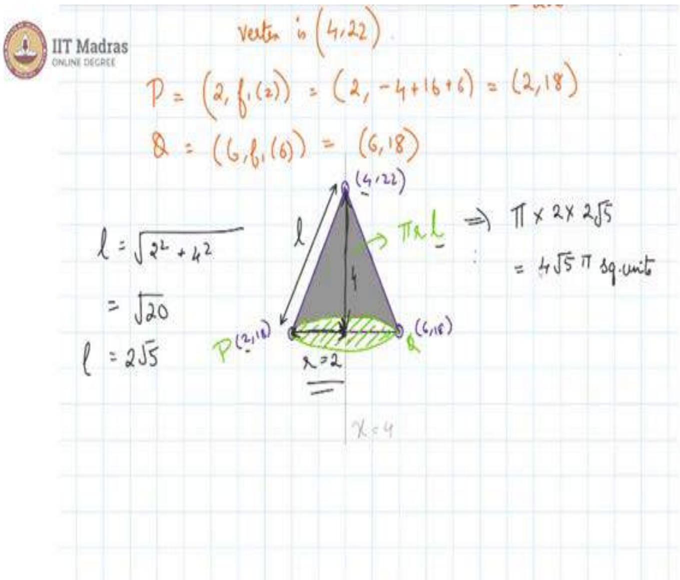


In our sixth question, we are given this particular quadratic function $f(x) = -x^2 + 8x + 6$. And we are told that two points P and Q, which are on this parabola such that they are two units away from the axis of symmetry. So, let us try to find out what the axis of symmetry is for this parabola. Our equation is $f_1(x) = -x^2 + 8x + 6$. And that would mean, in a standard form $a = -1$, $b = 8$, and $c = 6$.

And that would give us the vertex is at $x = \frac{-b}{2a}$, which in our case will then become $a = -1$, $b = 8$, so we will get 4. And the functions value at 4 is $f_1(4) = -(4)^2 + 8 \times 4 + 6 = 22$. So, the vertex is (4, 22). Further, we are told that P and Q are two units away from the axis of symmetry. So, the axis of symmetry is along $x = 4$, which means P and Q will be at $x = 2$ and $x = 6$, $4 - 2$ and $4 + 2$.

So, these points are going to be P$(2,\ f_1(2)) = (2, -4 + 16 + 6) = (2, 18)$. And the point Q is going to be $P(6,\ f_1(6))$ and from symmetry we know that this is also going to be 18, so $P(6,$ 18). And it is now told to us that the triangle PVQ is rotated 180 degrees about its axis of symmetry and we are being asked the curved surface area of the resulting cone. So, let us look at what this looks like.

(Refer Slide Time: 3:08)



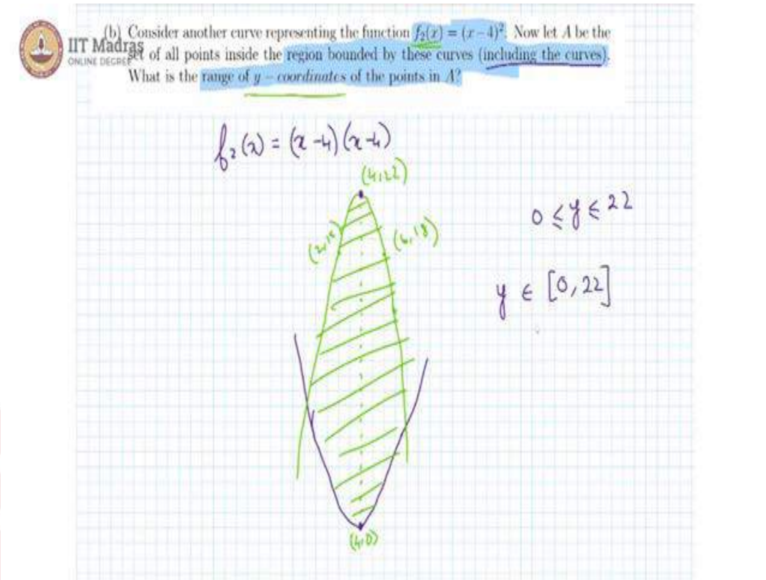
So, now let us suppose that this point here, let us call this our (4, 22), in that case 18 is 4 units below, so this will be the horizontal line passing through 18 and 2 will be here. So, (2,18) is here and this gives us (6,18) is here. This is (2,18) and this is (6,18). And that gives us a parabola which looks something like this, obviously a smoother curve than I have drawn, but something like this. And the triangle we are interested in is an isosceles triangle, which looks roughly like this.

This is the triangle that is being rotated 180 degrees about its axis of symmetry and its axis of symmetry is $x = 4$. I am erasing the parabola in order to focus on the triangle alone. If this triangle were to be rotated, this point which is our P, this is our Q, this point P basically goes around and reaches Q, whereas Q comes around and reaches P. And in this way, we have a cone in our hands and we want the curved surface area and that would be this region and the base circle is this flat surface below this is the base circle.

And we are interested in the curved region whose surface area is given to be $\pi r l$. So, what is r, r is the radius of the base circle. Which is basically then this quantity, this is r, which we can tell is $4 - 2$, so it is 2. And what is l over here, that is the slant height, which is basically this height, that height can be obtained as the hypotenuse of this base radius and height here, which is as we can see 4 units. So, $l = \sqrt{2^2 + 4^2} = \sqrt{20} = 2\sqrt{5}$.

So, we have $r = 2$ and $l = 2\sqrt{5}$, this gives us a curved surface area is $\pi \times 2 \times 2\sqrt{5} = 4\sqrt{5}\pi$ square units.

(Refer Slide Time: 6:46)



For the part B of our question we have another curve which is also quadratic and whose roots are basically 4 repeated. So, $f_2(x) = (x-4)(x-4)$. So, x being equal to 4 makes $f_2(4)$ =0. So, therefore, our root is 4 and it is repeated because coming twice here. So, let us now try to look at what they are asking. Now, let A be the set of all points inside the region bounded by these curves, including the curves. So, we are saying the region bounded by these curves and including the curves.

And they would like the range of y coordinates of points in it. We know already that (4, 22) is the vertex for our previous parabola. And it also passed through (2, 18) and (6, 18). And about this new parabola, the $f_2(x)$, we know that 4 is repeated root so there is only 1 root and therefore, at 4, that is 22, this would be 21, this is 20, this is 19, 18 17, 16, 15, 14, 13, 12, 1, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1, and 0. So, this is going to be the repeated root and the vertex of our other parabola.

So, if one parabola is like this, f_1 had negative x^2 coefficient so it is a downturned parabola, then the other parabola $f_2(x) = (x-4)^2$ is an upturned parabola which is going to be something like this. So, these curves are going to intersect in some way this way. And we are interested in the range of $y-coordinates$. So that would be, what are all the $y-coordinates$ possible in this region.

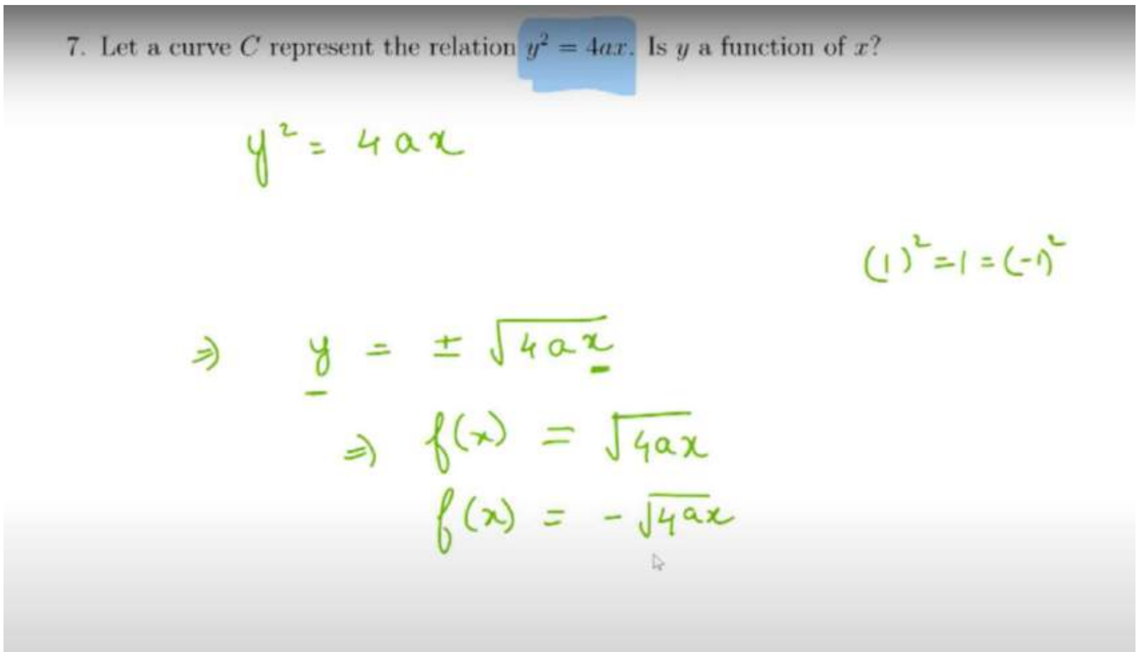
So, if this is the region we are looking at, then clearly this is the upper bound of our $y-coordinates$ and this is a lower bound. So, $y-coordinates$ in our region range between 0 and 22. And they said including the curve, so 0 is also included, 22 is also included, so we can write the same thing as $y \epsilon [0, 22]$.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial - 07

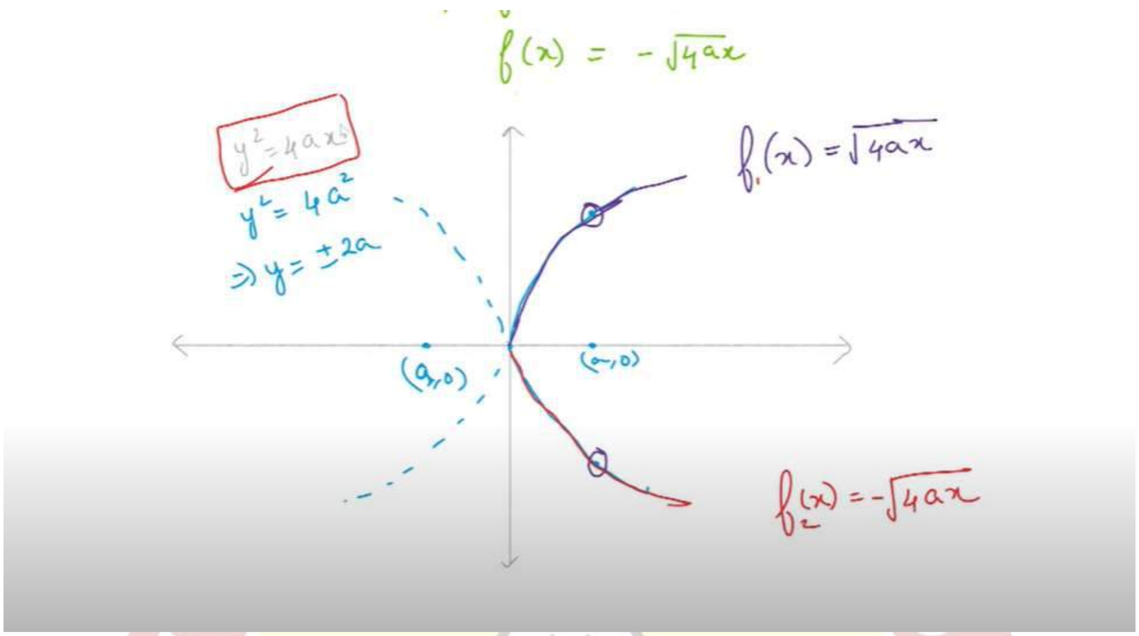
(Refer Slide Time: 0:04)



In question number 7, we have one relation given this way, $y^2 = 4ax$. And they are asking a very simple question, is y a function of x. So, we have $y^2 = 4ax$. And the interesting thing about square roots is, if I did the square root of 1, it is not just 1, it is actually ± 1. So, both $(+1)^2 = 1$, which is also equal $(-1)^2$.

So in this case, we need to consider the fact that y $=\pm\sqrt{4ax}$. Which means for the same x, I might have 2 different y's. So, put it this way, I am basically saying f(x) assuming it is a function is equal to $\sqrt{4ax}$ and f(x) is also equal to the $-\sqrt{4ax}$. And this is not allowed, for a single element in the domain, for a function, you should have only one image in the range. But here we have 2 different images for the same element in the domain. Therefore, this is not a function.

(Refer Slide Time: 1:55)



If we looked at it in terms of the plot, we have $y^2 = 4ax$, that is what we are trying to plot. And for x = 0, we get y = 0. So, this curve passes through the origin definitely. And for the next x value, I am going to take a, so therefore, $y^2 = 4a^2$, which gives $y = \pm 2a$. So, if a is positive, this is (a, 0), then 2a is going to be somewhere here like this, and -2a is going to be somewhere here like this.

And so, we have a parabola which goes something like this. And if a were to be negative, then this would have been (a, 0) and we would have a similar parabola in the negative direction. Either way, it is pretty clear that for a given value, you have two corresponding y values, for a given value of x you have two corresponding y values and that is not allowed for a function.

Independently $f(x) = \sqrt{4ax}$, which is this part of the curve, that can be treated as a function and $f(x) = -\sqrt{4ax}$, for convenience let us call this as f_1 and this is f_2. This is also possible to be treated as a function independently, but their combination which gives us this relation, that is not a function.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

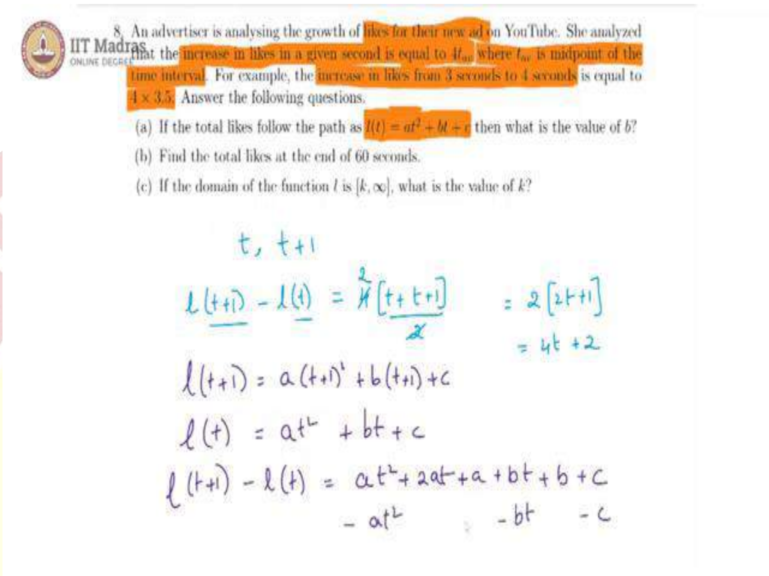
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Week - 04
Tutorial - 08

(Refer Slide Time: 0:15)

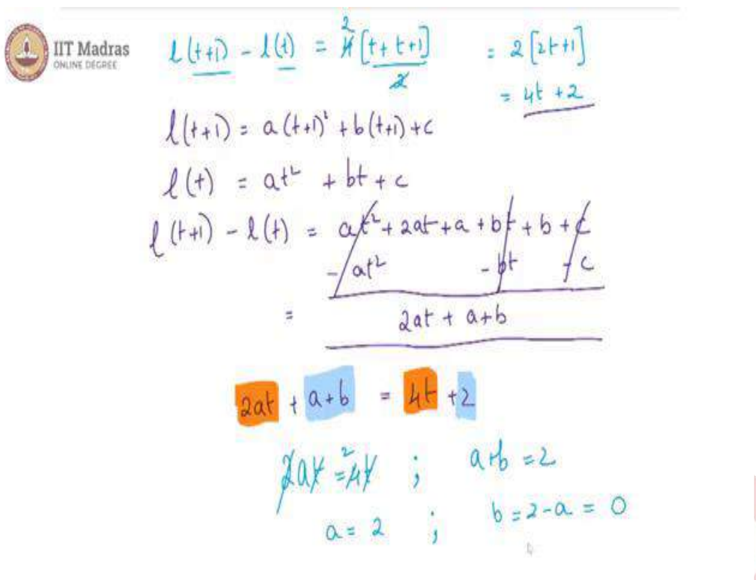


For our eighth question we have an advertiser who is analyzing the growth of likes for their new ad on YouTube. She analyzed that the increase in likes in a given second is equal to 4 times t_{av} , where t_{av} is the midpoint of the time interval, that is the average time in that time interval. And so we are given an example to explain what this is. The increase in likes from 3 seconds to 4 seconds. So, from the time $t = 3$ to the time $t = 4$, there is a number of increase in likes, which is equal to 4×3.5 and 3.5 is the midpoint of 3 and 4.

So, one way to write this is, let us look at time t seconds and the time $t + 1$ seconds. Then it is given to us that the likes at time $t + 1$, so, number of likes is a function of time. So, $l(t + 1) - l(t) = 4 \times t_{av} = 4 \times \frac{(t+t+1)}{2} = 4t + 2$, this is the difference in the likes from time t seconds to $t + 1$ seconds.

Now, it is further given to us that this particular function is a quadratic function. So, $l(t + 1) = a(t + 1)^2 + b(t + 1) + c$ and $l(t) = at^2 + bt + c$. Then $l(t + 1) - l(t) = at^2 + 2at + a + bt + b + c - (at^2 + bt + c) = 2at + a + b$

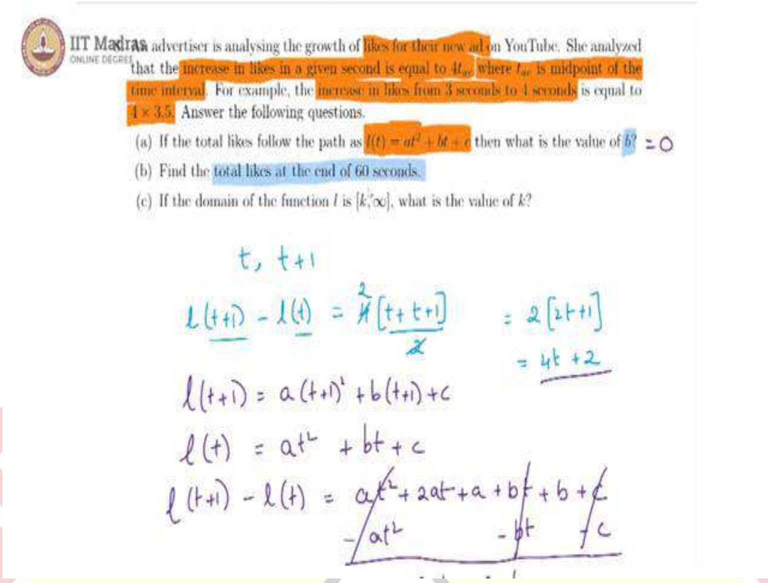
(Refer Slide Time: 3:08)



This quantity is basically equal to $2t + 4$. So, we are saying that $2at + a + b = 2t + 4$. Now, what are we supposed to acknowledge here is that the term with the t in it, that is the time dependent term is going to be same on both sides. Whereas the term which is constant is going to be same on both sides.

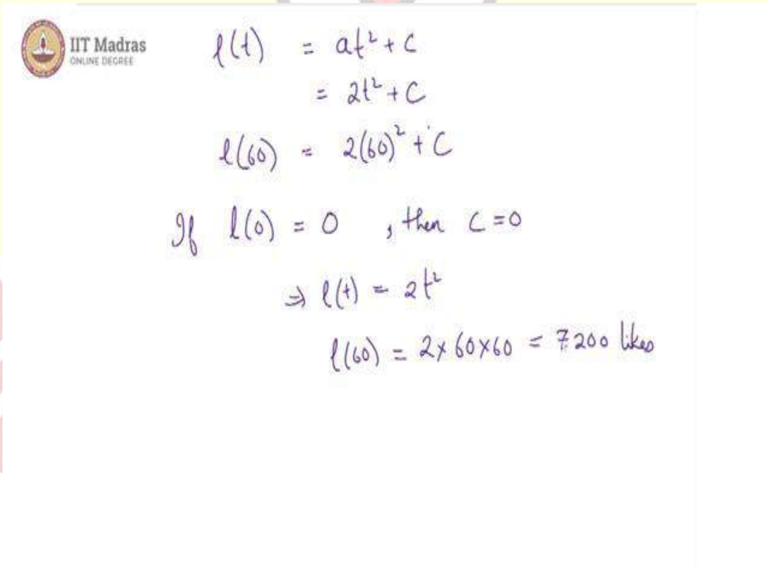
Thus, we are saying $2at = 4t$ and $a + b = 2$. This gives us 2 times t and t cancelled. So, we know $a = 2$ and that would imply $b = 2 - a = 0$.

(Refer Slide Time: 4:20)



And our question is asking us what is the value of b. So, we know this is equal to 0. Second question, the second part of the question is asking what is the total number of likes at the end of 60 seconds.

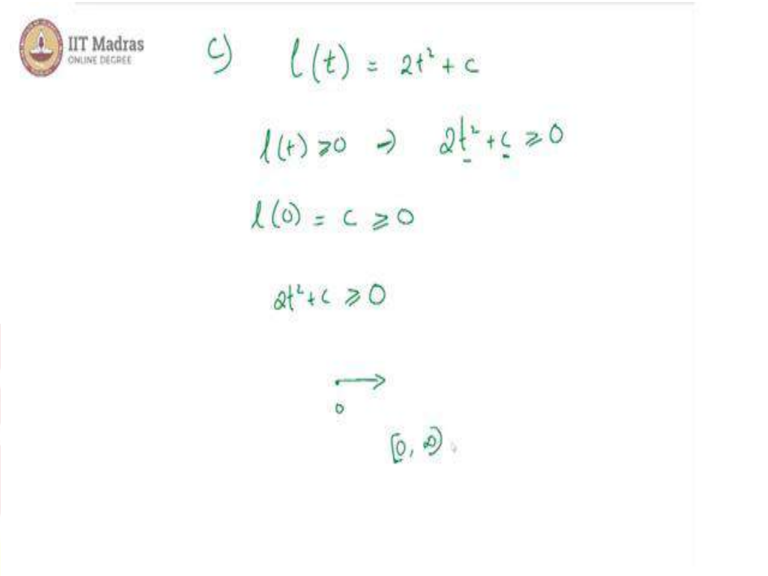
(Refer Slide Time: 4:49)



That would be impossible to calculate because we have the values of a and b, so we know that our $l(t)$, in this case we want l of 60. $l(t) = at^2 + 0t + c = 2t^2 + c$. But we do not know what c is. So, $l(60) = 2 \times 60^2 + c$. Now, if we made further interpretations that there were 0 likes at time $t = 0$. So, if $l(0) = 0$ then $c = 0$. So, this is a particular assumption we are making, we are assuming that the timer started when the likes were 0 and that would imply your $l(t) = 2t^2$.

So $l(60) = 2 \times 60 \times 60 = 7200$, that is 7200 likes at the end of 1 minute.

(Refer Slide Time: 6:18)



And lastly, for Part C, we are being asked the domain of the function is $[k, \infty)$, what is the value of k. We know that $l(t) = 2t^2 + c$. Now only real requirement we have is that our likes be greater than or equal to 0. So, $l(t) \geq 0 \rightarrow 2t^2 + c \geq 0$. Another thing we have is clearly that $l(0) = c \geq 0$, because at 0 time, it is not like you can have negative likes. So, $c \geq 0$.

Now we know that $t^2 \geq 0$ and now we also found that $c \geq 0$. So, $2t^2 + c \geq 0$, which means any time that is 0 or greater than 0. So, we are looking at the timer being started at a particular time and from there on, if this is 0 from there on your function is well defined and the number of likes will be greater than or equal to 0. So, the domain will be all the time from 0 seconds to ∞.

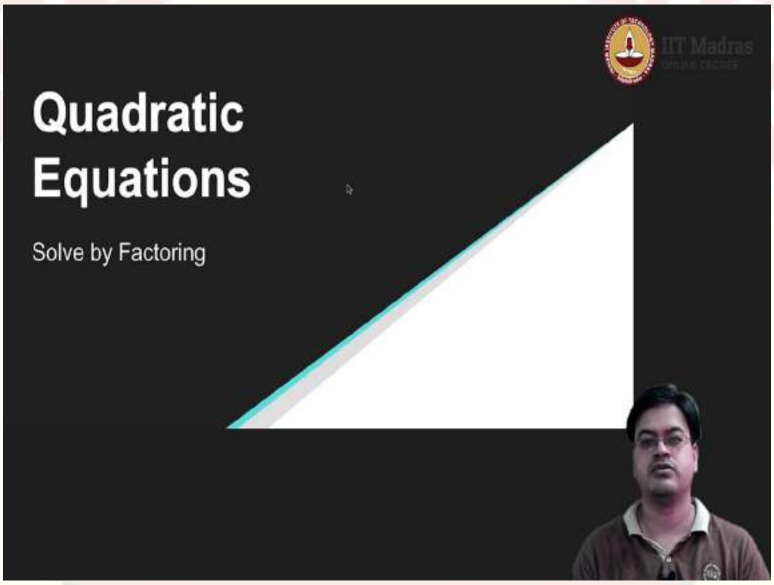
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

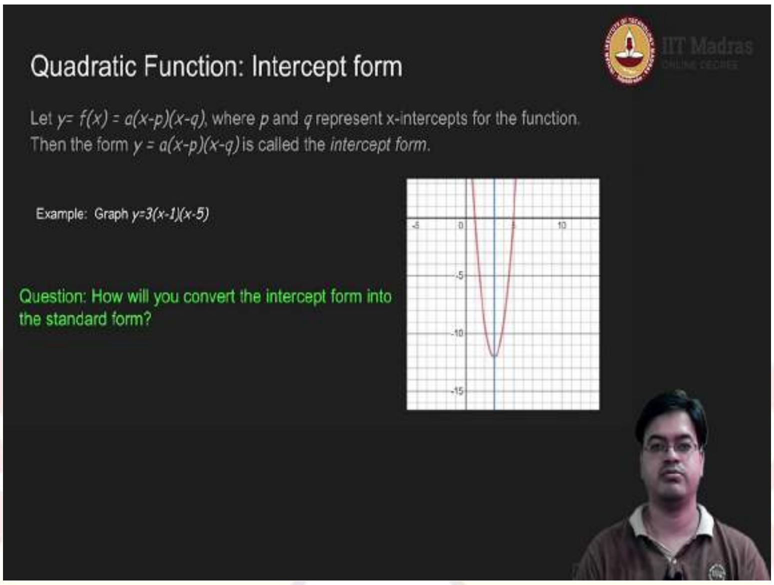
Lecture – 27
Solution of Quadratic Equation Using Factorization

(Refer Slide Time: 00:15)



So, in the last video, we have seen how to find the roots of a Quadratic Equation by graphing the associated quadratic function.

(Refer Slide Time: 00:37)



In this video, we will see how to find the solution of a quadratic equation by a well known method called factoring method. For that, we will define one new form of a quadratic function that is intercept form.

What is an intercept form? If $y=f(x)$ which is a quadratic function is written in this form , where and $(x-q)$ are called binomials ok. Where this p and q are nothing but x intercepts of the quadratic function.

So, essentially what you have done is, you have seen a graph of a function and you have located the two intercepts x intercepts of the function. Whenever the expression is possible in this form you are writing it. So, $y=a(x-p)(x-q)$, and this form is called the intercept form.

So, let us try to see one example of intercept form which is let us say you have been given this intercept form $y=3(x-1)(x-5)$. And you also know that these 1 and 5 are x intercepts of the quadratic function, that means, you have been given two values. Can you find the third value? The answer is yes.

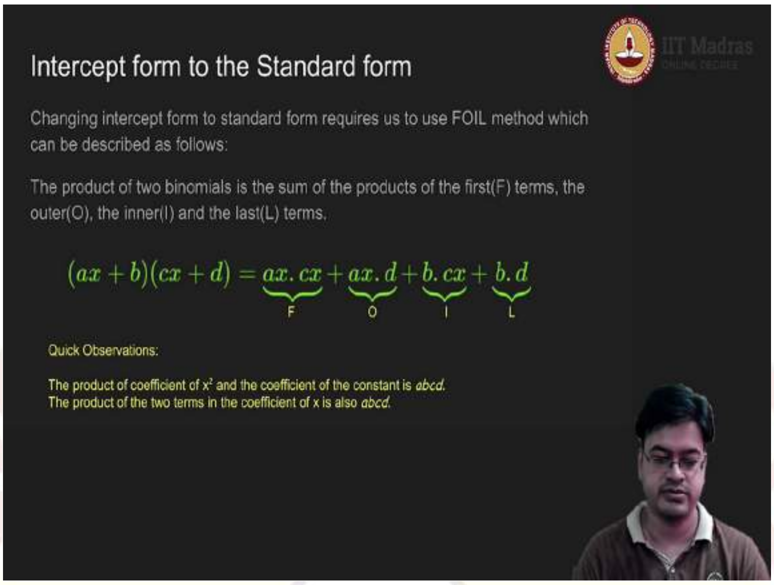
So, at point 1, when , the value is 0; at point 5, the value is 0. Now, using the logic that I gave you in the previous video, you can actually see that there will be some axis of symmetry between this 1 and 5, because 1 and 5 both take value 0 right. So, the axis of symmetry will be nothing but the distance between these two points divided by 2. So, $\frac{1+5}{2} = 3$. I am sorry it will not be a distance, it is just the sum of these two points divided by 2, average of these two points, that should be the correct terminology. So, it should be average of two points. So, $\frac{1+5}{2} = 3$

Now, $x = 3$ will be the axis of symmetry for this particular function if at all the this are the roots of the this are the x intercepts of the function right. So, now, I have given you $x = 3$ is the axis of symmetry. So, just substitute the value 3 in this particular graph in this particular equation, and you will get 3 - 1 = 2, and 3 – 5 = -2, that means, $2 \times -2 = -4, -4 \times 3 = -12$. So, you got three values. What are those three values? (1, 0), (5, 0), (3, -12).

So, based on this information, you can easily plot the graph which will look like this. As you can see this is the value -12 here, value -12 is here and 0, 0. So, you can easily plot this graph. You can connect the smooth curve using this.

Now, the main question is once given this kind of expression, how to convert this expression into a standard form? So, that is the question that I will post down. How will you convert the intercept form into the standard form? Just by multiplying the two binomials. So, for multiplying the two binomials, we have one rule which I will state in the next slide.

(Refer Slide Time: 04:17)



So, how the conversion from intercept form to standard form will happen using a method called FOIL method which is described below. So, the product of two binomials as I already mentioned that $(x-p)$ and $(x-q)$ these are the two binomials. So, the product of two binomials is sum of the product of first terms the outer, the inner and the last terms. So, let me make it more precise by demonstrating it.

So, let us consider this expression which is . Now, $(ax+b)(cx+d)$ what I will do is I will first take the first term of this expression, and the first term of this expression, and multiply them together that is the first term over here by the sum of the product of the first terms I mean this term.

Then I will take the inner term ok, then I will take the outer term, sorry, then I will take the outer term that is b is the outer term here, sorry, not b is not the outer term, b is the inner term. You have ax which is the outer term, and d which is the outer term. So, now you just multiply them together which gives me $ax.d$ which is the outer term product of the outer term.

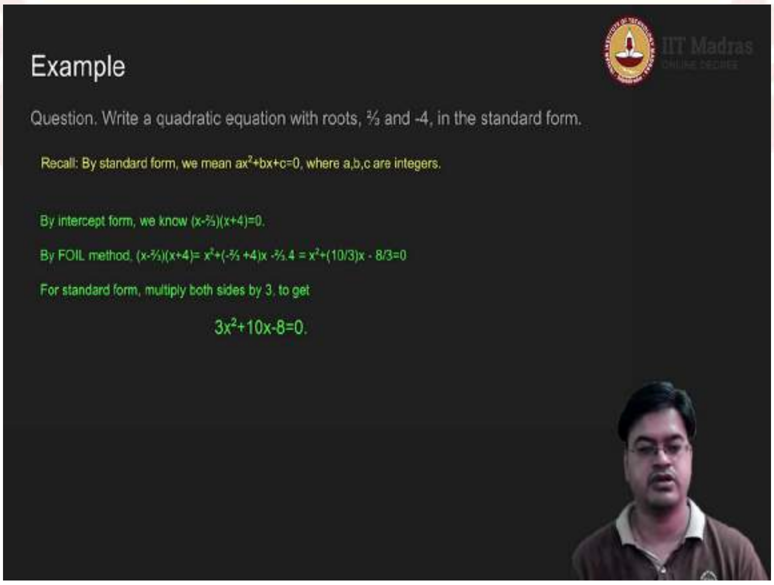
Then you take the inner terms that is b and cx. So, $b.cx$, this is the inner term. And $b.d$ are the last terms. First term, outer term, inner term and last term that way we will multiply these things together. That means, I will get the first term as acx^2 ; second term is $(ad+bc)x$; and the third term as bd .

Now, if you look at, so basically the ac is the term which is the coefficient of x^2 , $ax+bc$ is the $ad+bc$ is the term which is the coefficient of x , and bd is the term which is the which is the constant term right.

So, now, a quick observation you can make is if you look at the product of this first term and the last term, what will you get $ac \times bd$, so it is $abcd$, the product is $abcd$ right. So, the product of the coefficient of x^2 and the coefficient of constant is $abcd$. In a similar manner, if you consider the coefficients of x , a d and b c are the coefficients of x, $ad+bc$, so if you take product of these two terms, again it will be $abcd$. This is a crucial observation which we will need while converting a standard form to intercept form and vice versa.

Just remember this the product of the coefficient of x^2 and the coefficient of constant is $abcd$. And the product of the two terms of the coefficients of x is $abcd$; both of them are $abcd$. So, this we will use to convert our expression into standard form, and convert our expression in intercept form in various ways. So, that observation is very crucial for us.

(Refer Slide Time: 08:03)



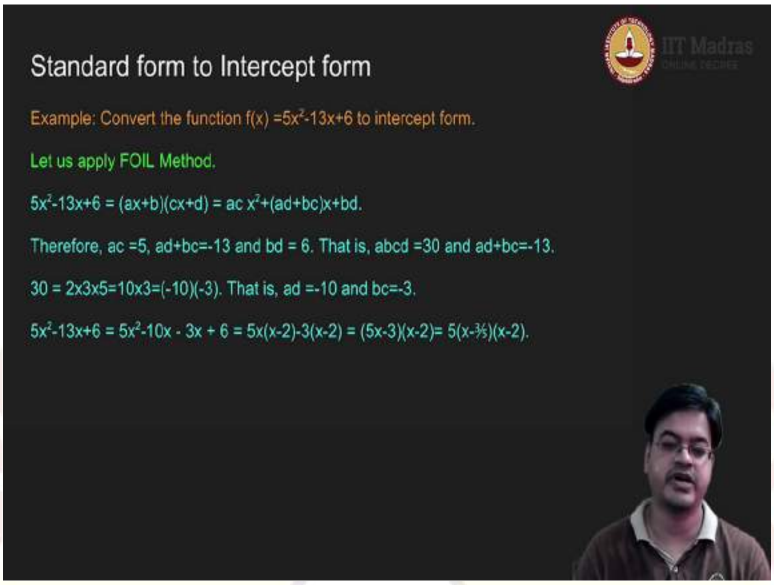
So, let us do take one example and see how we can apply our knowledge which we have gained in this particular video along with the previous videos to solve this problem. So, the question is the write a quadratic equation with roots 2, $\frac{2}{3}$ and -4 in the standard form ok. So, let us recollect what is a standard form. Standard form is of $ax^2+bx+c=0$; and a, b, c all are integers; and $a \neq 0$. This is the standard form; we have already seen that ok.

So, now, if I want to write this, we will use our knowledge about intercept form, and we can easily write this expression as $(x-\frac{2}{3})(x+4)=0$ because $\frac{2}{3}$ and -4 are the roots. Yes, but this equation is not in the standard form. So, now in the previous slide, we have seen that in order to convert this into a standard form, we will use a FOIL method. So, let us try to use a FOIL method. So, the what is a here? $ax+b$ that is a is 1, b is $\frac{-2}{3}$, c is 1, d is 4 ok.

Now, you use FOIL method that is first terms. So, first terms is 1×1 , so it will retain $1x^2$, Then $ad+bc$, so $(\frac{-2}{3})(1)$ and a is 1, 4 $4.-\frac{2}{3}$ that is the product here $\frac{-2}{3}+4$ this is the term which have coefficient of x , and then $\frac{-2}{3}.4$ which is the term here. So, this is successful application of FOIL method.

Now, let us rewrite all these things that is you can sum this and write the sum that is x^2 , so $4\times 3=12, 12-2=10$, so $x^2+\frac{10}{3}x-\frac{8}{3}$ right. Is this equation in the standard form? No, because for standard form a, b, c, all must be integers. So, what I will do is, I will multiply this equation with 3 on both sides. So, if I multiply on both sides with 3, then I get $3x^2+10x-8=0$ that is the solution to this question. So, the quadratic equation in standard form is $3x^2+10x-8=0$. So, we have solved.

(Refer Slide Time: 11:09)



Let us now go further and try to see how I will convert a standard form into an intercept form. Again we will use FOIL method, but in a reverse manner. So, I want to convert a function quadratic function which is given to me $5x^2-13x+6$ to intercept form, that means, I want to write $a(x-p)(x-q)$. So, how will I convert this?

So, what I will do is, I will take this particular function $5x^2-13x+6$, and apply FOIL method to it. How to apply FOIL method to it? I will equate this to be equal to $(ax+b)(cx+d)$. Then based on FOIL method, I have this expression which is ac^2 sorry $acx^2+(ad+bc)x+bd$. Now, remember we have done some observations that is the product of this and this is $abcd$ right.

So, now, I can equate this equation with this equation. So, term containing x^2 will be equated with term containing x^2. So, I will get $ac=5$, $ad+bc=-13$ and $bd=6$. Then from this expression I can also derive an information that is $abcd$ that is the product of the first and the last term and the product of the terms contained in the sum is 30. So, $5\times 6=30$; and $ad+bc=-13$.

Now, my job becomes crucial. My job is to guess what those two terms will be ad and bc right. So, that their product is 30, and if you sum over them, then it must be -13. For that

I will use the prime factorization theorem that was introduced in week-1. So, if you look at this expression 30, I will get prime factors as $2\times 3\times 5$.

Now, I want the product to be equal to 30, and I want the sum to be equal to -13. So, based on this, what I can derive is if at all this, this term has to be negative, I should have some negative factors over here and both of them should be negative factors. In particular if I combine 5 and 2, I will get 10 and 3, and 10 + 3 =13, but it is not giving me -13.

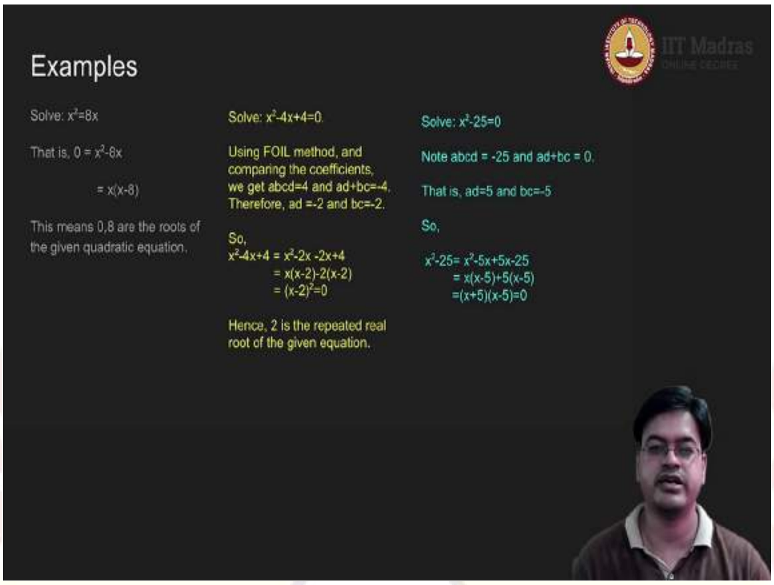
So, I will use a trick that multiplication of two negative numbers will become a positive number. So, it is $(-10)(-3)$ which will give me 30; at the same time, it will be the sum will be -13. So that means, my ad is -10; bc is -3. It does not matter, you can switch also. You can write bc as -3, and ad as -10 also, it does not matter.

So, now I will substitute these values into this expression, essentially I will rewrite this expression. So, I will write this expression as $5x^2-10x-3x+6$ ok. Then what I will do is I will look at the first two terms; first two terms, and I will take the greatest common factor from these two terms that is $5x$. So, I will take $5x$, whatever is remaining I will put in a bracket that is $x-2$.

Here also I will do take the greatest common factor out that is -3, so $-3(x-2)$. Now, you can see these $(x-2)$ s are same. So, essentially this expression will come if I have $(5x-3)(x-2)$. Now, is this in the intercept form? No, still it is not in the intercept form. What is the intercept form? It is $a(x-p)(x-q)$. So, I will just divide everything by 5 in this expression and take the 5 out. So, $5(x-\frac{3}{5})(x-2)$ this is the intercept form.

So, using FOIL method, I have converted this expression into an intercept form. An expression was given to me in standard form; I have converted it into intercept form. Let us see few more examples as this concept is quite intricate. You may need some practice, you solve as many problems as possible, but I will give you some demo cases, so that it will be easy to distinguish for you.

(Refer Slide Time: 16:05)



So, let us take let us say you have you have been asked to solve this equation; $x^2=8x$. Now, here you do not need, you do not really need a FOIL method. What you need is, just rearrange $x^2=8x$ and you just take out the greatest common factor which is x. So, x this will give you $x(x-8)$. So, if I want to solve this, I know $x-0$ and $x-8$ are the things. So, 0 and 8 are the roots of this given quadratic equation. Simple, this solves our problem for such a simple case, where the constant term is absent right.

Now, let us take another example, $x^2-4x+4=0$. Now, in this case, you will use FOIL method obviously but the essence of FOIL method reduces to that the coefficients of x are of the form $ad+bc$, and the product of this and this is $abcd$. So, the product $abcd$ is 4, and $ad+bc$ is -4, this is what it reduces to ok.

So, if abcd is 4, and ad + bc is -4, is there any other way out 4 can be factorized only in one way that is 2×2; 2×2. And ad + bc is -4, that means, both of them should be negative -2×-2. So, ad is -2, bc is -2. Substitute it in the master equation where you can write $x^2-2x-2x+4$. So, you have substituted it in master equation.

Now, you go ahead and take out the greatest common factors out, the first expression will have x out, the second expression will have 2 out. Then again these are product of binomials. So, it will be $(x-2)^2$ = 0 given in the expression. So, what is the root of this equation? 2; 2 is the real repeated root of this equation.

Let us go ahead and solve one more example. $x^2-25=0$. This is quite interesting 25, you can see is a perfect square 5, and I want to find the root of this equation. Again I will use FOIL method, abcd is -25, and ad + bc is 0 right. 5 is a perfect square. So, 5×5 is the factorization, but -25 is there. So, one will be, one 5 will be with a positive sign, another 5 will be with a negative sign. So, ad can be +5, and bc can be -5 or vice versa, it does not matter.

Substitute this substitute this knowledge into this expression. So, you take $x^2-25-5x+5x$, take out the greatest common factors that is x and 5 respectively, you will get this kind of expression. And then you just rewrite them as product of two numbers that is $(x+5)(x-5)=0$, and you have solved. Remember in all these expressions we have written this in intercept form.

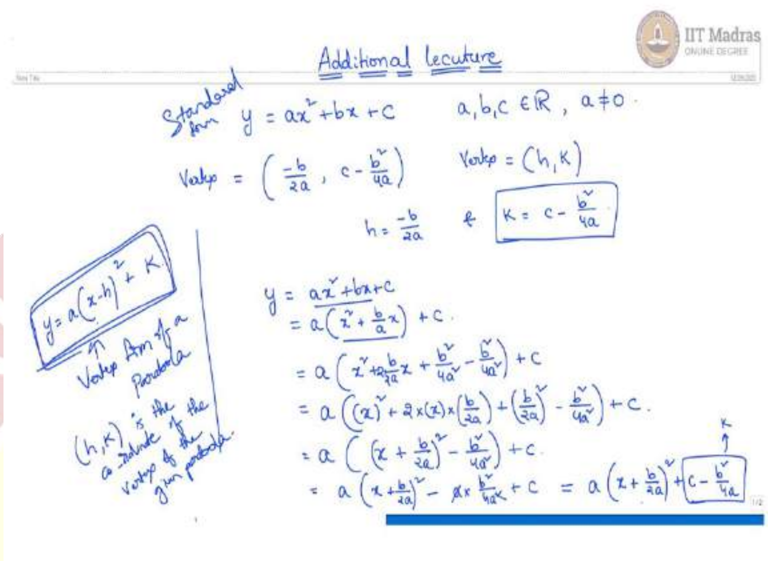
So, all these expressions are written in intercept form. Once you write an expression in intercept form, it is very easy to find the roots of the equation, or in fact once you write in the intercept form you have already figured out the roots of the equation.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 05 - Additional Lecture

(Refer Slide Time: 00:34)



Hello, everyone, today we will discuss a small topic related to the forms of parabola. In other words we are going to see the relation between the standard form of a parabola and the vertex form of the parabola, we already know the standard form of a parabola which is given by $y = ax^2 + bx + c$ where these a, b, c belong to real and $a \neq 0$. From this standard form we will try to derive the vertex form of a parabola.

Now, from this equation we know that the coordinate of the vertex of the parabola is vertex will be x coordinate will be $\frac{-b}{2a}$ and y coordinate will be $c - \frac{b^2}{4a}$, this we already see in the previous lecture. Now, let us denote this coordinate of the vertex as (h, k). So, our h will be nothing but $\frac{-b}{2a}$ and k will be $c - \frac{b^2}{4a}$ .We have obtained the required data, now let us start the deriving.

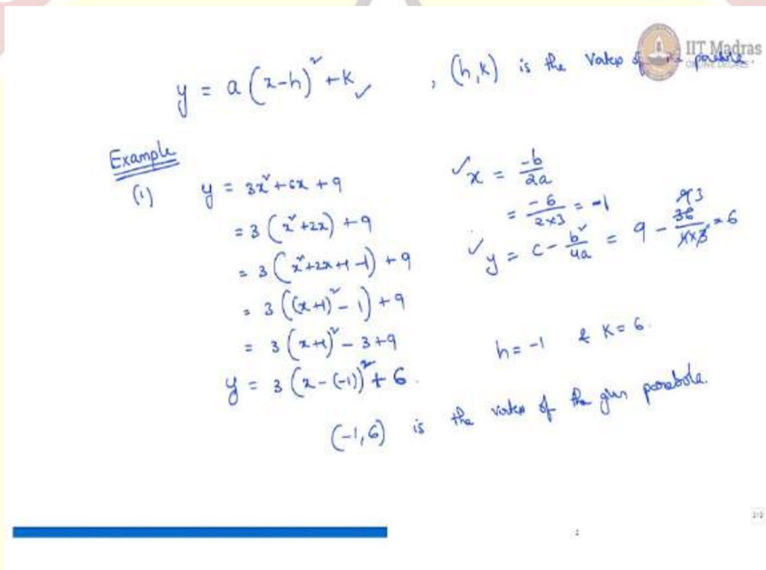
We have $y = ax^2 + bx + c$, I will take a common from these two terms, I will get $a\left(x^2 + \frac{b}{a}x\right) + c$, also I will add and subtract $\frac{b^2}{4a^2}$ to this term, I will get $a(x^2 + \frac{b}{a}x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2})$+c.

Now, also I will multiply with 2 and divide by 2 here, now I will rewrite this $a(x^2 + 2 \times x \times \frac{b}{2a} + (\frac{b}{2a})^2 - \frac{b^2}{4a^2})$+c. So, if we observe this $x^2 + 2 \times x \times \frac{b}{2a} + (\frac{b}{2a})^2$, this is in the form of $p^2 + 2pq + q^2$, we can write this as $(p + q)^2$.

So, writing like that, we get $(x + \frac{b}{2a})^2 - \frac{b^2}{(2a)^2} + c$. Now, if I multiply a we get $a[\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a^2}]$+c, a and this cancelled, finally I will obtain this will be equal to $a\left(x + \frac{b}{2a}\right)^2 + c - \frac{b^2}{4a}$.

If we observe $c - \frac{b^2}{4a}$ is k here and $\frac{b}{2a}$ will be $-h$, so if we substituted h and k in this equation we get $y = a(x - h)^2 + k$, this is the vertex form, vertex form of the form of a parabola, where this $(h\,, k)$ is the coordinate of the vertex of the given parabola, this is the standard form and from this standard form we derived the vertex form.

(Refer Slide Time: 05:43)

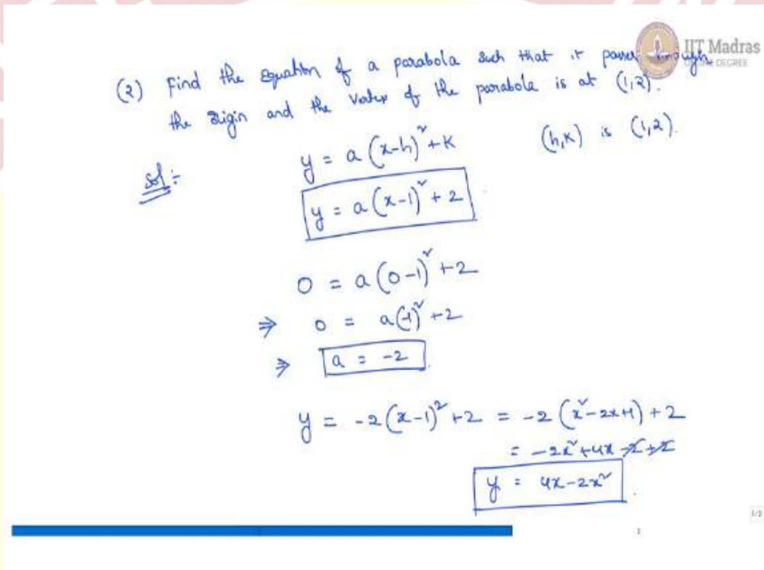


So, we have got the vertex form of the parabola which will be like this $a(x - h)^2 + k$ where $(h\,, k)$ is the vertex of the parabola. Now, let us see one example to understand this vertex form clearly. So, suppose we have an equation of a parabola given like this $y = 3x^2 + 6x + 9$ now we try to write in vertex form, so I will take 3 common from the first two terms, I will get $x^2 + 2x + 9$, so in order to make this a perfect square I will add 1 and subtract 1.

So, $3(x^2 + 2x + 1 - 1) + 9$, so 3 times this can be written as $3((x + 1)^2 - 1) + 9$, which gives us $(3(x - (-1))^2 + 6)$. So, we have got the equation $y = (3(x - (-1))^2 + 6)$. So, if we equate it with this vertex form we get $h = -1$ and $k = 6$. So, our vertex will be at point $(-1\,, 6)$ is the vertex of the given parabola.

So, we will just cross verify it, we know that if we have a standard form we can calculate the x coordinate of the vertex, so x coordinate of this vertex will be $x = \frac{-b}{2a}$, so here b is 6 and a is 3, so if I substitute that$-6\ by\ 2\ \times\ 3$ which I will get $x\ = -1$ and we know the y coordinate as $c - \frac{b^2}{4a}$ here we have c is $9\ -\ b$ is 6 so b^2 is 36 / $4a$ is 3 so 4 9's 9 3's, so I will get 6. So, my vertex point will be at $(-\ 1\ , 6)$, if we solve y, solve through standard form also.

(Refer Slide Time: 09:02)



Now, let us see one more example, find the equation of a parabola such that it passes through the origin and the vertex of the parabola is at $(1\ , 2)$. So, as we know the vertex form given by y is equal to a times x minus h whole square plus k, here we have given that $(h\ , k)$ is nothing but (1 , 2).

So, if we substitute that our equation will be simplified to $a(x - 1)^2 + 2$, also it is given that this equation passes through the origin that means 0 , 0 should satisfy this equation. So, if we substitute it we get the value of a, so $0 = a(0 - 1)^2 + 2$, this implies $0 = a(-1)^2 + 2$, which implies again $a = -\ 2$.

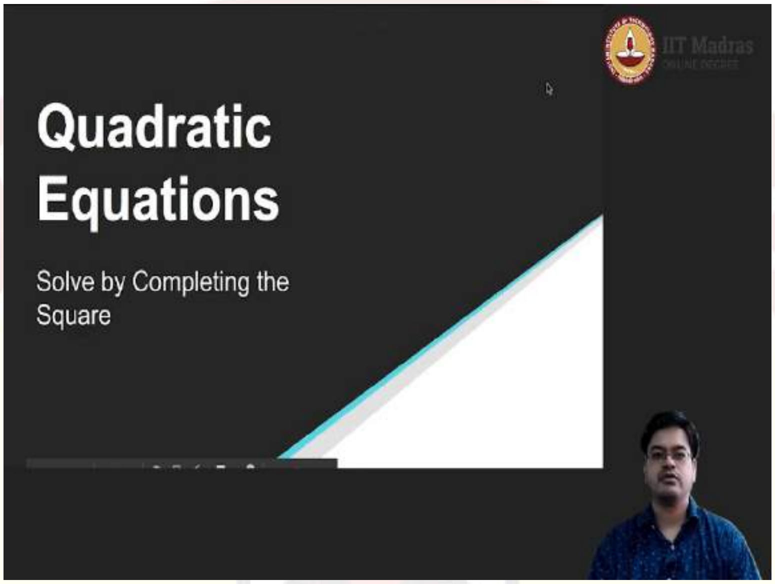
So, our final equation of the parabola will be equal to $y = -2(x - 1)^2 + 2$, if you open that $y = -2x^2 + 4x -\ 2 +\ 2$, so this will be get cancelled and $4x - 2x^2$, so $y = 4x - 2x^2$ is the equation of the parabola that passes through the origin and the vertex of this parabola will be at $(1\ , 2)$. Thank you.

IIT Madras

ONLINE DEGREE

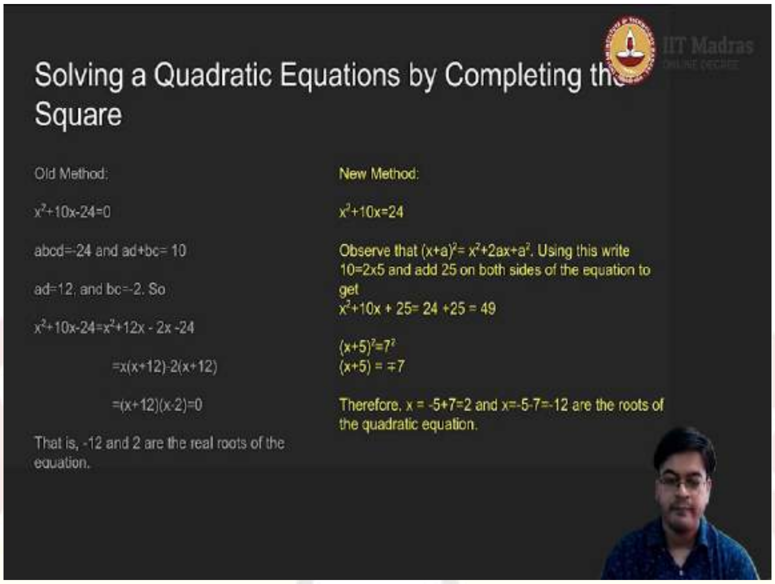
Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Solution of quadratic equation using Square method
Indian Institute of Technology, Madras
Lecture 28

(Refer Slide Time: 00:14)



In this video we are going to learn one more interesting method called Solving quadratic equations or for finding the roots of quadratic equation. The method named completing the Square method also it has a very good connection with a very well-known or very popularly known as Quadratic formula. So, we will explore that connection towards the end of this video.

(Refer Slide Time: 00:44)



So, let us start I will demonstrate this method through some examples. So, let us first understand or revise what we done in the earlier stage or in the earlier video. We have used a method of Factoring that I have called as old method. So, let us take this example $x^2+10x-24=0$. If you use the method of factoring you need to identify what is this term 24 and 1. So, -24×1 so - 24 is the product of the leading coefficient and the constant term and $ad+bc=10$.

So, I have this setup which is $abcd=24$ and $ad+bc=10$. So, I will essentially use a prime factorization theorem and get the prime factors of -24 so that if you rearrange the prime factors in such a way that the sum should be equal to 10. One such rearrangement is 12 and -2. So, ad is 12 and bc is -2. So, I got this and then based on our factorization technique I will substitute this 12 and 2 for this coefficient if x and I will get this expression which is $x^2+12x-2x-24$.

Then I will use the greatest common factor technique that is I will take out x common, 2 common from the last 2 terms and I will get this expression. So, finally, I got $(x+12)(x-2)=0$ and therefore, I will get the roots of the equation are -12 and 2. Now, somebody came up and thought, that why should I bother what is this last term? It is just a constant right, so I will replace this constant with something and I will work on it.

So, from that particular though comes the new method, which is the method of completing the square. So, what that person did is just rewrite this expression into this form that is $x^2+10x=24$. Now, the next question that person asked is if I look at this x^2+10x, do I know something that will make this particular expression as a complete square. So, what do I mean by complete square let us understand?

So, complete square means $(x+a)^2$ if I want to write $(x+a)^2$ then what I need to do here is to add some number and subtract some number or to add this same number on both sides right. So, in this case if you look at this expression that is $(x+a)^2$ which is $x^2+2ax+a^2$. So, this a is the number that I am looking for. Now, in this case if I consider this expression and if I want to add a number which will typically be a^2 what that a^2 should be, Is the first question.

So, to answer that let us equate this 10 to this 2a, so $10=2a$ therefore $a=5$ is the answer. So, what will be a^2? a^2 will be 5^2 which is 25. So, now I got a number a^2 to add and subtract from both sides. So, what I will do is I will add the number 25 on both sides so once I add the number 25 on both sides for this expression. I get $x^2+10x+25=24+25$, it turns out here in this case that the number is 49 which is also a perfect square.

But it need not be the case all the time. So, now what I know here is this number this particular expression is nothing but $(x+5)^2$ from this formula. Formula that is given here and then what is other side is 7^2. So, I can rewrite this expression as $(x+5)^2=7^2$ wonderful.

So, I got something in terms of squares, now had it been only one square then in the situation was easy I would have equated to $x+5=\pm 7$. But there will be two situations because both the terms on the left-hand side and on the right-hand side are squares. Now, you just write $x+5=\pm 7$ and $-(x+5)=\pm 7$ that will give us four cases.

But if you look at these two expressions, they will eventually reduce to the same two expressions that is $x+5=\pm 7$. So, it is sufficing to consider only two equations $x+5=\pm 7$. Once I have considered this then I know the solution right, so it is just a matter of substituting the values and doing some little bit of algebra. So, you subtract 5 from both sides so $x = -5 + 7$ which will give me 2, and $x = -5 - 7$ which will give me -12.

These are the roots of the quadratic equation, these exactly match with the roots that we have got -12 and 2, and here 2 and -12. So, the solution set is same therefore now it is a personal choice which method to prefer but what is a choice that is available if you have some difficulty in factoring this. Let us say this is not 24 but some absurd number and you have some difficulty in factoring this finding prime factorization.

What you are doing here is you are not using this particular property when you are doing this example. When you are solving this example through this method you are not using this particular property so you can get rid of this property and you do not have to worry about. One note of caution is you cannot use this method when the number given here becomes negative in this side because square root of a negative number is not defined.

So, after adding this a and the number still remains negative you cannot use this method because according to this method there is no real solution whereas we do not know. So, this method had some limitations but it is quite powerful in solving the problems okay. We will come to how to overcome the limitations in a certain in the next few slides and we will see its beautiful connection with the quadratic formula.

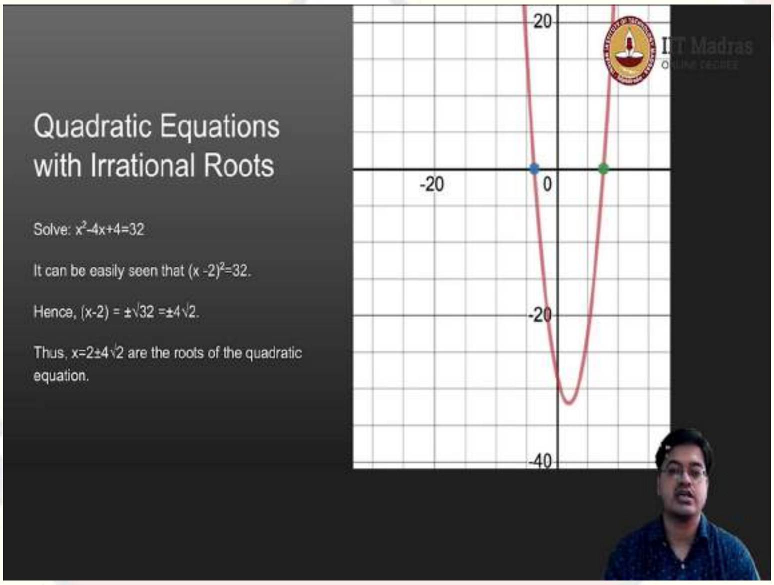
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 29
Quadratic formula

(Refer Slide Time: 00:15)



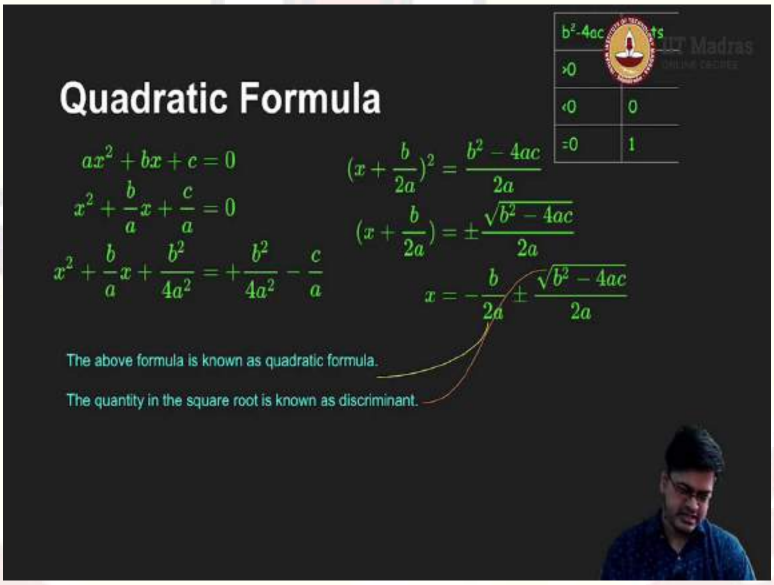
So, let us now go when the right-hand side is not a perfect square. In this case the right-hand side was a perfect square. So, when right hand side is not a perfect square, let us take this expression where; $x^2-4x+4=32$. So, I have already done the first step, I have added it. So, you can see the left-hand side is nothing, but $(x-2)^2$ which is equal to 32 and 32 is not the perfect square.

So, in such case what will happen? So, you will equate you will go in a similar manner $(x-2)^2=32$, you will take a positive square root of the left-hand side and $\pm\sqrt{32}$. $\sqrt{32}$ can be decomposed into 16×2. So, $\sqrt{16}$ is 4. So, it is $\pm 4\sqrt{2}$.

So, you will get 2 two roots; $2\pm4\sqrt{2}$ are the roots of the quadratic equation and they are irrational roots because, $\sqrt{2}$ is an irrational number ok. It is interesting to verify this result using a graph because, that will give us the clear cut understanding where does this $2+4\sqrt{2}$ are mapped. The two green dots over here represent the location of the roots ok. So, this is how you will solve a quadratic equation using the method of completing the squares.

Now let us explore the relation between quadratic equations method finding the roots using quadratic equations. Sorry; finding the roots of quadratic equations using completing the square method and its connection with quadratic formula.

(Refer Slide Time: 02:05)



So, for this let us take a general quadratic function equated to 0 that is; $ax^2+bx+c=0$. I n is the second step is because $a\neq0$. I can easily divide a I can easily divide by a; so that will give us the second step.

Now, as far as we can understand what we have here is, $\frac{c}{a}$ is the constant term. So, as per if we go by our method of completing the square, we will push this $\frac{c}{a}$ on the other side. So, it will take a negative sign that is what you are seeing here $\frac{-c}{a}$.

And, now, $\frac{b}{a}$ was the term $\frac{b}{a}$ was the term, if it is a complete square $\frac{b}{2a}\times 2$ will have come. So, I will get a term which is our a in the earlier expression that will be $\frac{b}{2a}$. So, $(\frac{b}{2a})^2$ will be $\frac{b^2}{4a^2}$, which is the term that I will add on both sides. So, I have added on both sides $\frac{b^2}{4a^2}$.

Now, look at this expression carefully, what is this expression? This is $(x+\frac{b}{2a})^2$ and this is some constant. So, I will use this that is; $(x+\frac{b}{2a})^2$ is equal to we can rearrange this term, $4a^2$ is the LCM. So, you just multiply by 4a over here you will get $4a^2$ there and divide by 4a so you will get $\frac{b^2-4ac}{2a}$. This is what you will get if you rearrange these terms hm. Sorry, it is wrong.

It is $\frac{b^2-4ac}{4a^2}$ it is. This is wrong. It should be $4a^2$ and then once you take the square root of this then you will get $(x+\frac{b}{2a})=\frac{\pm\sqrt{b^2-4ac}}{2a}$.

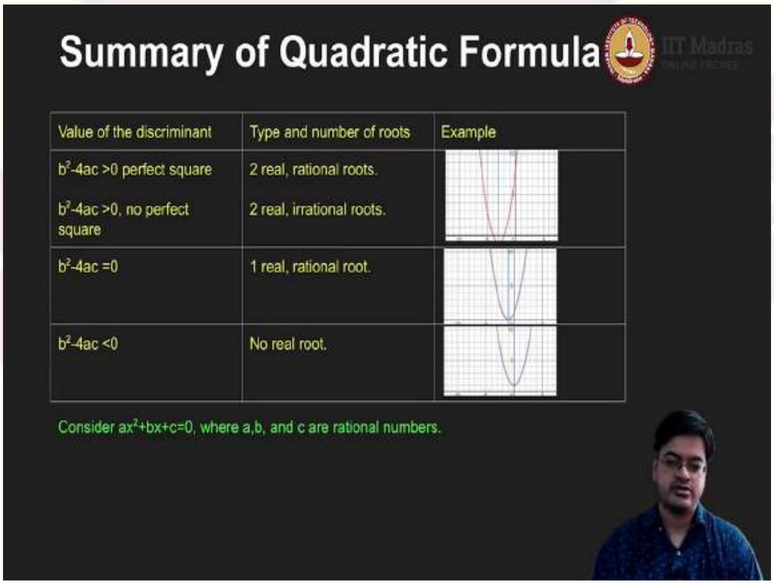
So, effectively using the same method of completing the square, the root of this equation will be $x = \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a}$; as easy as this. So, this method is very powerful; this is what we have done using method of completing the squares and it gives us a general formula which is called quadratic formula. So, this formula is known as quadratic formula and the term over here in the square root is known as discriminant.

What is the quadratic formula? This complete expression on the right-hand side is a quadratic formula. And the term in the square root since $b^2 - 4ac$ is called the discriminant. Why? Because it discriminates.

Let us see, if this $b^2 - 4ac > 0$; that simply means we have two real roots. If this $b^2 - 4ac = 0$; that means, we have only one repeated root. And if $b^2 - 4ac < 0$, then we are actually taking square root of a negative number which will go to the complex domain. So, it has no real roots.

So, let us summarize this method or the summary of this method into a table.

(Refer Slide Time: 06:03)



So, value of the discriminant suppose $b^2-4ac>0$ and the discriminant b^2-4ac is a perfect square; that means, I know the square root of it then we have two real rational roots. If $b^2-4ac>0$, but it is not a perfect square; then I have two real irrational roots.

We have already seen in week 1 that real number line is divided into rational numbers and irrational numbers. So, this is the splitting which will help. So, if $b^2-4ac>0$, but not a perfect square I will get irrational number. If b^2-4ac is a perfect square, I will get a rational number. If $b^2-4ac=0$, then I will get one real rational root. And if $b^2-4ac<0$, I do not have any root.

Let us demonstrate it through some graphs which we have seen which we have already seen. So, here is the example; where $b^2-4ac>0$. These are the two roots of this quadratic equation which are given here.

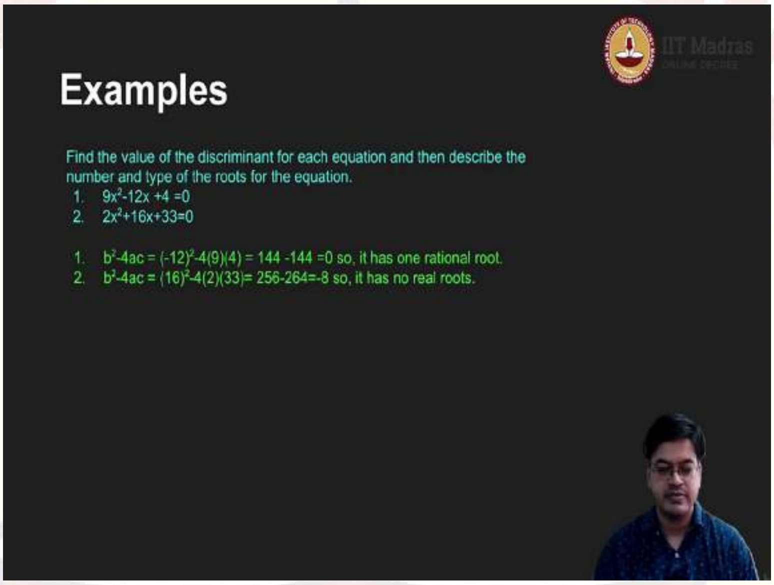
Let us say $b^2-4ac=0$, this is the root only one root it has and it is repeated. And if $b^2-4ac<0$, our example was x^2+1. So, in this case right it never touches the minimum; so you will get this particular expression ok. $b^2-4ac<0$, it never really touches the x axis. That is the verification that we have something like this ok.

So, let us go to the next slide which is actually yeah of course, these are the conditions that are required where a, b, c are rational numbers because, I am telling that this is this is will be a rational root. So, if it is not rational then what you need to do is; you suppress this, you do not need to say anything about this.

If they are rational numbers then whatever I am saying over here is true and whatever I am saying over here is true. If they are not rational numbers still you will have two real roots, but I cannot say whether they will be rational or irrational. And you will you may have one real root, but it can be irrational also. For example, $(x-\pi)^2$ or $b^2-4ac=0$ and then still you will get root as a π which is an irrational number.

So, that is so in order to distinguish between rational and irrational you need a condition that a, b, c are rational numbers. If you do not want to distinguish between rational and irrational numbers you do not need this condition. You can have a, b, c as real numbers; only condition that will prevail is $a \neq 0$.

(Refer Slide Time: 09:21)



So, let us go ahead and see some examples where I will use the discriminant formula or quadratic formula to distinguish between the roots. So, the question itself says; find the value of discriminant for each equation and then describe the number and type of roots for the equation.

So, let us take the first example is $9x^2-12x+4=0$. So, I want to evaluate b^2-4ac . So, it is b is -12, a is a is 9 and c is 4. So, b^2-4ac I want to evaluate for this. In a similar manner let us take the next equation which $2x^2+16x+33=0$; where b will be 16, a will be 2 and c will be 33.

Let us evaluate b^2-4ac for the first equation that is; $(-12)^2-4\times 9\times 4$. So, 9 4s are 36, 36 4s are 144 and 12^2 is also 144. So, 144 - 144, if you refer to the previous table it has only one repeated rational root.

You go to the second example $b^2 - 4ac$; b is 16, a is 2 and c is 33. So, if you look at it 16^2 is 256, 33×8 is 264 yes. So, I got 256 - 264; that means, I got -8. Therefore, the $b^2 - 4ac < 0$, and hence it will have no real root ok. This is the summary of using the discriminant method.

The discriminant method or the quadratic formula actually gives you a number of ways to handle the problem. So, in short what we have seen today is, we can solve an equation given that I know the values of a, b and c using the quadratic formula. So, let us summarize what are all the methods that we have studied in this particular example ok.

(Refer Slide Time: 11:45)

So, the summary of concepts is; let us say I have a method which is called graphing method. This is the method which we started with. When do we use this method? The graphing method actually unless your solutions are integers will not give you a good result.

But it is a best method to verify your results or verify the results that you have actually found algebraically. If there is any calculation mistake or something it will be revealed very easily. So, the graphing method is very helpful when you want to verify the result, but you can also use it to find roots of the equation occasionally.

Similarly, factoring method also suffers from the disadvantage that; it may not be helpful if the factors are not easily visible. For example, you may get the constant term to be equal to 26.2 in the quadratic equation. The constant term is 26.2 and then, you may have tough time in visualizing the factors.

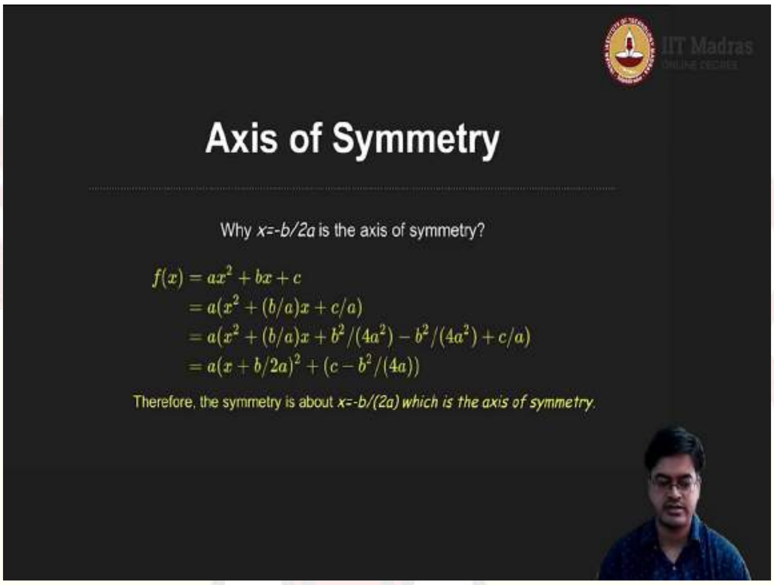
So, in such cases factoring method need not be used, but it is very helpful if the constant term is 0 or the factors are easy to find you can actually guess the factors if the nice numbers like 49, 24 all these nice numbers are there then you can very well go with factorization method.

The method of completing the squares all the time it works; it is very easy when b is even. Otherwise, you will have problems if the right-hand side in the method of completing the square is negative, that is you will go to complex domain and then you may have some problems which we are not dealing with in this course. So, for us it can be always used when b is even ok.

Now last method is quadratic formula, which is derived from which is derived from method of completing the square. And this gives you all the time this is this will give the answer all the time it is always helpful for. So, for our purposes when we are studying these methods these two methods; completing the square and quadratic formula will always give the answer irrelevant of whether the coefficients are rational numbers, irrational numbers, or some absurd numbers ok.

Now, let us go to one more concept which is called axis of symmetry. I have not given any derivation about axis of symmetry yet.

(Refer Slide Time: 14:33)



So, let us start with axis of symmetry. We already know while graphing the graphing the quadratic function it is very important to know the axis of symmetry. And we have boldly claimed at $x=\frac{-b}{2a}$ is the axis of symmetry. Now I will answer the question why $x=\frac{-b}{2a}$ is the axis of symmetry. This is an application of method of completing the square.

So, let us assume that I have been given a general quadratic function $f(x)=ax^2+bx+c$. $a\neq 0$. So, I will pull out a common and therefore, my expression will become $a\left(x^2+\frac{b}{a}x+\frac{c}{a}\right)$. Now, when I was completing the square I was throwing $\frac{c}{a}$ on the right hand side, but this time that provision is not there.

So, I will retain $\frac{c}{a}$ only thing is I will split the entire expression. So, when I split the entire expression, I will get a as it is and this expression as it is here $\frac{b}{2a}$ should come;

that means, $\frac{b^2}{4a^2}$ I will add and subtract $\frac{b^2}{4a^2}$ ok. And therefore, I will get this expression. Once I get this expression, I will recognize this term this entire term so these three terms together as $(x+\frac{b}{2a})^2+c-\frac{b^2}{4a}$ ok.

Now you look at this equation correctly. For example, this is a quadratic function written in a different form. What is this number? This is some constant ok. This is some constant and now you look at this number this is actually deciding the symmetry around x symmetry on x axis.

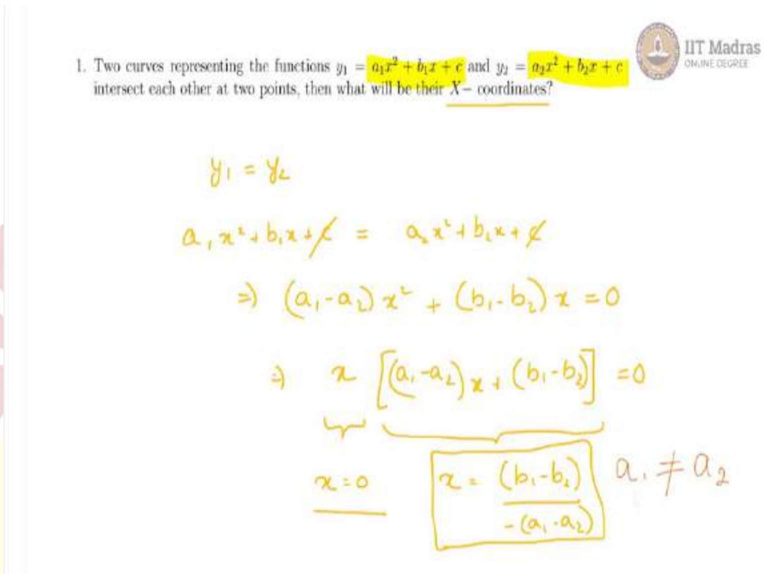
If you put $\frac{-b}{2a}$ as one vertical line; everything because $y=x^2$ it is symmetric about that y axis, everything will be symmetric about $x=\frac{-b}{2a}$. This expression defines the symmetry of the relation or the symmetry of the function because this is nothing, but just a constant on y axis.

Therefore, this x is equal x. So, basically you will write $(x+\frac{b}{2a})^2=0$. So, $x=\frac{-b}{2a}$ that vertical line is the axis of symmetry for this expression. So, the symmetry about $x=\frac{-b}{2a}$. Therefore, this is known as axis of symmetry that answers the quadratic equation axis of symmetry question. Why $x=\frac{-b}{2a}$ is the axis of symmetry?

This ends our topic on quadratic functions and quadratic equations.

Mathematics for Data Science 1
Week 05 - Tutorial 01

(Refer Slide Time: 0:14)



Hello Mathematics students, in this week's tutorial we will look at question related to quadratic functions. In our first question here we have two quadratic functions given to us and they intersect each other at two points and what will be there x coordinates. Clearly if they intersecting each other that means the x and y will be same. So, that is mean $y_1 = y_2$ and this is what we are trying to solve for.

So, $a_1x^2 + b_1x + c$ should be equal to $a_2x^2 + b_2x + c$ and the x is supposed to be same. So, anyway we can cancel off the c here. So this gives us $(a_1 - a_2)x^2 + (b_1 - b_2)x = 0$. This would imply this is us $x[(a_1 - a_2)x^1 + (b_1 - b_2)] = 0$. So, this corresponds to two different solutions. So, if we took this part to be 0 then $x = 0$ as one solution.

And the next will give us $x = \frac{b_1 - b_2}{-(a_1 - a_2)}$. So, this is a product of two terms and that product of two terms is 0 which means either of two terms has to be 0. So, one solution is x being 0 and the other one you get this as the solution. Now this is only a valid solution if a_1 and a_2 are not equal because a denominator cannot be 0. Therefore, $a_1 \neq a_2$ is a condition that needs to be satisfied.

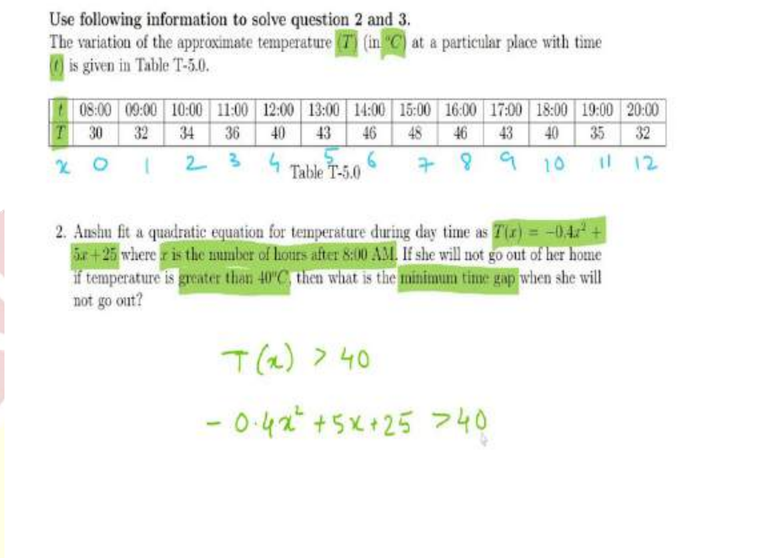
So, these are the two x coordinates one is 0 and the other is $\frac{b_1 - b_2}{a_2 - a_1}$.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 05 - Tutorial 02

(Refer Slide Time: 0:14)



Use following information to solve question 2 and 3.
The variation of the approximate temperature (T) (in °C) at a particular place with time (t) is given in Table T-5.0.

t	08:00	09:00	10:00	11:00	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00
T	30	32	34	36	40	43	46	48	46	43	40	35	32
x	0	1	2	3	4	5	6	7	8	9	10	11	12

Table T-5.0

2. Anshu fit a quadratic equation for temperature during day time as $T(x) = -0.4x^2 + 5x + 25$ where x is the number of hours after 8:00 AM. If she will not go out of her home if temperature is greater than 40°C, then what is the minimum time gap when she will not go out?

$T(x) > 40$

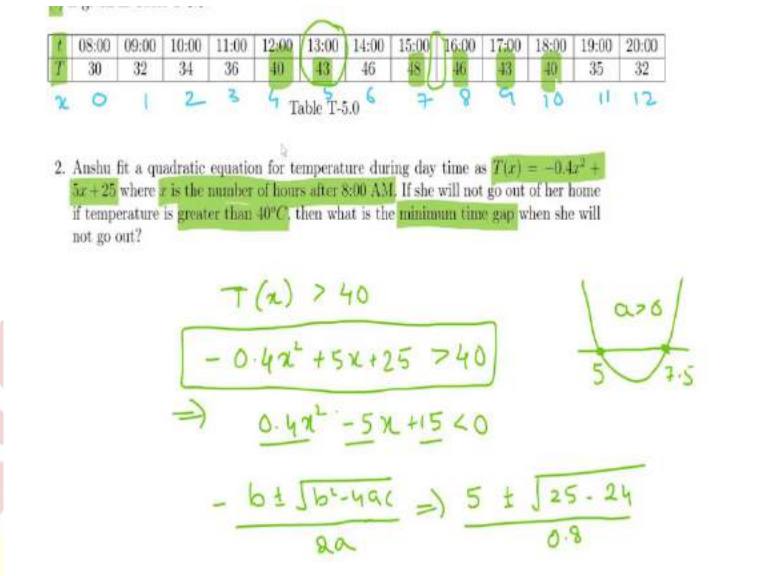
$-0.4x^2 + 5x + 25 > 40$

We are supposed to use this information this particular table to solve question 2 and 3. We will do a question 2 now. And this table will give us the variation of approximate temperature T. So, this is a temperature T in °C at a particular place with time small t, so this is the time. So, the time and the respective temperatures are given in this table.

And Anshu fit a quadratic equation for temperature during day time as $T(x) = -0.4x^2 + 5x + 25$. Where x is the number of hours after 8 am. So, x begins from 0 for $x = 0$. If we wrote additionally here this is 0, this is 1, this is 2, this is 3, this is 4, 5, 6, 7, 8, 9, 10, 11 and 12. So, we have x going from 0 to 12. If she will not, so if Anshu will not go out of her home the temperature is greater than 40 degrees.

So, greater than 40 degrees and Anshu will not go out of the home. Then what is the minimum time gap when she will not go out? Which means what is the time when the temperature is greater than 40. And this is on the basis of this particular quadratic equations. So, we are essentially trying to solve this as $T(x) > 40$. So, that means $-0.4x^2 + 5x + 25 > 40$.

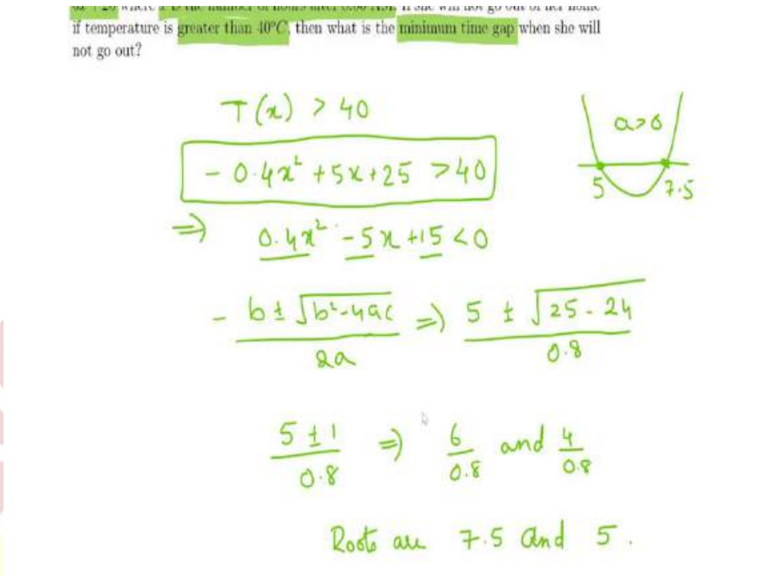
(Refer Slide Time: 2:23)



And that would indicate that $0.4x^2 - 5x + 15 < 0$. So, if we took all the LHS to the RHS this is what you will get and this is an upward facing parabola. So, the parabola will be like this and we are looking for the portion where you have the value, the y value to less than 0. So, that would be happen between the roots for this we find out the roots using the formula $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$.

And how did we know that this parabola is upward facing because a is greater than 0, a here is 0.4 b is -5 and c is 15. So, these roots will come out to be $5 \pm \sqrt{25 - 4 \times 0.4 \times 15}$ that is 16×0.4 that is 6 ×4 that is 24. So, divided by $2a$ is 0.8.

(Refer Slide Time: 3:46)



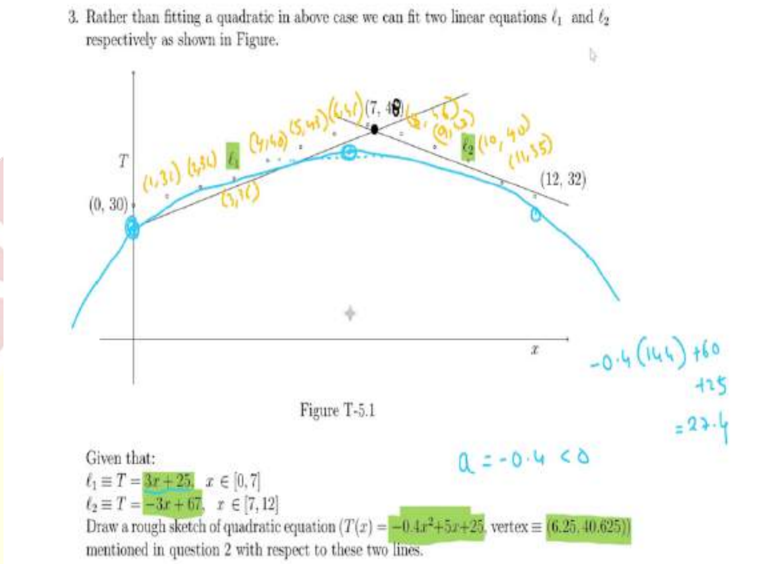
So, our roots are 5 plus or minus 1 divided 0.8 which is one is 6/ 0.8 and the other is 4/ 0.8. So, that gives us the roots as 7.5 and 5. So, these are the roots 5 and 7.5 and that means, this condition that is the temperature being greater than 40 is satisfied between 5 hours and 7.5 hours. That would be from here till somewhere in between here that is 15, 30. So, from 1 pm to 3:30 pm is the time suggested by the curve fit that Anshu has drawn but clearly this is wrong because it is already 43 here, and 48, here and 46, and 43, and 40, and 40 so it is a much larger duration where the temperature is greater than 40 degrees Celsius. So, this particular curve fit is pretty bad.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 05 - Tutorial 03

(Refer Slide Time: 0:14)



A third Question is related to the second question. So, in case you have not the second question please go back and see it. And here we are trying to say that instead of fitting a quadratic we can fit two linear equations l1 and l2. They have already provided us with the two equations which are this and this. So, $l_1 : y = 3x + 25$ and $l_2 : y = -3x + 67$ and the curves are already given.

Now the question is asking us to draw rough sketch of the quadratic equation that was fit and the vertex is also provided for us with respect to these two lines. So, I think it is useful if we can just mark out the points here. So, (0 , 30),(7 , 46) and (12 , 32) are already given. The remaining ones were, this is (1, 32), this is (2 , 34), this one is (3 , 36), this one is (44 , 40), this is (5 , 43), this is (6 , 46). So, this question has a problem here this should be (7, 48) that point.

And this is (8, 46) again and this is (9 , 43) and this point is (10 , 40) this is (11 , 35) and that may have (12 , 32). So, these are points and for us to do the rough sketch. The vertex is at 6.25 so the vertex should somewhere here and it is at 40.625. So, this is the horizontal (4, 40), then vertex is somewhere here. So, clearly our quadratic is below the points that we have been given. And this being the x^2 coefficient is -0.4 which is less than 0 so it is a down turned parabola.

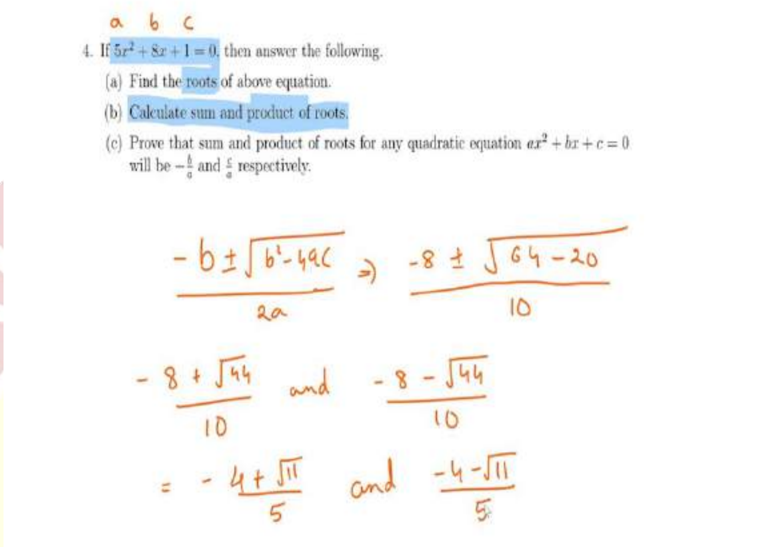
And let us look at the two points that we know for sure 0 and $x = 0$ this parabola is going to give us 25 the quadratic equation going to give us 25. Which is definitely below so somewhere here it appearing to be intersecting with this line. So let us look at, we have $3x + 25$ so at $x = 0$ the quadratic equation and the l1 line meet. And $x = 12$, we have 0.4 into $-0.4 \times (144) + 60 + 25$ which give us about 27.4, I think.

So, that is below this somewhere here so our quadratic is going to look like this. Something like this so it is quite inaccurate for the given data. So, this is a very bad curve fit.

IIT Madras
ONLINE DEGREE

Mathematics for Data Science 1
Week 05 - Tutorial 04

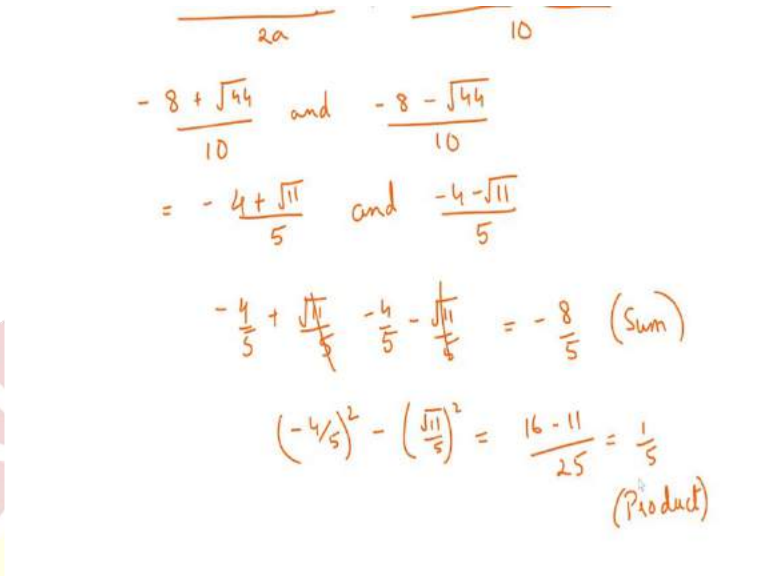
(Refer Slide Time: 0:14)



This is the pretty straight forward question, we are given a quadratic equation and we are asked to find the roots and also calculate the sum and product of roots. So, the roots we are going to get from the formula again which is $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$. So, that gives us, here in this case a is 5, b is 8 and c is 1. So, we have $\frac{-8 \pm \sqrt{64-4\times5}}{10}$.

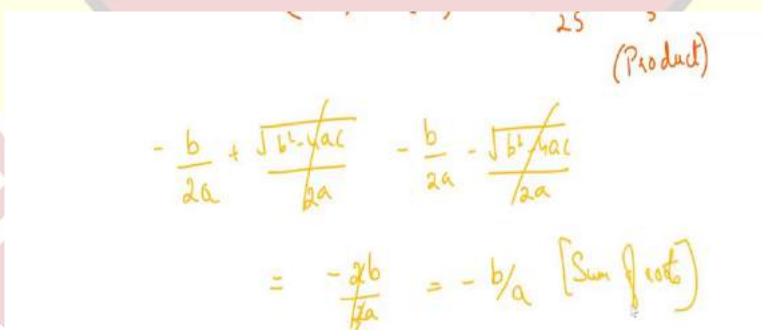
So, we have $\frac{-8+\sqrt{44}}{10}$ and $\frac{-8-\sqrt{44}}{10}$. And if we simplify it taking 2 common out, you will get $\frac{-4+\sqrt{11}}{5}$ because the 4 comes out of the square root and becomes 2. And $\frac{-4-\sqrt{11}}{5}$.

(Refer Slide Time: 1:44)



And the sum of these roots is if you just add them up you will get $\frac{-4}{5}+\frac{\sqrt{11}}{5}$. So, these get canceled so you get $\frac{-8}{5}$ is the sum. And in terms of product you basically doing $(a+b)\times(a-b)$ so you will get $(\frac{-4}{5})^2-(\frac{\sqrt{11}}{5})^2$. So, that gives us $\frac{16-11}{25}=\frac{1}{5}$. So, this is the product of the roots. $\frac{-8}{5}$ and $\frac{1}{5}$ are just sum and product of the roots respectively.

(Refer Slide Time: 2:43)



$$\left(\frac{-b}{2a}\right)^2 - \left(\frac{\sqrt{b^2-4ac}}{2a}\right)^2$$

$$= \frac{b^2}{4a^2} - \frac{(b^2-4ac)}{4a^2}$$

$$= \frac{\cancel{b^2}-\cancel{b^2}+\cancel{4}\cancel{a}c}{\cancel{4}a^{\cancel{2}}} = \frac{c}{a}$$

[Product]

The question is asking us to prove that the sum and product of roots for any quadratic equation will be this and this respectively. $\frac{-b}{a}$ and $\frac{c}{a}$ respectively. So, all we need to do for this is to sum $\frac{-b}{2a}+\frac{\sqrt{b^2-4ac}}{2a}$. This we are summing with $\frac{-b}{2a}-\frac{\sqrt{b^2-4ac}}{2a}$ a. So, clearly these two cancel off and you are left with $\frac{-2b}{2a}$, 2 and 2 cancel off and you have $\frac{-b}{a}$ is sum of roots.

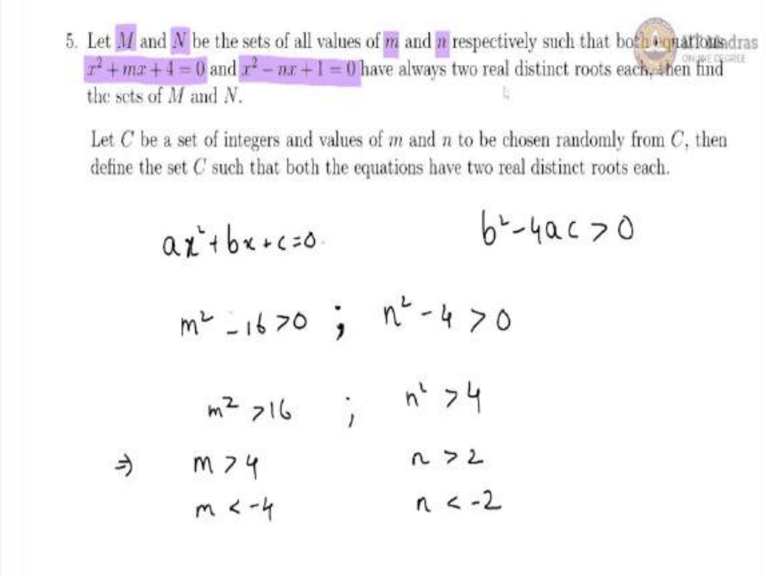
And when we do the product again it is in the $(a+b)(a-b)$ form so we will get$(\frac{-b}{2a})^2 - (\frac{\sqrt{b^2-4ac}}{2a})^2$ which gives us $\frac{b^2}{4a^2}-\frac{(b^2-4ac)}{4a^2}$. So, that gives us $\frac{b^2-(b^2-4ac)}{4a^2}$, b^2-b^2 cancels off then we have 4 4 going away $a \ and \ a$ going away so you were left with $\frac{c}{a}$. So, this is the product of roots for a quadratic equation.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 05 - Tutorial 05

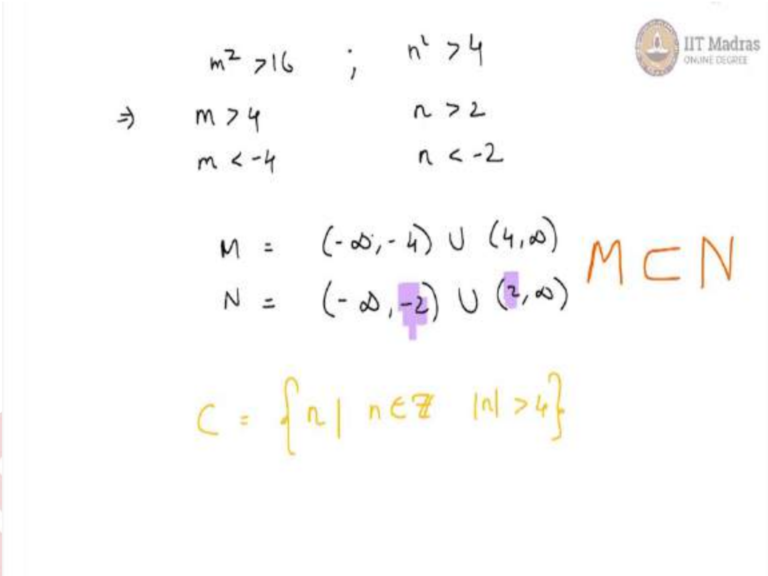
(Refer Slide Time: 0:14)



In this question we have 2 sets capital M and capital N which are sets of all values of small m and small n respectively such that these two equations have always two distinct real roots each, then find the sets M and N. Let us finish this part first. So, for a quadratic equation $ax^2 + bx + c = 0$ to have distinct real roots, the discriminant which is basically the value $b^2 - 4ac > 0$.

So, for this first equation that would be $m^2 - 16 > 0$ and simultaneously, we need for the second equation $n^2 - 4 > 0$. So, $m^2 > 16$ and $n^2 > 4$ and this would imply m is positive and greater than 4 or m is negative and lesser than -4. And here this would imply similarly n is positive and greater than 2 or n is negative and lesser than -2. So, these are all the possible values for which you will have two real distinct roots for these equations.

(Refer Slide Time: 2:02)



So, your set M would be the union of two intervals, one is a $(-\infty, -4) \cup (4, \infty)$. And set N is similarly $(-\infty, -2) \cup (2, \infty)$. Now the next part of the question, C is a set of integers and values of m and n are to be chosen randomly from C, then define the set C such that both equations have two distinct real roots each.

So, this is necessarily one single set we are taking and m and n should be chosen from that set. So, we clearly cannot have m being -2 or 2 or even -3 or 3. The set we are looking for is some sort of an intersection of capital M and capital N because both small m and small n should be drawn from this. And in this case that intersection will just be the set capital M because M is necessarily a subset of N.

However, C is also given out to be a set of integers, so it is not just the intersection of m and n, it is the set of integers which belong to the intersection and this case that intersection is only capital M where therefore, we have this set coming up as C.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

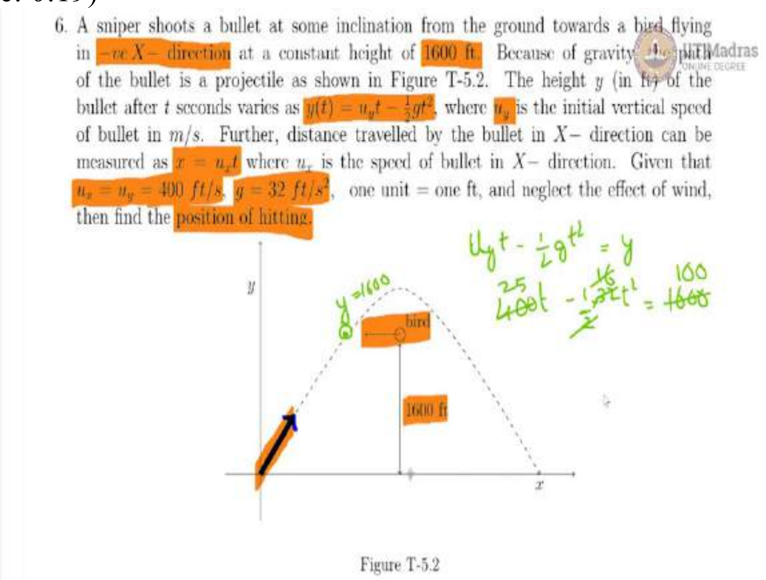
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 05 - Tutorial 06

(Refer Slide Time: 0:19)

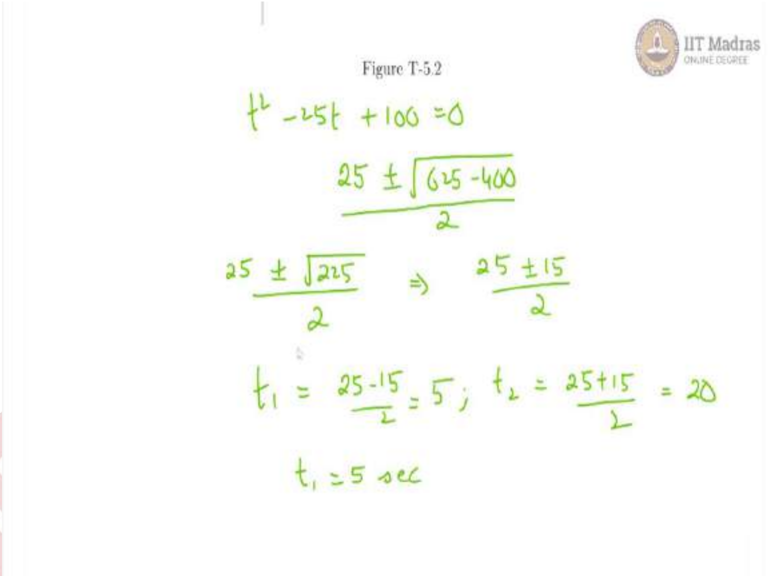


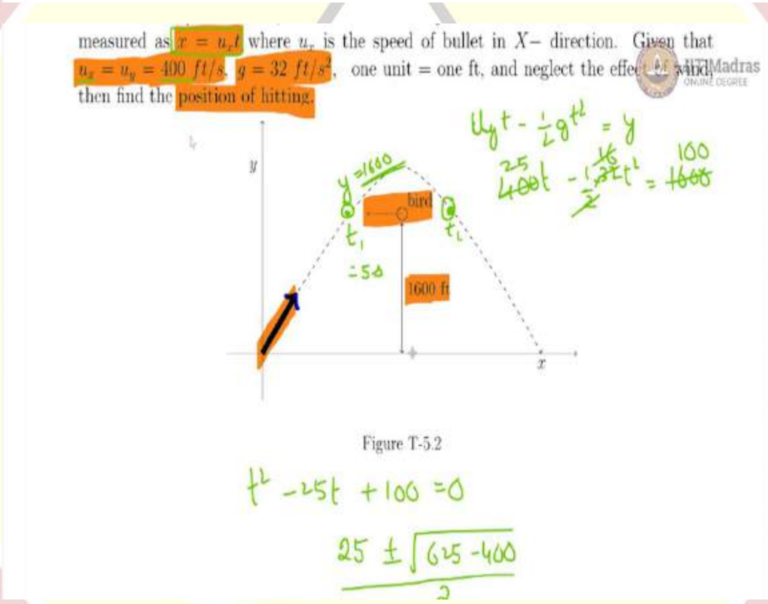
In our sixth question, there is a sniper who shoots a bullet at some inclination from the ground towards a bird flying in the $-$ x direction, some bird is flying in the $-$ x direction at a constant height of 1600 feet. Because of gravity, the path of the bullet is projected as shown in this diagram. So, this is the bullet, it is going in this particular parabolic path and this is the bird which is going in the $-$ x direction at a constant height of 1600 feet.

Now, they have given the height y of the bullet at t seconds as this function, this is a quadratic function $y = u_y t - \frac{1}{2}gt^2$, where u_i is the initial vertical speed and that is also given here, it is equal to 400 feet per second and the value of g is also given here, 32 $feet/s^2$. And then, the distance travelled by the bullet in x direction is given by $x = u_x t$ and $u_x = u_y$ = 400 feet per second, neglecting the effect of the wind and everything, find the position of hitting?

Where will the bullet hit the bird and that would be here where y = 1600 for the bullet. So, let us use the y equation and the y equation is $u_y t - \frac{1}{2}gt^2$. So, $u_y t - \frac{1}{2}gt^2 = y$, so we know y is supposed to be 1600 and u_i is 400, so we get $400t - \frac{1}{2}g, g\ is\ 32t^2$. So, 2 ones and 2 16s, now you can cancel off 16 here with this is equal to and this becomes 100 and this becomes 25.

(Refer Slide Time: 2:34)

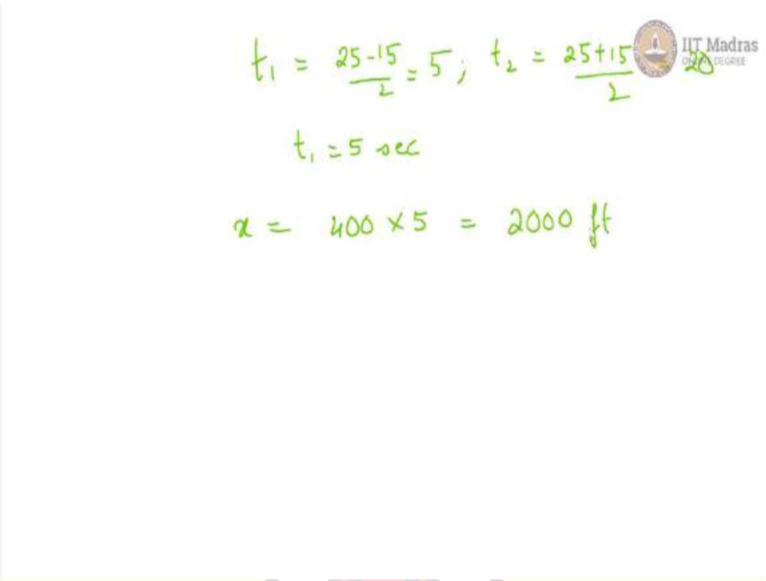




So, we get a quadratic equation which is $t^2 - 25t + 100 = 0$ and if we solve for the roots of this equation, we get the time when y is 1600 and we will get 2 times because y is 1600 twice on this path. So, we will get t_1 and t_2, we are looking for t_1 because that is where the bullet will hit the bird. So, your two roots are using the formula $\frac{-b \pm \sqrt{b^2-4ac}}{2a}$, here you will get it as $\frac{25 \pm \sqrt{625-400}}{2}$.

So, that gives us $\frac{25 \pm \sqrt{225}}{2}$ which then given us $\frac{25 \pm 15}{2}$. So, we have one solution, $t_1 = \frac{25-15}{2}$ and $t_2 = \frac{25+15}{2}$. So, this is equal to 5 and this is equal to 20. Clearly, t_1 =5 seconds is where our bullet will hit the bird. This is t_1 =5 seconds. And we already know the y coordinate of this place so, for finding the position what is left is to find the x coordinate which we will get from $x = u_x t$ where, u_x is already given to be 400.

(Refer Slide Time: 4:23)



So, $x = 400 \times t_1$ which is 5 and that is equal to 2000 feet.

(Refer Slide Time: 4:35)

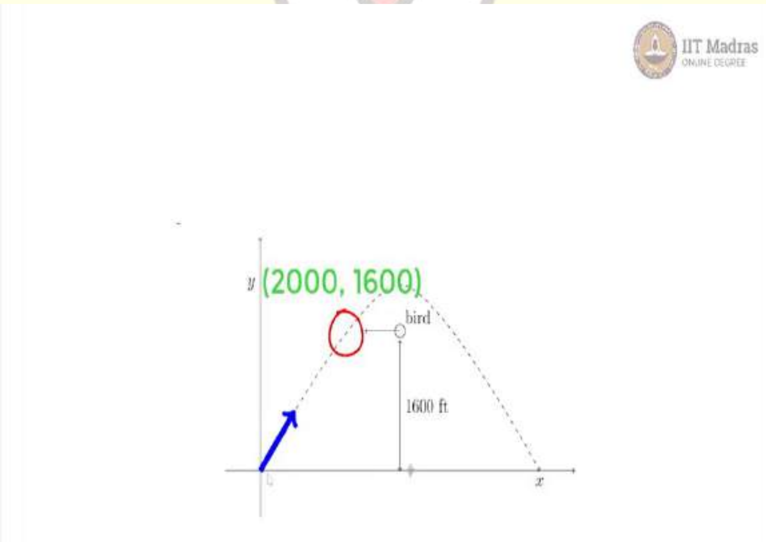


Figure T-5.2

Thus, the x coordinate for the point of hitting is 2000 and the y coordinate is 1600 feet. And this is the point where it hits.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 05 - Tutorial 07

(Refer Slide Time: 0:14)

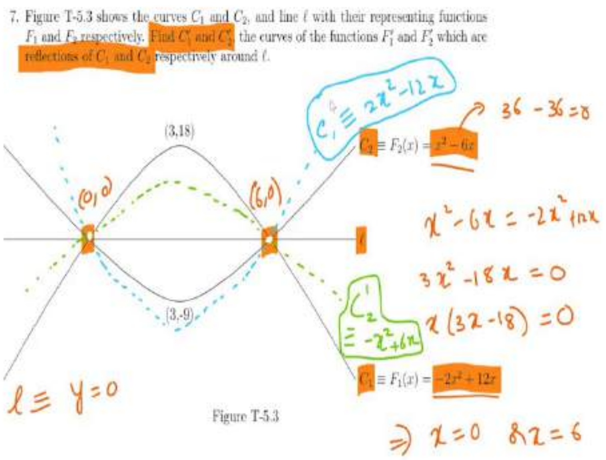


Figure T-5.3

In this question there are these two curves C_1 and C_2 which are both quadratic curves and there is this line l which is passing through these two intersection point. So, line l is passing through the intersection points of these two parabolas. They are asking find C_1' and C_2', the curves of the functions F_1' and F_2' which are reflections of C_1 and C_2 respectively around l which means for C_1 the reflection would be something like this, about l it would be something like this and for C_2 the reflection would be something like this and these are what we are trying to find out, C_1' and C_2'.

So, this should be C_2' and this would be C_1'. For all of these, we have to first find the line l and that we can find when we solve for the equality of these two functions. So, we are taking $x^2 - 6x = -2x^2 + 12x$. And that gives us $3x^2 - 18x = 0$ and that further gives us $x\,(3x - 18) = 0$ that indicates $x = 0$ or $x = 6$.

So, this point has coordinate $x = 0$ and this point has coordinate $x = 6$. We need to find the y coordinates for these points now. For that we substitute $x = 0$ and we get in this equation or this equation I wrote this and we get $y = 0$. So, this point is essentially the origin. Whereas, for this point we substitute $x = 6$ and we get $36 - 36$ which is 0. So, this point would then be (6, 0).

So, essentially this is a horizontal line which is $y = 0, l$ is $y = 0$. So, now we are just looking for reflections about the x axis because $y = 0$ as the x axis. And that would give us directly the negative coefficients of the same things. So, C_1' would then be $2x^2 - 12x$ whereas, C_2' would now be $-x^2 + 6x$. Thank you.

IIT Madras

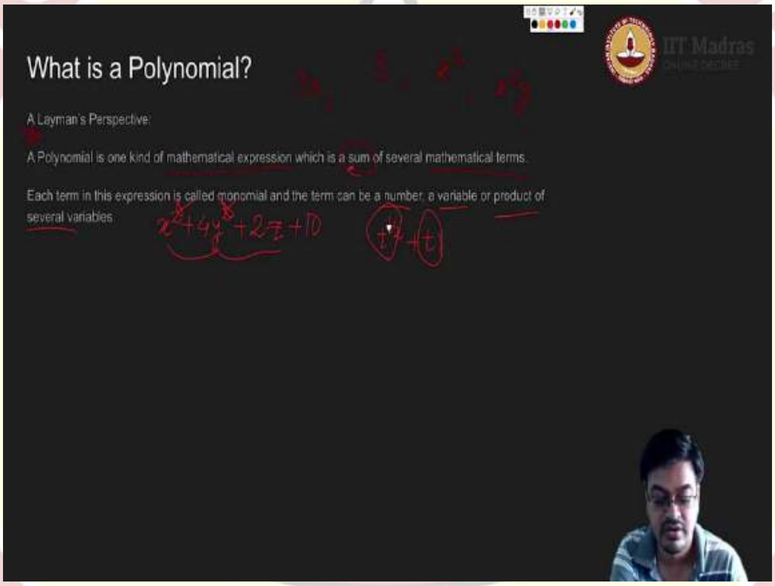
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 30
Polynomials

Let us introduce Polynomials. So, today we are going to see how the polynomials look like, how they behave. So, let us start with polynomials. Let us go ahead and see what expressions do we call as polynomial.

(Refer Slide Time: 00:29)



So, for that let us take the first point where we will take a Layman's perspective and, we will try to understand what a Layman will think of a polynomial. In order to do that let us first take a Layman's perspective and see what Layman will think. So, for a Layman, a polynomial it is some kind of mathematical expression which is a sum of several mathematical terms.

Then we asked Layman what do you mean by mathematical terms? The answer is each term in this expression, each term in this expression, that is mathematical term in this expression can be a number, a variable, or a product of several variables.

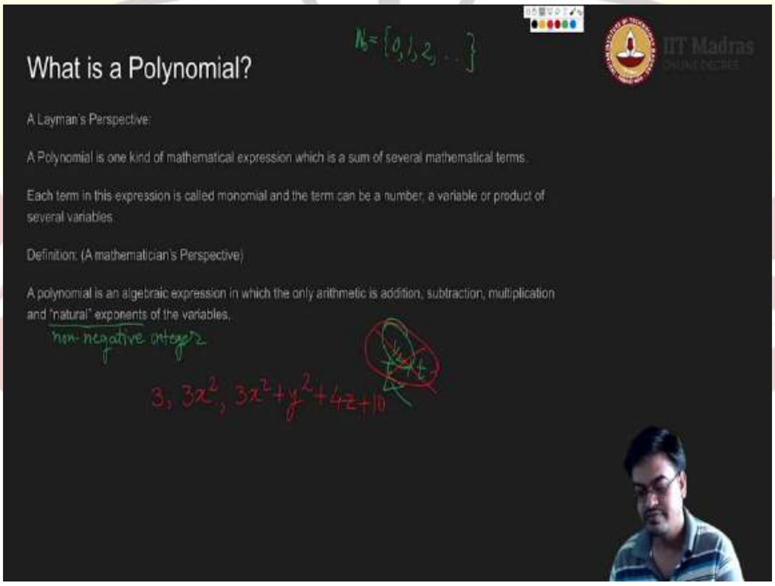
So, according to Layman each term this mathematical term can include a number, a variable, or product of several variables. These are the things that are allowed. So, basically then, let us take one example for this. And let us see if I have $3x$ this is a number and some variable. So, I have a constant 3, I have some number like x^2, I have some term like x^2y, all this contribute to something called polynomial ok.

Now, take a more significant number that is say $x^{2} + 4y^2 + 2z + 10$; will this contribute to be a polynomial? Yes, because it is sum of a number which is 10 here, a variable. There are many variables 1, 2, 3, there are three variables, and product of several variables; in particular here we have x^2 and here we have y^{2}. So, this also qualifies to be a polynomial.

Then according to this, suppose what I will write is here, let us say some expression of the form $t^{\frac{1}{2}} + t$; is this expression a polynomial? Layman will say yeah, it can be a polynomial because, $t^{\frac{1}{2}}$, if you square this number you will get this, correct. So, this is one variable this is another variable and therefore, we are actually having a polynomial.

So, then we went and asked mathematician, what is a mathematician's perspective of a polynomial?

(Refer Slide Time: 03:35)



So, that we will define as a definition. So, a mathematician said a polynomial is nothing, but an algebraic expression in which only arithmetic is addition, subtraction, multiplication and this is interesting, he mentions it as natural exponents of the variables. Natural, by natural I mean the way we defined a set of natural numbers in our first week I mean natural means 0, 1, 2 and so on.

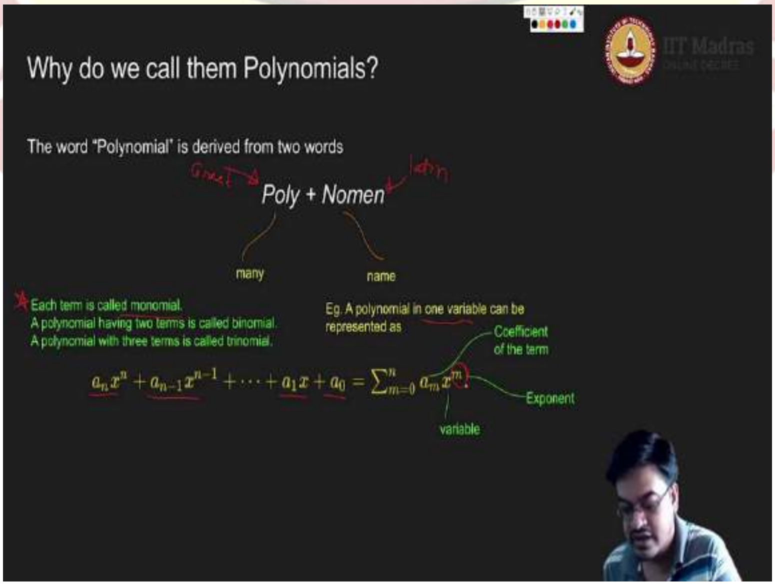
So, this is my set of natural numbers which include 0, if you want to emphasize it you can put it as N_0. Otherwise, you can call this set as set of whole numbers or set of non-negative integers. So, the definition can be twisted like this will have non-negative integer exponents. If you do not want any ambiguity, we can say that non-negative integer exponents of the variables.

So, then we if we go back to that earlier expression which is $t^{\frac{1}{2}} + t$, the all other expressions will qualify to be a variable, but this expression will not qualify to be a variable, why? Because this $t^{\frac{1}{2}}$ by definition is not a natural number, it is a rational number. We will come to it later, but this cannot qualify definitely so this cannot qualify as a polynomial if we go by this definition.

We have already seen many examples. Let us re-iterate them; one example was constant 3, another was $3x^2$, another one was $3x^2 + y^2 + 4z + 10$, all these are qualified to be polynomials. But this expression does not qualify to be a polynomial. So, this we will consider later as well and we will give our rational reason why it is not a polynomial.

So, we now we know what is polynomial. Now, it is time to see why is the name polynomial? Why do we call this as polynomial? That is what we will see now. So, let us go ahead and see something about the nomenclature. Why do we call them polynomials?

(Refer Slide Time: 06:17)



So, polynomials essentially is derived from two words one word is poly, second word is nomen. This is a Greek word this word and this has roots in Latin. The word poly essentially means many and the word nomen essentially means names or a terms. So, in our case it turns out to be terms. So, an expression having many terms is called polynomial ok.

Now, each term of the because it has many terms each term of the polynomial will be called as monomial, each term of a polynomial will be called as monomial. Then, if the polynomial has only two terms then you will call it as binomial. If the polynomial has only three terms then you will call it as trinomial.

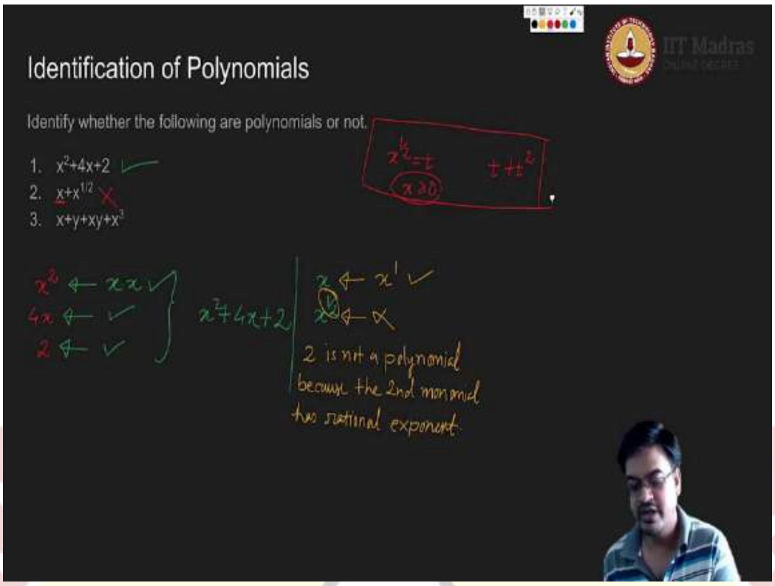
So far, if you can label them you can label them, but in general we will treat them as polynomials. And remember that we will include this also; a monomial is also a polynomial for us. We will not distinguish between monomial and polynomial. Of course, monomial enjoy some different set of properties, but we will keep them with polynomials.

So, let us take one example. For example, a polynomial in one variable can be represented as $a_n x^n$ this is the highest term, $a_{n-1}x^{n-1}$, $a_1 x^1$, a_0 right. I am assuming that this a_n's not equal to 0. Otherwise the if they are 0, then the polynomial may extend to infinity. I do not want that. So, I am assuming that these $a_n{}'$s are not 0.

Now, if you can rewrite this using the notation of summation in this manner and therefore, this a_m will have a specific name and it will be called as coefficient of the term. Because this is a polynomial in one variable, x is the variable we are interested in x is the variable, and this m is the exponent of the variable.

Now, remember in order that this term to be a polynomial, this m should always be a natural exponent; by natural I mean the one that is non-negative integer. If it is not nonnegative integer then I cannot classify this as a polynomial ok. Let us go to the next step and see some examples of polynomials and try to identify whether the given expressions are polynomials or not.

(Refer Slide Time: 09:35)



So, here is the question about identification of polynomials. Identify whether the followings are polynomials or not. The 1st one; $x^2 + 4x + 2$, what can you say about this? So, the first let us take term by term x^2, $4x$ and 2. So, the all these are monomials involved in this polynomial.

So, when I take x^2 it is nothing, but variable x multiplied by x. So, it is a product of two variables. So, this is ok. When I take $4x$ it is a number and a variable. So, this is also I do not have any and finally, this is just a number. So, together and the expression given it is sum of these that is; $x^2 + 4x + 2$, expression given it is sum of this. Therefore, this is a valid polynomial form.

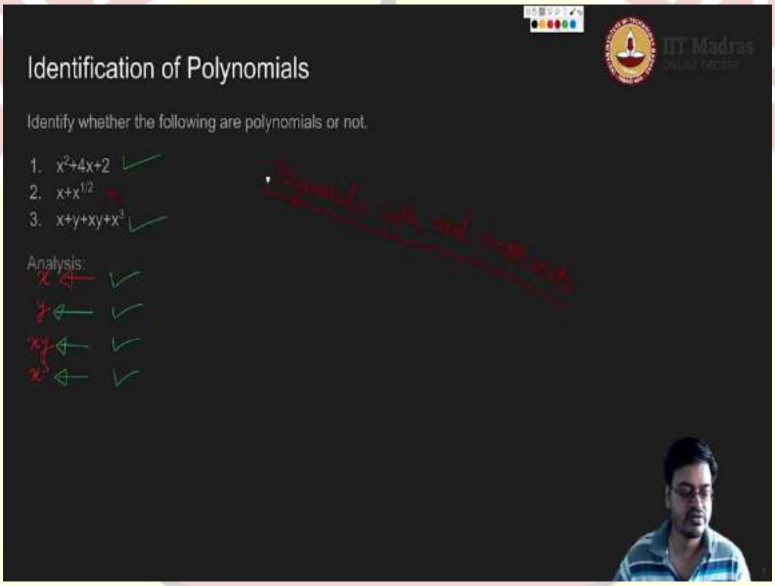
Let us go ahead ok. Again, the same expression has come $x + x^{\frac{1}{2}}$. So, now, you look at the terms that are involved x and $x^{\frac{1}{2}}$. Now, if you look at the terms involved x and $x^{\frac{1}{2}}$, then this it is simply a variable raised to 1, x^1, right. So, I do not have any problem, this is a valid term because there is no issue with this; $x^{\frac{1}{2}}$ this term is not a valid term because, it has rational exponent. So, this is not correct.

So, this second expression 2 is not a polynomial, why? We need to justify we need to write a reason because, the 2nd monomial has rational exponent. This is an interesting observation. So, this does not qualify to be, to be a polynomial. So, I can erase this. This is not a polynomial.

Now, some people may say that what is a big deal? I can put $x^{\frac{1}{2}}$, let's say t and you can rewrite this expression as $t + t^2$, but remember when I am putting x raised to half as t, I am putting an explicit assumption on this x that is; this x should be greater than or equal to 0. So, I cannot define this polynomial on the entire real line.

So, we will refrain from doing such assumptions and therefore, it would not be a polynomial. Let us go to the next example, this example ok. So, let me erase the previous ones so that I will have some space.

(Refer Slide Time: 13:15)

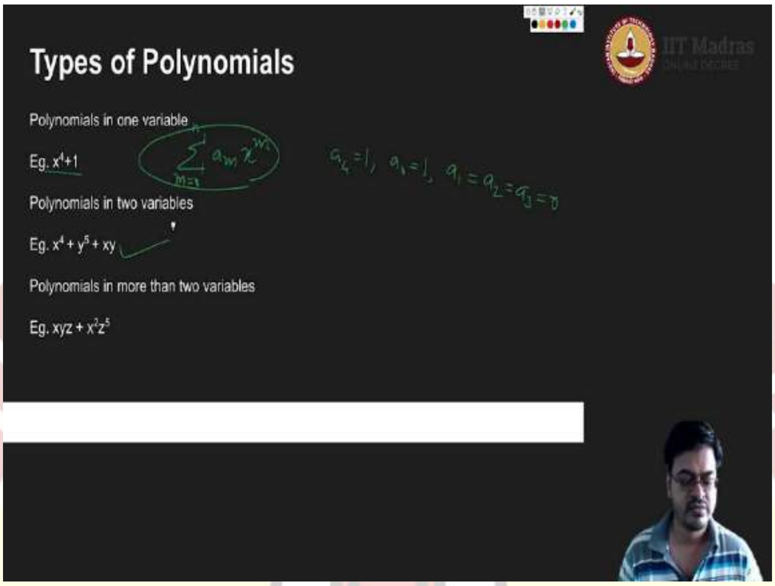


We will write all the terms as usual $x,\ y,\ xy,\ and\ x^3$. We will analyse the each monomial one by one, is this valid? Let me change the colour. Is this valid? Yes, it is valid because it is just a variable y? Yes, it is valid just a variable; product of several variables; product exponent natural exponent of single variable? Yes, it is valid.

So, according to me this and this qualify as a polynomial. And this do not qualify as a polynomial. So, we have identified what are the polynomials and how they look like. So, our identification part is complete. In particular we are dealing with polynomials having real coefficients because, all the numbers that I am giving you are real numbers.

So, just remember this fact we are only handling polynomials with real coefficients. If you go to the further branches of mathematics you may have polynomials with simply integer coefficients, you may have polynomials with complex coefficients, we are not dealing with them. So, this is how we will identify whether a polynomial whether a given expression is a polynomial or not.

(Refer Slide Time: 14:51)



Let us go ahead and try to describe what are the types of polynomials, that we can encounter. We have already seen them; we are just enlisting them for the sake of completeness. So, there can be polynomials in one variable which will typically look like $\sum_{m=0}^{n} a_m x^m$. So, that let me rewrite it.

So, that was our expression; summation over $\sum_{m=0}^{n} a_m x^m$. So, this particular thing falls into that category. What is what will be here in this particular case a_4 = 1, a_0 = 1 and all others like a_1, a_2 and a_3 , all of them are 0. So, this is how we will describe the polynomial. So, this is a polynomial in one variable.

You can encounter a polynomial in two variables. For example, we have already seen some examples this is a polynomial in two variables and you can have similar expression, but now, you will have $a_m b_m$ and a_{mn} or something of that sort, to indicate the powers of these exponents. So, we will not indulge into a mathematical representation of this, but you can have polynomials in two variables.

In a similar manner you can have polynomials in three variables or more than two variables. So, here is an example of a polynomial in more than two variables. And these are the types of polynomials that you may encounter with real coefficients. So, this summarizes the topic of representation of polynomials.

Now, let us go ahead and see some further properties of these polynomials.

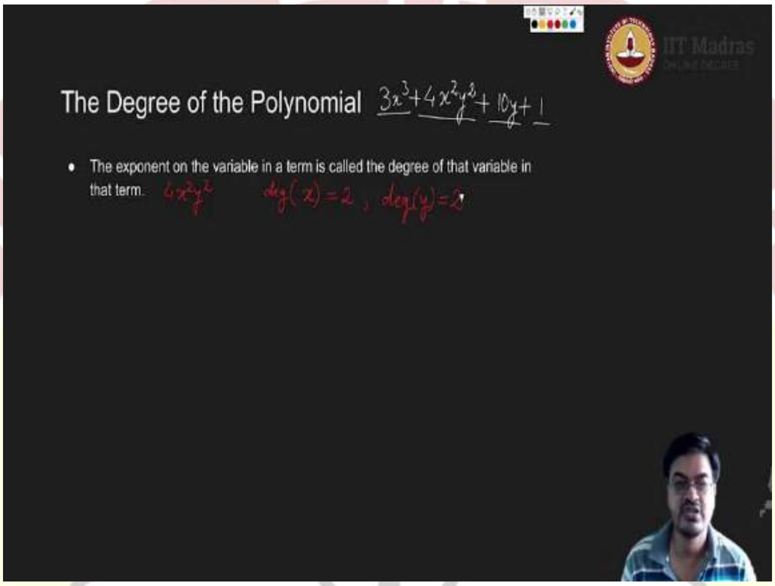
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 31
Degree of Polynomials

(Refer Slide Time: 00:15)



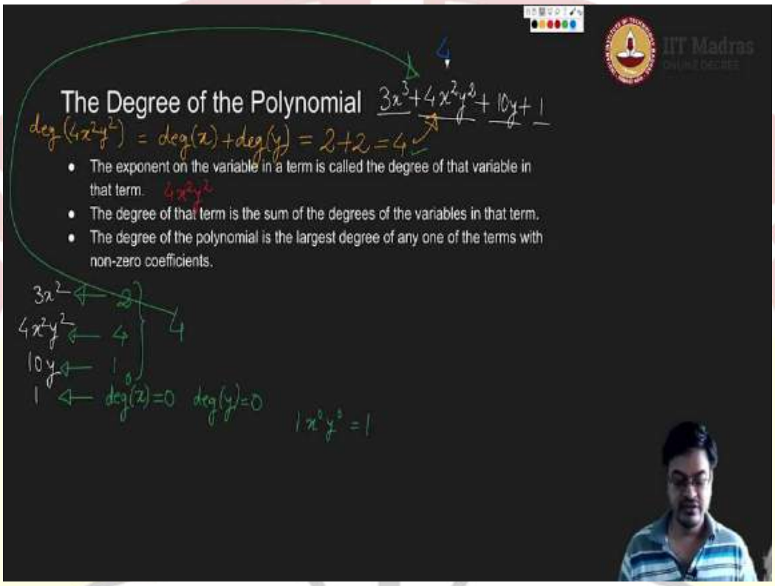
So, in particular, if I want to tell something about a Polynomial, an important property is a degree of the polynomial. So, what is the degree of the polynomial? For demonstration purposes, let me take one example. Let us say my example is $3x^3 + 4x^2y^2 + 10y + 1$, this is my example.

Then, I say this is the example. So, if I want to decide the degree of the polynomial, we have already seen each term itself is a polynomial. So, $3x^2$ is one; $4x^2y^2$ is one; $10y$ is one and this 1 is one. So, I want to identify the degree of each term as well.

In each term, there are many variables. For example, you would look at this term, if you look at this term then there are 2 variables. So, I want to have a complete understanding. So, in order to define the degree of a polynomial, I will start with defining the degree of the variable. So, the exponent on the variable in a term, the exponent on the variable in a term is called degree of that variable in that term. So, for demonstration purposes, let us take the expression $4x^2y^2$.

So, in this particular expression or in this particular monomial, how many variables are involved? One variable is x, second variable is y. So, what I am saying is now in this term, the degree of x; the degree of x let me abbreviate it as degree; degree of x is 2 and the degree of y, variable y is also 2, ok. So, this is how I will describe the degree of the variable.

(Refer Slide Time: 02:39)



Now, let us take this term as a term and say what is the degree of this term, right. So, let me erase this particular portion which is actually blocking our view ok. So, the degree of that term, this term, we have already seen the degree of x is something and degree of y is something.

Degree of x was 2 and degree of y was 2, the degree of the term is the sum of the degrees of those variables in the term. That means if I look at this expression which is $4x^2y^2$, then and I ask for the degree of this term, then it is essentially degree of x plus degree of y that is 2 + 2 which is equal to 4.

So, degree of this term the second term in this expression is 4, fine. Now, we will answer the question, what is the degree of a polynomial? So, the degree of the polynomial is the largest degree of among these all the terms of any one of the terms with nonzero coefficients or terms will exist only when there are nonzero coefficients.

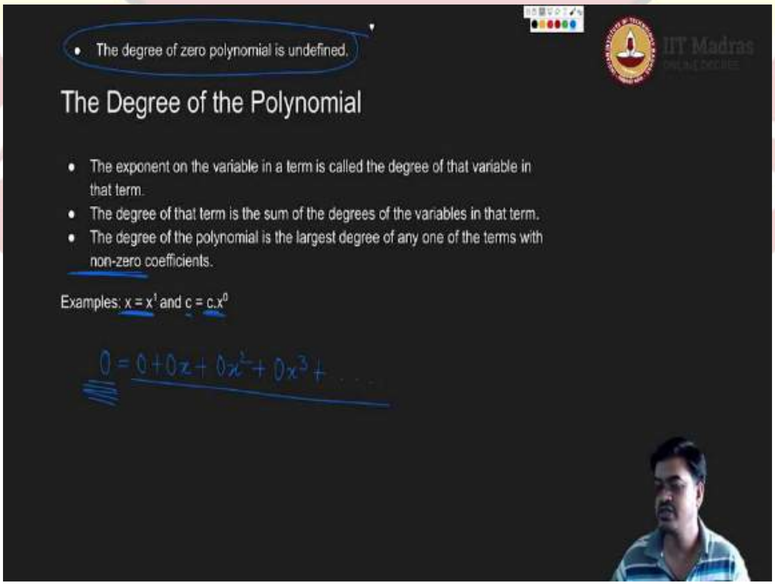
So, let us try to see how we can solve this problem. So, we will try to list all the degrees. So, if I take the first term that is $3x^2$, second term is $4x^2y^2$, then the next term is $10y$ and the last term is 1 which is the constant ok.

So, now, we will talk in terms of degrees. So, what is the degree of this particular term? It has only one variable x which has which is raised to the second power. So, the degree of this term is actually 2. What is the degree of this term? We have already seen here, the degree of this term is 4.

What is the degree of this term? The degree of this term is again y, means y^1. So, the exponent is 1, Interesting. Now, what is the degree of this term? Now, remember this is an expression in two variables; x and y. So, what then, I will ask a question what is the degree of x and what is the degree of y?

Now, you can also see that $1x^0y^0 = 1$. So, degree of x is naturally equal to 0 and degree of y is also equal to 0, right. Therefore, I can write the degree of this expression is 0 ok. Now, which one is the largest among these four? 0, 1, 2, 4? 4 is the largest. So, the degree of this particular polynomial is 4. So, this degree is actually 4, then write it here 4. So, this is a polynomial of degree 4 ok.

(Refer Slide Time: 06:31)



So, in this contest in while finding the degree of this particular polynomial, we have seen two things. What are those two things? If the coefficient is if the variable is x, then this is

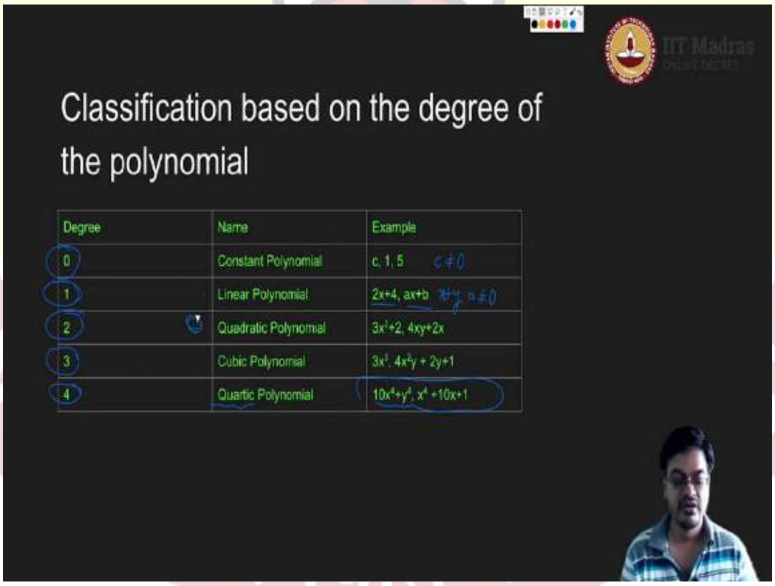
x^1, if it is a constant, then $c \times x^0$. In our case because the polynomial was having 2 variables, it is cx^0y^0.

Interesting question comes when we try to see polynomials at some other things. Let us say if I want to describe 0, when c is nonzero, it is ok, but if $c = 0$, then what? Then, you can see $0 = 0 + 0x + 0x^2 + 0x^3$ ... , the matter is complicated further and so on right.

It will continue. So, if the point given, then this number is 0, then we will call this as 0 polynomial and we cannot define the degree of this polynomial because for a degree, we need a nonzero coefficient, just remember this in mind.

Therefore, the degree of 0 polynomial is always undefined this is an interesting fact which will be used when we use the division algorithm. The degree of 0 polynomial is undefined. So, you can use it in a more interesting manner that is what I can say. So, the degree; this is what? Degree of 0 polynomial is undefined.

(Refer Slide Time: 08:19)



So, we have understood what is the degree of the polynomial. So, in particular, based on the degrees, now we have introduced one classification. So, based on the degrees, how the polynomials can be classified. So, if the polynomial has degree 0, then it is constant and this constant can never be equal to 0.

This is an important assumption. Then, if the polynomial is of degree 1, linear polynomial, then you will have a polynomial in this form. When I write this, then I should write $a \neq 0$, if I have a quadratic polynomial, if the polynomial ok.

So, here these are the polynomials in one variable, then I am considering a linear polynomials. If I am considering a polynomial of the form $x + y$, this is still a linear polynomial; but it is a polynomial in 2 variables. So, you can also encounter such polynomials in linear, but the crucial fact is degree is 1.

Second one is a quadratic polynomial which is of this form and here you can have polynomial in two variables, three variables or whatever way you want. Then, you will get a cubic polynomial which will have all terms containing degree 3, the highest term, highest monomial will have degree 3.

So, this is these are the examples of degree 3 polynomials. Similarly, degree 4 polynomials are called quartic polynomials and they will be given in this form and similarly, degree 5 polynomials are called quintic or quantic polynomials which will be represented with degree 5 polynomials, right.

So, and in general, you have a general term which is called polynomial. So, to be if you want to be specific, you can use this classification and say it is a quadratic polynomial, then you are giving more information about it.

This is what today's lecture meant to be. So, we have introduced the topic of polynomials.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS • सिद्धिर्भवति कर्मजा •

IIT Madras

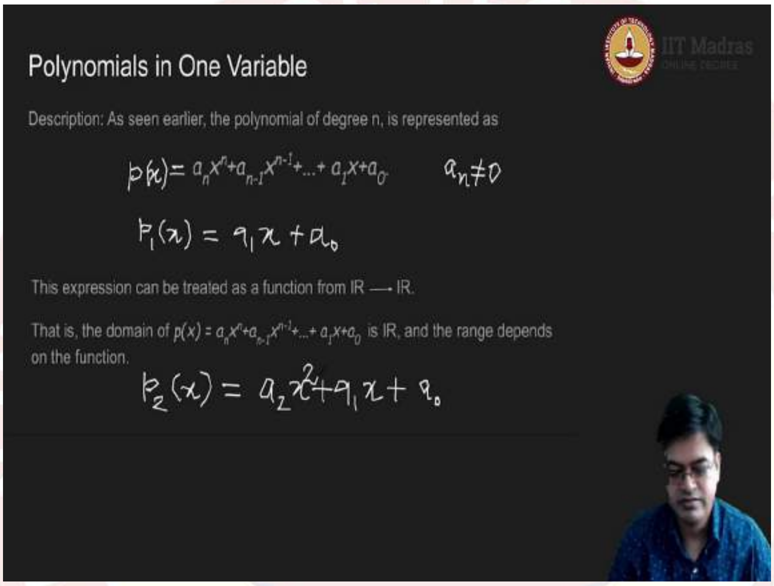
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 32
Algebra of polynomials: Addition & Subtraction

In this video, we will start with polynomials and we will try to do some Algebra with Polynomials. Or in other words you can say we will try to understand some operations on polynomials like Addition and Subtraction. So, let us move on.

(Refer Slide Time: 00:34)



In order to simplify our calculations, we will only focus on polynomials in one variable; whereas all the operations that we are discussing can be done on polynomials with multiple variables. In order to pinpoint the thing, we will recollect how polynomials in one variable look like.

So, a polynomial of degree n in one variable can be represented in this form; $a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$. So, you can actually correlate this with, let us say this is a monomial with 0 degree, this is a monomial with 1 degree and so on, if you go on this way, this is the monomial with n th degree. In order that this polynomial to qualify as a polynomial with n th degree, we need something, we need one condition; that condition is actually I need this to be a polynomial of the term of degree n to be non-zero.

So, that forces me to write $a_n \neq 0$, this is a condition that require that is required for writing a polynomial of nth degree. Remember here the argument is only one that is the variable is only one x. So, I can also assign this as something called $p(x)$, and now you can as well treat this $p(x)$ as a function of one variable which is interesting.

So, if you assign this as a function of one variable, the next question is; how is this function, what is the domain and co-domain and range of this function? So, the function runs from $\mathbb{R}$ to $\mathbb{R}$. So, it is a function from real line to real line; whereas the range typically depends on function.

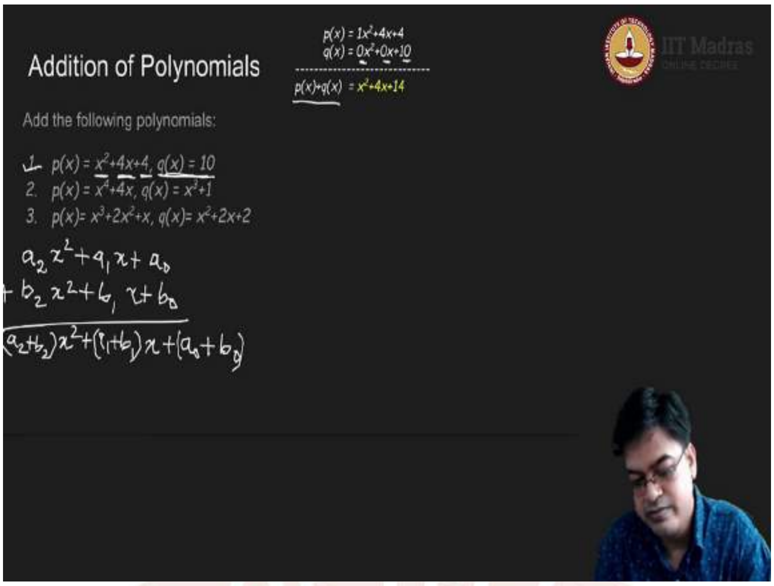
For example, if I take a function like let us say $p_1(x)$ is one function, which is $a_1 x + a_0$ and if I take this function, then it is a linear function; we have already seen this function, this is an equation of a real line, equation of a line. And if a 1 is not equal to 0, then this function actually represents a real line. If $a_1 \neq 0$, it also represents a real line, but it is some constant.

So, it is a horizontally real line. So, now, the range of this function for $a_1 \neq 0$ is entire real line. Whereas if you look at some other function, let us say $p_2(x) = a_2 x^2 + a_1 x + a_0$. Now, this particular function represents a parabola, which we have seen in our topic on quadratic functions.

And you know depending on the sign of a_2 , the parabola can open upward or downward; if it opens upwards, the range is the minimum value and any point beyond that; if it opens downwards, the range is the maximum value and any point below it. So, depending on the choice of the function, the ranges may differ. We will deal with polynomials as function when we will study the graphing of polynomials.

Right now, we are interested in algebraic properties of this polynomial. So, we will focus ourselves on the algebra of the polynomials that is addition, subtraction, multiplication, division.

(Refer Slide Time: 04:53)



So, let us move ahead and try to understand addition of polynomials. We have done this in some sense, for example, currently the polynomial that we have written that $a_n x^n + a_{n-1} x^{n-1} + \dots$, is also addition of some kind; but it is addition of monomials. So, let us try to see, if I have been given to two polynomials, how will I add them?

To help us in understanding and developing general theory for addition of polynomials, we will consider these three examples. Remember, the first example is actually one polynomial added with another monomial; second one both are two polynomials, but there are no clashing terms, like they the exponents are different for both the polynomials, you can check and the third one has few clashing terms.

So, we will demonstrate the addition of polynomials through these three things and we will formalize this into a theory. So, let us start with the first expression, $p(x) = x^2 + 4x + 4$. So, we are starting with this, this particular expression. So, $p(x)$ is x^2. So, if I am writing x^2, then it essentially means I am multiplying this with 1, the coefficient is 1; if I am starting with $4x$, then I do not have to do anything and this is 4.

So, in this case in our standard form for a quadratic polynomial, what is a standard form for a quadratic polynomial? $a_2 x^2 + a_1 x + a_0$ this is our standard form. So, in this particular thing, you can identify $a_2 = 1, a_1 = 4$ and $a_0 = 4$.

In a similar manner, I will look at this particular expression which is $q(x)$. Now, you notice the fact that $q(x)$ is just a constant polynomial, $q(x)$ do not have any terms which are related to square or related to a linear term. And, I want to add this polynomial to a given expression.

So, how will I add? So, let us bring in the terms related to square term and related to linear term; if I bring in those terms, the associated coefficients will be 0 right, the associated coefficients will be 0. So, I can write this term as $0x^2+0x+10$.

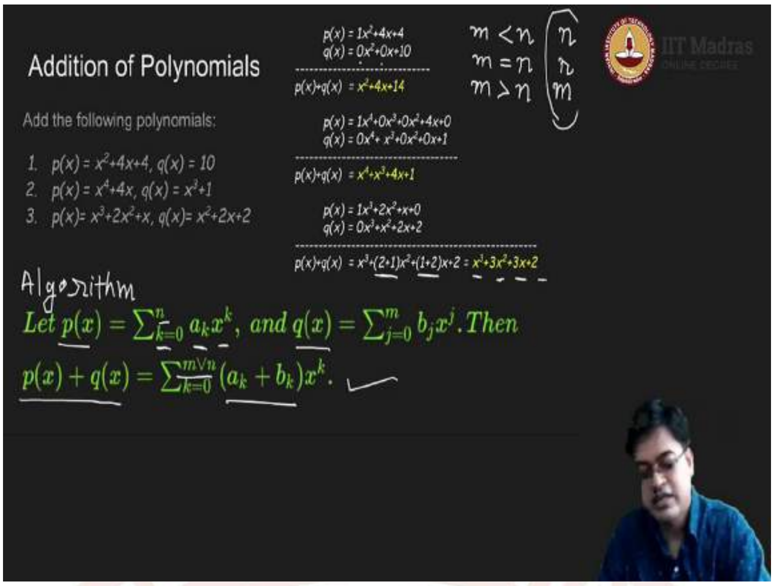
Now, because of this, let us write it in a generalized setting; $b_2x^2+b_1x+b_0$, right. So, now, I am trying to add these two polynomials. So, what is our recipe? We will consider the terms with like powers that is like exponents, ok. So, let me try to add the things.

So, if I consider this, this particular expression that is given here. So, I have $1x^2$, 1 minute. So, let me bring in my mouse pointer here. So, I have $1x^2+0x^2$. So, 1 + 0, I will get again singleton x^2; then 4 + 0 that will give me $4x$, 4 +10 will give me 14. So, essentially I can see that this expression should have a formulation which is of the form $x^2+4x+14$.

So, if I now try to do it in a more general settings, then how will I compare with this general setting. Let us see it here. So, I want to add these two. So, just add. So, in a similar manner, if I add these two; what I am getting is $(a_2+b_2)x^2+(a_1+b_1)x+a_0+b_0$, just to remember this format. So, what I am writing here is essentially.

Another point that to note with this example is; the first one was a polynomial of degree 2, the second expression $q(x)$ was a polynomial of degree 0. Now, the resultant expression that is $p(x)+q(x)$, what is the degree of this polynomial? It is a polynomial of degree 2. So, it is the maximum of degree of the first polynomial and degree of the second polynomial. So, we have roughly understood the settings that we need maximum of 1 and 2, let me write the findings in a different way.

(Refer Slide Time: 10:49)



So, if I have a polynomial of degree m and let us say if I have a polynomial of degree m and degree n, these are the two polynomials; and if they satisfy a relation that m is less than n, then the resultant polynomial will have a degree n. If I have a polynomial where degree m is equal to n ok, then the resultant polynomial will still have a degree n.

And if I have a case where m is greater than n, then the resultant what will be the; so if we switch this $q(x)$ to $p(x)$ and $p(x)$ to $q(x)$, this case will happen and the resultant polynomial will have degree m. So, just remember this in mind; that means it is always maximum of m and n, if the resultant polynomial is having polynomials underlying polynomials with different degrees. So, with this understanding, let us attack the second problem.

So, the second problem has $p(x)$ which is x which is a polynomial of degree 4. So, I have written all other terms which were not there in the polynomial by multiplying with 0. In a similar manner, I have written the second polynomial $q(x)$ which is a polynomial of degree 3; but we want the maximum degree to survive right or is essentially in this expression the maximum degree will survive, therefore I have this kind of expression.

So, the resultant polynomial we know from our discussion will be a polynomial of degree 4, and therefore I need to consider all the terms that correspond to each of the degrees. So, what is that term corresponding to degree 4? In the first expression that is

$p(x)$ is 1, the coefficient is 1 and the term corresponding to degree 4 in the second polynomial that is $q(x)$ is degree 0.

So, I will get 1 + 0, which is 1; $1x^4$. In a similar manner you can see, for x^3 it is $1x^3$; x^2 there is no survivor both are 0, so $0x^2, 4+0x$ and 0 + 1 times 1. So, it is 1. So, the resultant that you are interested in is x^4+x^3+4x+1 . Again I will reiterate, this time it will be; if you consider a generalized polynomial, it will be $a_4x^4+a_3x^3+a_2x^2+a_1x+a_0$.

And in a similar manner $q(x)$ will $b_4x^4+b_3x^3+b_2x^2+b_1x+b_0$. And if you sum over them, what we have written in yellow is essentially sum of $a_4+b_4=1, a_3+b_3=1, a_1+b_1=4, a_0+b_0=1$; a_2 and b_2 will sum to 0, because it does not have any non-zero coefficient. So, this is how we will handle the third, this is how we have handled the second problem.

And the term containing the highest degree survive, therefore the degree of polynomial is 4, ok. So, let us go back to the third problem, let us come ahead and solve the third problem; $p(x)$ and $q(x)$, again similar setting highest degree is degree 3. So, the term corresponding to degree 3 will survive. So, the polynomial with lower degree, I will bring it to degree 3. So, essentially, I will multiply with coefficient 0 for a degree 3 term.

Again by same logic, I will add the two terms and therefore, I will get the corresponding answer. So, there were cross terms, like the term corresponding to x^2 was crossing; for example, both polynomials had terms corresponding to x^2 .

So, you can see the difference here, we are just adding 2 + 1, 1 + 2. So, all these things are happening and together we are writing the result x^3+3x^2+3x+2 . So, from this we can derive, if you are clear with these three examples then we can derive a general formula; otherwise pause and look at each of the terms, you will be able to understand the general formula in a bit better manner if you pause and review these additions.

So, let us come to the general formula, you must have paused and understood the additions. So, if I have a polynomials of the form $\sum_{k=0}^{n} a_k x^k$. And if you assume that $a_n \neq 0$, then this is a polynomial of degree n.

In a similar manner you have taken a second polynomial $q(x)=\sum_{j=0}^{m} b_j x^j$. Now it does not matter whether m is greater than n or m is less than n, this particular thing will give you the answer, ok. What is the resultant? So, if I want to add these two polynomial functions $p(x)$ and $q(x)$, then $p(x)+q(x)$ will essentially show this kind of representation.

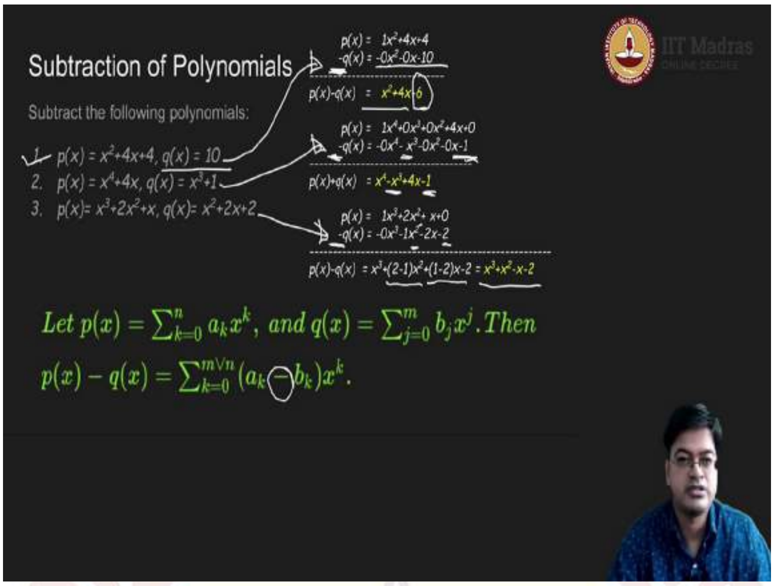
So, what we are actually saying is, choose which one is the maximum m or n ok? Whichever is the maximum? Take that maximum. So, it is maximum of m and n ; sum will run from k is equal to 0 to $n.$ Let us say for argument purposes this is the highest degree that is n is the highest degree, then match the degree of the highest degree for other coefficients, for example, for j equal to $m+1$ to n put all $b_j's$ to be equal to 0.

If you do so, this is what we have done here. So, in this case the first degree was 2 and the second degree of was 0. So, in this case we matched the degree and substituted all the coefficients here to be equal to 0. So, you do a similar thing over here in general and then just add the coefficients $(a_k+b_k)x^k$, and this should give you the final answer. So, this is in fact an algorithm for adding the polynomials, so this is algorithm for adding the polynomials.

What is the, what are the steps in the algorithm? First identify degrees of both polynomials, choose the polynomial with highest degree that will be the degree of the resultant polynomial. Take the polynomial of least degree that is step 2, take the polynomial of least degree, add all the coefficients which are of the degree higher than the polynomial and multiply them with coefficients 0.

Once you do that you are ready to do the addition, add the two using this formulation. So, this is how you can program, you can actually program into a computer for addition of polynomials. Now, let us try to understand this with subtraction. What is the difference between subtraction and addition? Both are essentially same, but in subtraction you are multiplying the second polynomial by -1.

(Refer Slide Time: 19:19)



We have already seen how subtraction will happen. So, here is a quick overview of these examples. So, this is, these are the polynomials, same polynomials now we are subtracting. And what we are doing by subtracting? I mean we have multiplied with -1, just look at here all these terms.

The procedure is exactly the same, it is just that first we have to multiply by -1 and put the polynomial appropriately. So, let us start with first example, but it will be a quick run, because $p(x)$ is this. So, there is no change, but I want to subtract $p(x)$ from, I want to subtract $q(x)$ from $p(x)$. So, this polynomial will be multiplied with -1 that is what is done here, so $-q(x)$.

So, correspondingly all coefficients are negated, just look at these terms; all coefficients are negated and therefore. Because there were no cross terms, so you will not find any difference in the first two terms; but the significant difference is there in the third term which is actually -6. In a similar manner, take the second question and you are multiplying $q(x)$ with -1.

So, $-0x^4-x^3-0x^2-0x-1$, right. Again because there were no cross terms, there were no additions. So, this also will have a minimal effect where the second term, these terms will be with the negative sign, right. So, this -1 was there. So, this will have a negative sign.

So, there is a minimal impact. The third example that we have taken will face a major impact; for example, $p(x)-q(x)$. Now, you take this $-q(x)$ here, if you take that $-q(x)$ here, then $p(x)-q(x)$; the first term will be as it is x^3 , because it is coming from this point here there was a clash of x^2 . So, 2 in when we added, it was 2 + 1. So, everything became 3; here it is $(2-1)x^2$. In a similar manner $(1-2)x$ and then the final term was -2.

So, essentially you got the expression in this form, which is x^3+x^2-x-2 ; but the key principles remain the same, except for multiplying with -1. Multiplying with -1, because that was a polynomial of degree 0 will not change the degree of the polynomial. So, once it is not changing the degree of the polynomial, all the rules which were possible for addition remain intact.

For example, you have to choose the degree which is maximum of the polynomial; there is no change in the degree except for the multiplication of a minus sign. So, that multiplication of minus sign is absorbed here. So, now, in the new rule it will be $p(x)-q(x)$ will be k is equal to 0 to maximum of m and n there is a remained intact; and earlier when we were adding it was $a(k)+b(k),$ now it is $a(k)-b(k),$ ok.

So, I hope you have understood addition and subtraction of the polynomials, both are essentially same and the resultant what we are getting is again a polynomial. In the next video, we will take a closer look at multiplication of polynomials.

Thank you.

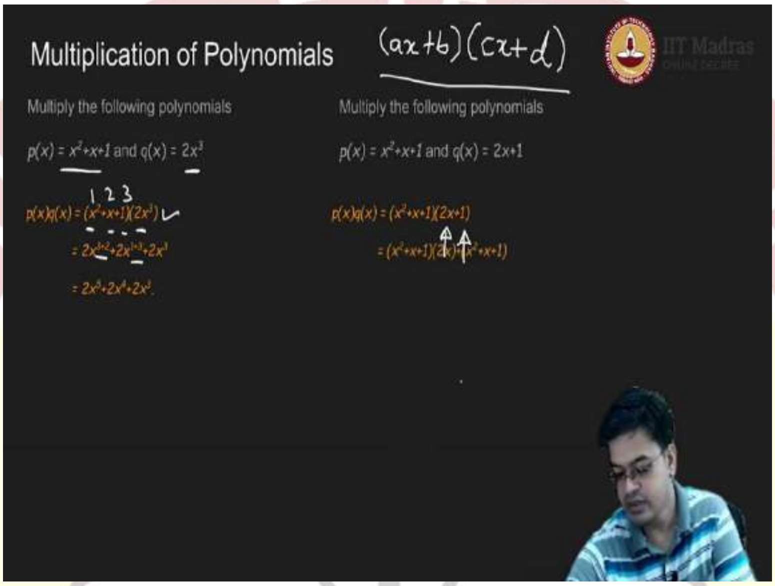
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture - 33
Algebra of polynomials: Multiplication

(Refer Slide Time: 00:14)



In this video, we will learn how to multiply two polynomials. Let us start with basics of multiplication of polynomials. We already know how to multiply two binomials. For example, if you have been given two binomials of the form $ax + b + cx + d$, then you know how to multiply these two binomials that is we will use the foil method. However, in this context, we want to generalize the settings for multiplication of polynomials of arbitrary degree.

So, let us see, let us start with some simple monomials with through examples. So, here is a polynomial given to you $p(x) = x^2 + x + 1$ and $q\ x = 2x^3$. The question is do I know how to multiply these two polynomials? Remember this one is called monomial, it has only one term. So, a standard rule of multiplication will mean we have seen this in our quadratic functions that I will consider the product in this manner.

Once I consider the product in this manner, what we will do is we will try to multiply each term of this $2x^3$ with each term of this polynomial. So, there are three terms. And for each term this $2x^3$will be multiplied. So, if I do that the law of exponents will apply.

For example, $x^3 \times x^2$ will mean x^{2+3}. So, once I apply we apply the law of exponents and add the exponents, obviously, 2 was a constant coefficient of x^3 which will be multiplied throughout the expression. And therefore, the resultant is this which we can simplify as $2\,x^5 + 2\,x^4 + 2\,x^3$. This is how we will multiply a monomial.

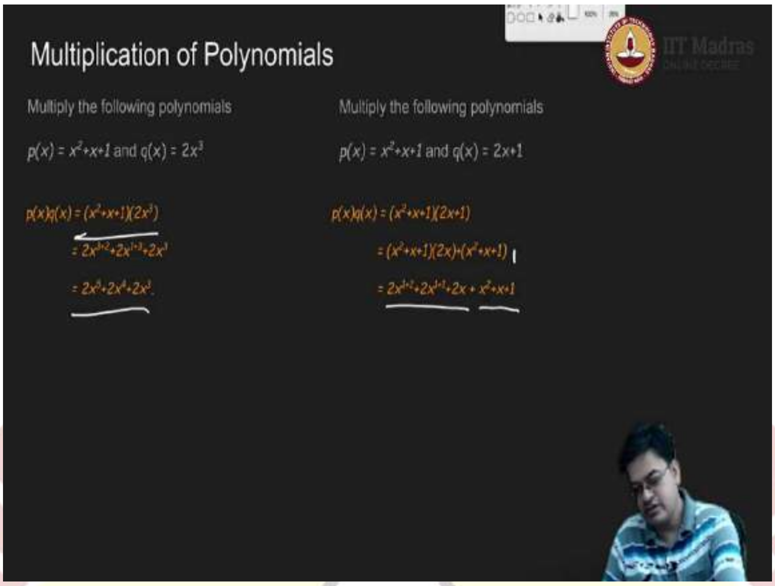
Now, as you can see the this polynomial has three terms 1, 2 and 3. So, it is not a binomial; it is a trinomial. So, my foil method will not work here. So, foil method will work only for these kind of expressions which are binomials. So, let us go ahead and try to consider a similar expression that is a quadratic expression and another binomial, and try to see how can I extend the basis of foil method right.

So, here is a binomial $2\,x + 1$. And here is a general polynomial quadratic polynomial which is $x^2 + x + 1$ same. Now, what will you do? So, naturally you will consider $p\,(x) \times q\,(x)$ which will be written in this form. Now, if I want to extend the basis whatever I did for monomial, that means, I need to convert this into two monomials.

So, what are those two monomials? One monomial is $2x$; another monomial is 1. So, if I treat them separately that is if I write them in this manner, let me erase this, that is I have written them in this manner.

Then what can I do about it, that means, now this turned out to be a same expression instead of x^3, here it is x that is all is the difference right. So, whatever I did here, I can do it here. And the last term is actually multiplied with 1 which it suppressed because multiplication with 1 will not change anything. So, I do not have to worry about the last term.

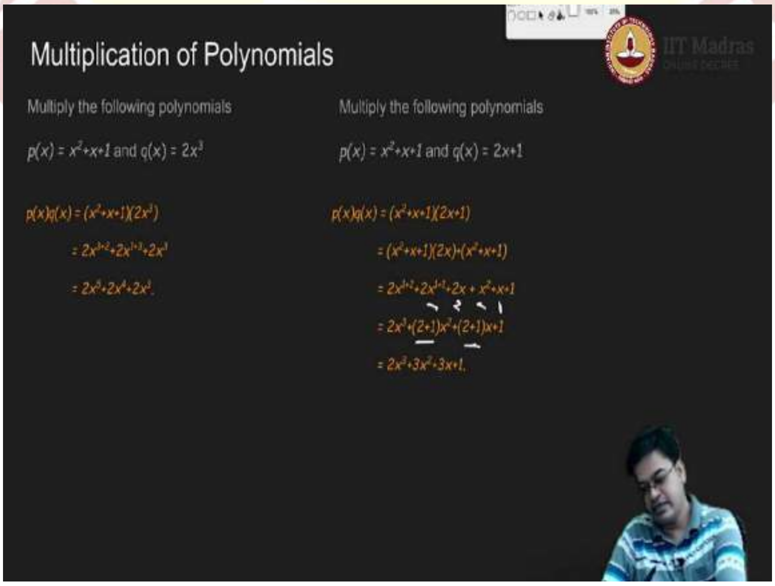
(Refer Slide Time: 04:05)



Now, I will multiply this $2x$ with all the terms in for of $p(x) = x^2 + x + 1$ which is similar to this particular thing. So, I will get $2\,x^{1+2} + 2x^{1+1} + 2x + x^2 + x + 1$. Now, the job is very simple.

You can treat this as one polynomial, and this one as a second polynomial, and then we have to add. How we add polynomials? We will add polynomials by matching the exponents, matching the exponents of x. So, if I want to add these two polynomials, what will I do, I will simply match the exponents and I will add them which is given here.

(Refer Slide Time: 05:02)

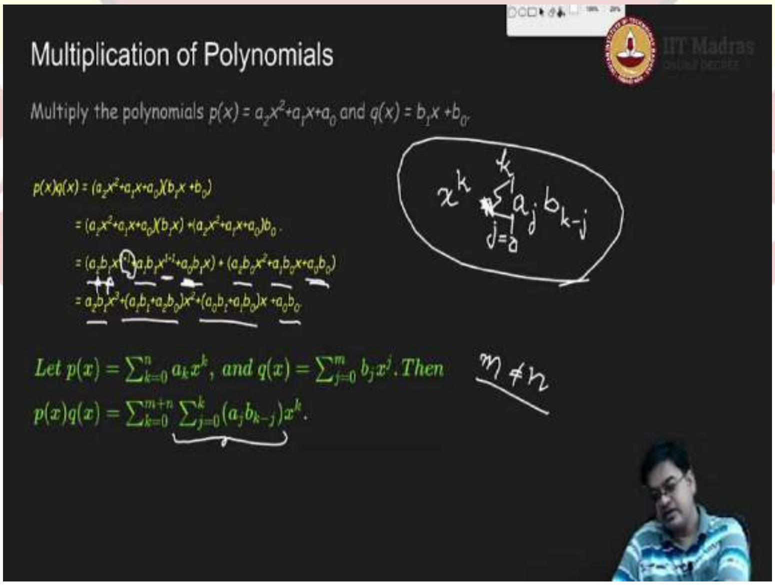


So, in this case $2\,x^3$, there is no competing term for x^3. So, it remains 2; x^2 comes here and here, therefore, I added the two which gives me 2+1, in a similar manner the terms containing x are these two. So, I have added these two, so 2+ 1 $x + 1$ which is similar to what we have seen in the last video of addition of polynomials. And therefore, we get the answer to be equal to $2x^3 + 3x^2 + 3x + 1$.

So, effectively what we have done is we know how to multiply the terms term by term. And finally, if at all I want to seek an extension of a foil method, it will be a term by term multiplication of polynomials, that means, you take the polynomial of least degree and multiply it with the polynomial of highest degree term by term, add those term match the powers and then write your answer. So, this is one prototype that we can follow for finding multiplication of polynomials or result of the multiplication of polynomials.

Now, the next question is can I generalize this method or can I answer it programmatically, that means, can I give a simple formula for what the coefficient of one part x^m will be? For example, in this case can I give a general formula what will be the coefficient of $3x^2$ provided I know polynomials $p(x)$ and $q(x)$. So, to answer that, let us go ahead and try to find a general formulation of this form of this formula.

(Refer Slide Time: 07:03)



Let us go ahead. And if you are asked given one quadratic polynomial and one linear polynomial, you are asked to compute $p(x) \times q(x)$, how will you go about this? This is what our task is now. So naturally I will write $p(x) \times q(x)$, and then I will convert each of them into monomials that is one monomial will be $b_1 x$, and second monomial will be b_0.

In this case, what will happen is we will simply multiply them as a separate term by term multiplication. So, in earlier case our b_0 was 1 when we studied one example. But here we are considering a general expression, and none of the expressions are 0 that is what we are assuming none of the coefficients at a_2, a_1, a_0, $b_1$and b_0 none of them are 0.

For example, if you consider $b_0 = 0$, then this term itself will vanish the second term itself will vanish; you will not have the second term. So, we are assuming that all terms remain in the loop ok. So, now it simple, the job is multiplying these two polynomials, and you will get some answers that is ok, but now our main worry is to find a pattern in these answers ok.

So, now, when I multiplied this, if you look at this particular expression that is $(a_2 b_1 x^{2+1} + a_1 b_1 x^{1+1} + a_0 b_1 x^1) + (a_2 b_0 x^2 + a_1 b_0 x^1 + a_0 b_0)$. Here you take a pause and examine the terms. For example, this term contains the coefficient of x^3, this is $2 + 1$.

So, x^3. So, in that case, what is happening here is if you look at the suffixes of the coefficients this is a_2, this is b_1, so together they will sum to 3. In a similar manner, you look at this term which contains x^2. And you look at the suffixes of the coefficients that is $a_1 b_1$, together they will sum to the exponent that is a 1+1=2. So, this should be a coefficient of x^2.

Then if this logic is correct, what should be the coefficient of a constant? The coefficient of the constant that is x^0. So, the coefficient of the constant must be $a_0 b_0$. In a similar manner you can ask the question what is a coefficient of x? If you asked that question, you will naturally get the answer you collect all the in all the coefficients such that their suffixes will sum to 1 that is $a_1 b_0 + b_1 a_0$. So, is there anything called $b_1 a_0$? Yes, it is here.

So, this what we have actually done is we have figured out a pattern; that means, if I want to find the coefficient of x^k, then better the sum should be some $a_j b_{k-j}$, so that they both will sum, they both will sum to it is not equal to the this is I am saying x raise to coefficient of x^k will be equal to of the will be of the form a_j+b_{k-j}. So, with this understanding, let us go further and try to rewrite this sum ok.

So, once I have rewritten this sum, my analogy is further amplified. For example, if you look at the coefficient of x^2, yes, it was it is a_1b_1 and a_2b_0 which is the coefficient of x^2, so that also means this means if I can sum over this j from 0 to what point to a point where I want the sum the exponent is raised to k, then I will get all possible combinations where sum is actually k.

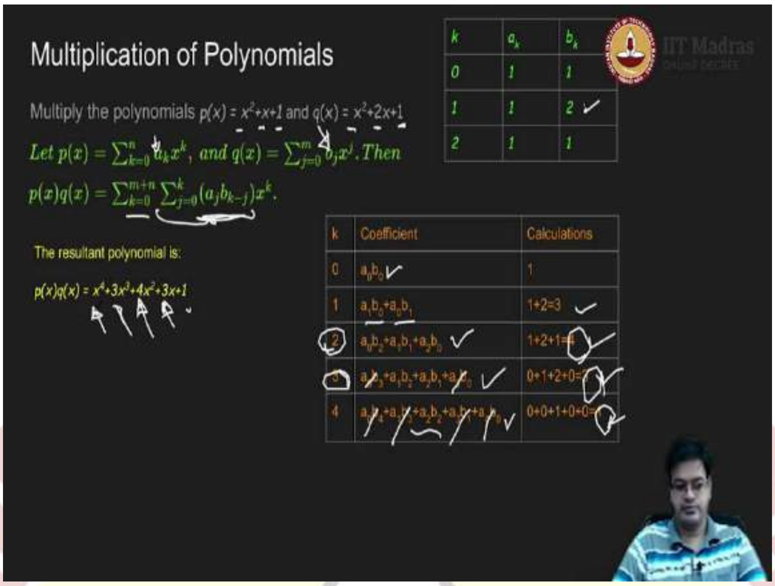
In a similar manner, you can pause this video and verify whether you are getting the same expression for x^1 and all others right. So, with this understanding, I am ready to generalize this demonstration or this theory for a polynomial of an arbitrary order.

Let us consider polynomials of degree n and m, and try to find the general answer for them, and that answer will be in this form. So, if you are given a polynomial of degree $n, p(x)$, and if you are given another polynomial of degree $m, q(x)$, let us say $m \neq n$.

Even if $m = n$ it does not matter, but for our purposes let us take $m \neq n$, then what will be the coefficient of each of the x^k's? The coefficient is actually given here, $\sum_{j=0}^{k} a_j b_{k-j}$ this is what we have figured out in this expression is the coefficient of x^k.

Then the question is how far the degree will go? The degree will go till $m+\ n\ m \neq n$; if $m = n$ then the degree will go to $2\,n$ that is ok. So, $k = 0$ to $m+\ n$, and each of the coefficient of x^kwill be $\sum_{j=0}^{k} a_j b_{k-j}$. Now, let us demonstrate this idea with one example. Let us go ahead and see one example of this idea.

(Refer Slide Time: 13:36)



So, now, you have been given two polynomials two quadratic polynomials and you are asked to compute the multiplication of these two polynomials. One way is very simple you will go with term by term multiplication, and it simply means you have to multiply the terms of second polynomial with the first polynomial in a term by term fashion, or you can actually use the formula that I have given you in the previous slide. So, you can pause this video, and try to compute by yourself or you can go along with me.

So, let us recall that formula again that is $p(x)$ is equal to sum a, so my polynomial is a polynomial of degree n, and $q\ (x)$ is a polynomial of degree m. In this case, in this particular example, the polynomial the first polynomial is of degree 2 as well as the second polynomial is of degree 2.

So, in order to find the product of these two polynomials, what do we need to find is we simply need to find the coefficients of x^k. So, let us first identify what are a_k's and what are b_k's, j is a dummy index. So, it does not matter.

So, let us first identify what are a_k's and b_k's. So, a_0 as you can see is 1, b_0 is 1, a 1 is 1 again, b_1 is 2, correct, this is correct, and then a_2 and b_2 both are 1. So, I have enlisted all the coefficients of this particular expression, $p\ (x)$ and expressions $p\ (x)$ and $q\ (x)$. Now, we need to use this formula, then this formula which gives me the sum. So, let us use this formula and figure out.

Remember, all the coefficients that are not listed here. For example, what will be a_4, if at all, I will write a_4, what will be a_4 in this expression? It will be 0. What will be a_3 in this expression? It will be 0. So, all the coefficients that are not listed here are 0s. Keep this in mind and try to answer the question.

So, now, computation of coefficient; it is very easy. So, let us start with 0th degree term that is constant term. So, here $k = 0$. So, the summation will actually go from $j = 0$ to 0, that means, it will have only one term which is a_0b_0.

What is a_0b_0? Look here 1 into 1, so it will give you 1 ok. Let us go for a degree 1 term. So, j is equal to 0 to 1, j is equal to 0 to 1, so it will have, $a_0b_1 + a_1b_0$ these two terms are there. So, let us compute them through this table a_1 is 1, b_0 is 1, so this will retain 1. a_0 is 1; b_1 is 2, so it will give you 2. So, together it is 1+ 2 = 3.

Let us go for a second order term that is the monomial with degree 2. So, in this case, j will run from 0 to 2. So, I will have a_0b_2, a_1b_1, a_3b_0, a_1b_1, a_2b_0, a_0b_2, a_1b_1, this is correct. Just go ahead and compute these terms, a_0 is 1, b_2 is 1, so you will get 1, $a_1$is 1, b_1 is 2, so you will get 2. And a_2b_0 that is a_2 is 1, b_0 is 1, so you will get another 1. So, you will get the sum to be 4.

Let us go for a third term x^3 term, and just simply substitute this. So, we need to find all possible combinations. So, if it is a degree 3 term and we start with a_0, it will be a_0b_3, a_1b_2, a_2b_1, a_3b_0, these are the terms. And then you simply compute them.

Remember here now we came up with b_3. What is b_3? b_3 is not listed here, that means, b b_3 must be 0. In a similar manner here a_3 must be 0 correct. So, these 2 terms are chopped off right away they are 0. So, let us focus on the other 2 terms the first term you can easily verify because b_2 is 1, and a_1 is 1. And a_2b_1, b_1 is 2, a_2 is 1, so it will be 2. So, 1+ 2=3; this is correct.

Now, the final term is a degree 4 term, correct. If you do a term wise multiplication, what you will come up with is because the degree 4 will be contributed by the highest order terms.

So, you will simply multiply $x^2 \times x^2$, and you will get only 1 term. But in this formulation what we are doing here is we are taking all possible terms of degree 4. So, even though they are 0, we will first list them, and we will put them as 0s.

So, now, when we consider degree 4 term, I will get a_0b_4, a_1b_3, a_2b_2, a_2b_2, a_3b_1 and a_4b_0. So, all these terms are here. And most of the terms will obviously, be 0 only 1 term is a contributor.

For example, a_0b_4 is 0, a_4b_4 will be 0, a_1b_3 is 0, a_3b_1 is 0. Why? Because b_4, b_3, a_3, a_4 all are 0 only term that will contribute is a_2b_2 which will be 1×1, so 1. So, this gives us a clear cut answer, and this is a systematic way to multiply two polynomials.

Therefore, the resultant polynomial $p(x) \times q(x)$ simply write the terms from this table, so this is a coefficient of x^0 is 1, so the constant term 1 is here coefficient of x^1 is 3, so $3x$ is here. So, in a similar manner x^2 coefficient of x^2 is 4. So, you will get $4x^2$ here ok; so $3x^3$ correct.

So, this is also done. And then x^4 has only 1 term as 1, so x^4. Therefore, you got the resultant polynomial to be equal to this. Now, remember one side note the multiplication of two polynomials will always fetch you a polynomial again ok. Next operation is division which we will see in the next video, but the division of two polynomials will not always lead to a polynomial. We will see that in the next video.

Bye for now.

Thank you.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

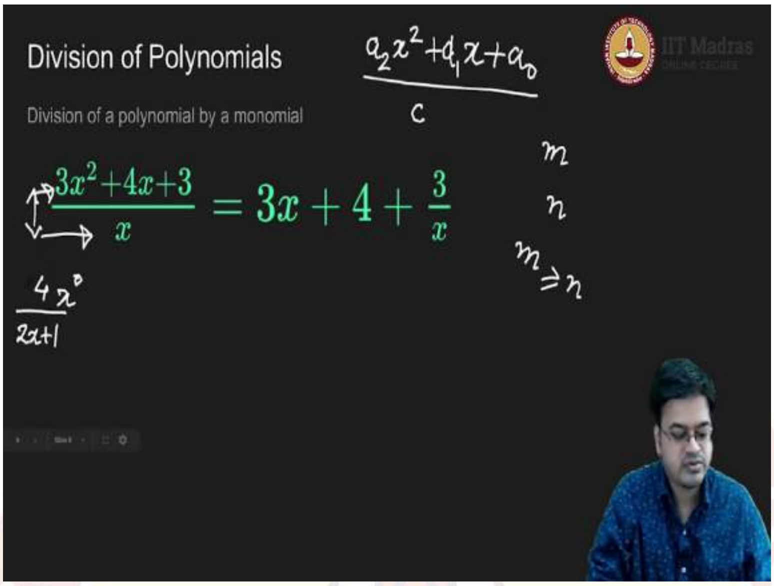
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 34
Algebra of polynomials: Division

(Refer Slide Time: 00:15)



In this video, let us have look at Division of polynomials. What is a division of polynomial? We have already familiar with division of polynomials, but we have not done in a rigorous manner and we do not know all possible cases that can occur while considering division of polynomials, that is why it is important to look at division of polynomials.

We know some cases like for example, if I have been given a polynomial say $a_2x^2+a_1x+a_0$ and if I am told that if this polynomial is divided by a constant say c. Then, I know what is the resultant polynomial. It will be $\frac{a_2x^2}{c}+\frac{a_1x}{c}+\frac{a_0}{c}$ that will be the polynomial. So, this case, we are already familiar with.

Now, let us go to one more level of extension. Suppose, this polynomial is divided by a monomial; that means, we are considering a division of a polynomial by a monomial.

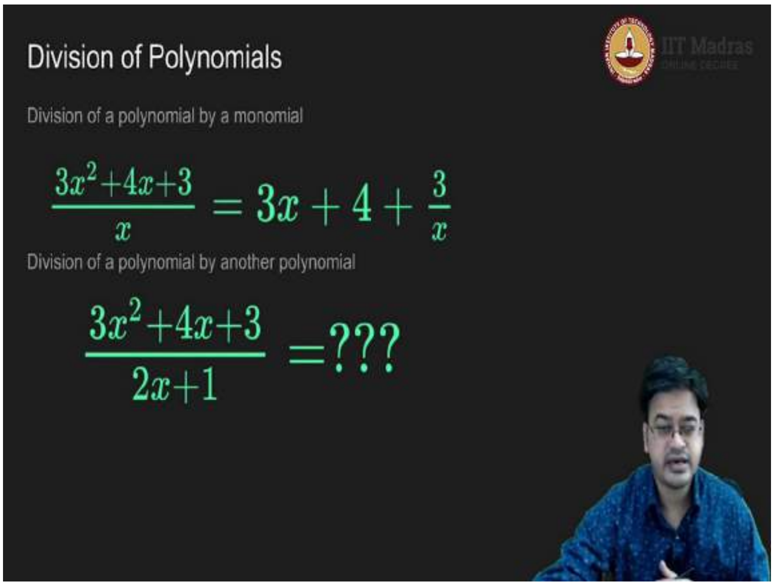
Monomial means, the polynomial that contains only one term, only one variable term. So, in that case, let us take this example. So, $\frac{3x^2+4x+3}{x}$, ok.

So, notice few factors here. In this case, when I am considering a division of two polynomials, the numerator and the denominator; the denominator should always have a degree smaller than the degree of the numerator. If it is not the case, let us say the numerator has degree m and the denominator has degree n , then what I am saying is the degree of the numerator m should always be greater than or equal to n. If it is not the case, then the division is not possible ok.

For example, let us consider one case, where I am considering a constant polynomial let us say 4 and I am dividing it by some polynomial which is $2x+1$. Here, I cannot divide this; I cannot divide by this polynomial because there is no corresponding x term. Here it is x^0 .

So, I cannot divide this polynomial because the degree of the polynomial plays a crucial role. So, in this case, the division is not possible, I have to keep this function as it is. Let us keep this point in our mind and consider division of polynomials. So, now, I am dividing a polynomial with a monomial, how will you handle this?

(Refer Slide Time: 03:09)



So, a monomial is simply x here. So, what I will do is I will split this with this addition sign, I will split each of them in separate terms. So, I will consider the term $\frac{3x^2}{x}$ that will give me a term, when I consider this term it will give me a term $3x$.

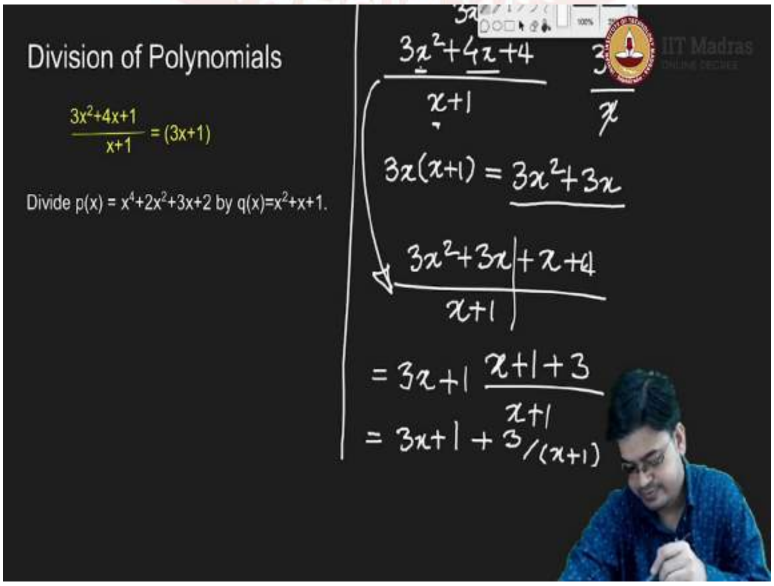
When I consider $\frac{4x}{x}$, I will get a term 4.

Now, as I mentioned earlier when I consider the term $\frac{3}{x}$, the degree of 3 is a constant because 3 is a constant polynomial the degree is 0. So, I cannot divide this polynomial. So, this will automatically influence this decision that it will remain as it is, that is $\frac{3}{x}$. So, these are some key things while dividing polynomial by a monomial.

Now, the key idea is I want to divide a polynomial with another polynomial. Let me erase this first. So, now, I want to divide a polynomial with another polynomial. So, how will I go about this?

That is I want to find something of this sort. Let us address this question in a video ok. So, apparently, I do not have any practical way to divide this right now; but from whatever theory I learnt about quadratic functions, can I derive something? That is what the question is. So, we will try to figure out some more methods in this video.

(Refer Slide Time: 04:51)



So, let us continue and take a question, this is the numerator that is given to me divided by another polynomial which is $x+1$ and I want to figure out what this will be equal to? Let me take this polynomial over here and try to figure out what this polynomial will be equal to. So, now, I have $3x^2+4x+1$ and it is divided by $x+1$. Now, if I want to divide the numerator by the denominator, what I should see is ok, the denominator has the highest degree which is x. The numerator has highest degree which is x^2 and now, how will I be able to get rid of the denominator for some at least for some terms?

So, in that quest, what I will see is I will simply take the first term over here and the first term over here and I will see like monomial, I will see what is $\frac{3x^2}{x}$. This I can do very easily because both are monomials. So, x vanishes with this square and I will I am left with $3x$. So, next thing that I will do is I will consider $3x(x+1)$. So, this actually gives me the answer $3x^2+3x$. Now, I will try to figure out this term in the expression that is given in the numerator.

So, if I want to figure out the expression that is given in the numerator, I can easily split this $4x$ as $3x+x$. If I can do so, that means, I can take this term and based on this logic, I can actually write this as $\frac{3x^2+3x+x+1}{x+1}$.

Now, I can intelligently split the term over here and I can divide this and I can take this as a separate term and divide this. So, now, you can readily see the answer will be here, $3x(x+1)$ that will get cancel off with $x+1$ and over here, it will be 1. So, the answer is $3x+1$. Therefore, such a division is possible, ok.

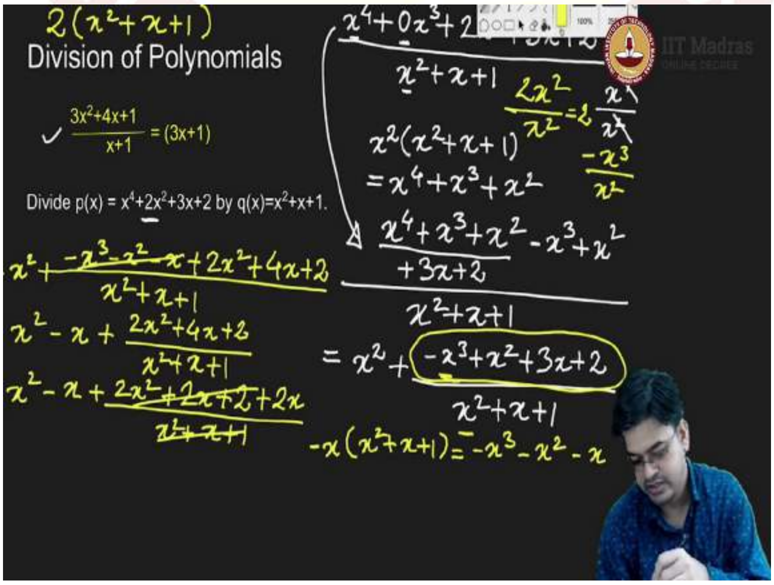
Let us verify whether the answer is $3x+1$; yes. So, I have demonstrated you how to divide a polynomial using simple method by the method of factorization that we have already used. Now, this is because $x+1$ was the factor of $3x^2+4x+1$.

What if $x+1$ is not a factor of $3x^2+4x+1$, what would have happened? Let us use this example to understand our findings. So, let me take a eraser and let me write that this instead of 1, let me put ok, let all other things remain constant, what makes it a factor; that $x+1$.

So, I will simply what I will simply do is I will simply change the term to 4 ok. Now, $x+1$ is no longer a factor, still I will continue with the same method, I will take this $\frac{3x}{x}$ 3x by x. So, I can consider this x^2+3x , only difference is this will be $x+4$. In that case, what happens is $3x^2+3x$. So, that will give me $3x$ into let me rewrite this as $\frac{1(x+1+3)}{x+1}$. So, that again gives me an edge that is this is nothing but $3x+\frac{1(x+1)+3}{x+1}$ getting cancelled.

So, this will remain as $\frac{3}{x+1}$. So, this is how even if it is not a factor, I can divide the polynomial. Now, as we have started by giving some for addition, multiplication, subtraction, we have given some algorithms. So, now, we need to identify such algorithm for division of polynomials. To do that, let us first solve this complicated problem in this simple manner and try to derive an algorithm and try to derive an algorithm for by solving this problem.

(Refer Slide Time: 10:15)



So, the problem is, I want to divide the terms divide a polynomial $p(x)=x^4+2x^2+3x+2$ by $q(x)=x^2+x+1$. So, the $p(x)$ is a polynomial of degree 4; $q(x)$ is a polynomial of degree 2. So, how will I go about this? So, again, I will apply a strain strategy that is I will start by writing x^4 plus remember here, it directly goes to $2x^2$. So, but I want the term containing x^3 also to be present $0x^3+2x^2+3x+2$ and this term is divided by x^2+x+1.

So, remember our first step in the last example was first you take the first term over here and take the first term over here. Then, take a consider a division of these monomials. This will give me x^2. So, what I will do now is, I will consider $x^2(x^2+x+1)$. This will give me actually $x^4+x^3+x^2$.

Now, as per our earlier strategy while solving this problem, we have adopted a strategy that I will add these terms over here. So, if I add these terms over here, then I will subtract appropriate terms over here. So, in this case, x^4 is already there, x^3 was not there and here, x^3 is there.

So, I need to subtract that x^3 from this expression and then, I need to subtract from $2x^2$, I will split this into two. So, let us rewrite this expression, that is the numerator of this expression, x^4 is already there. In order to cancel the denominator, I need x^3 over here.

So, I need to add x^3+x^2 ok; but this x^3 was not present here, it was $0x^3$. So, naturally the next step will be to eliminate x^3 from here. So, that it will retain a legacy of this term. So, if I have eliminated x^3, then it is $2x^2$ of which $1x^2$, I have taken out, so this will be another x^2+3x+2 as it is.

So, let me write that term as it is $3x+2$ and now, if I divide this term by x^2+x+1, then what I will get here is take these first three terms and keep the remaining term as it is ok. So, if I do that, then what will happen is this term x^2 will come out as common plus now, what happens?

This term vanishes, this term vanishes, this term because x^2, I can take out common; from these three terms, I can take out x^2 common that is what I have written and it cancels with the denominator. So, whatever is remaining are the remaining term that is $-x^3+x^2+3x+2$ and this thing is divided by x^2+x+1.

Now, is our division over? No, because the numerator over here has a higher degree than the denominator. Therefore, our division is not over. So, again, I will follow a similar step, I will simply change the color so that I will have a better view ok. So, let us change the color and have a better view of this.

So, let me write it here from this; from this step, I can go here and say ok. So, this is in fact, equal to x^2 plus now you look at this term $-x^3$ and x^2. So, you divide $\frac{-x^3}{x^2}$ which will give you $-x$. So, in this case, you will multiply $-x(x^2+x+1)$.

So, if you multiply $-x(x^2+x+1)$, what you will get over here is $-x^3-x^2-x$, this is what you will get. So, you write this term as it is, that is $-x^3-x^2-x$.

Now, from this term, you adjust the terms. So, $-x^3$ is already there, so I do not have to compensate for this term. But there is a plus x^2, there is a plus x^2 and here there is a $-x^2$. So, that will give me plus $2x^2$ because I am compensating for this extra $-x^2$ added in this term, then there is a $-x$ and over here it is plus $3x$.

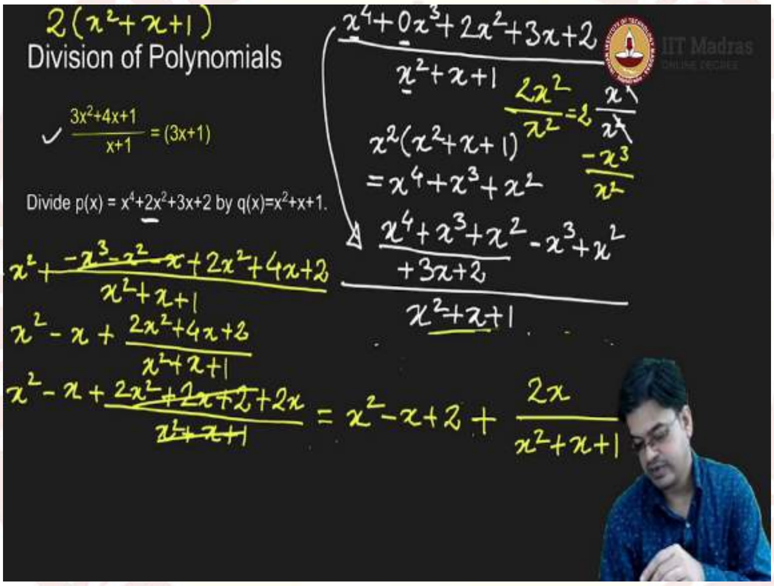
So, I have to add one x for this $-x$. So, that will give me plus $4x$ plus and there is no competition for a constant term upon x^2+x+1. Now, you can take out x common and this will cancel off, this term will cancel off with this term by taking x common. So, it is x^2-x is in common plus what you are left with here is $2x^2+4x+2$ upon x^2+x+1 ok. Again, you will apply a similar procedure that is you will actually divide $\frac{2x^2}{x^2}$. So, you will get 2. So, essentially what you; so, when you do that, when you divide $\frac{2x^2}{x^2}$, you will get 2.

So, when you will multiply this number by 2, let me write it here that is $2(x^2+x+1)$ ok. So, in this case, what you will get is $(2x^2+2x+2)$ of which $2x^2$ is already there. So, I will continue over here itself x^2-x+2x^2 is already there, $2x^2$. So, let it be $2x^2+2x$, over here there is plus $4x$. So, I can split $2x$ over here plus $2x$

plus 2. So, that will again come plus 2 as it is here. So, what is remaining now is $2x$ upon x^2+x+1 .

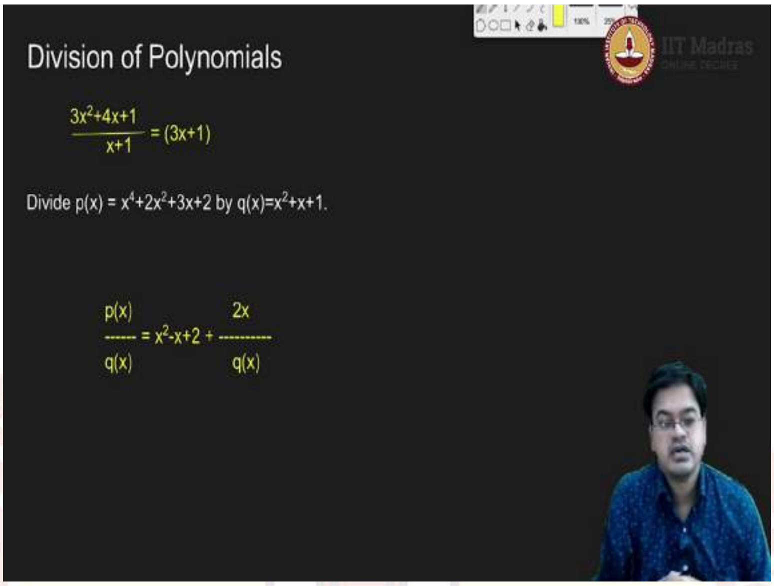
So, now if you look at this term, what you will get is you can take out 2 common and this will cancel off with this denominator and therefore, the final expression, I am running short of space. So, let me erase some terms over here. Let me erase some terms over here so that I will get some space.

(Refer Slide Time: 19:43)



So, you can rewrite this as to be equal to $x^2-x+2+\frac{2x}{x^2+x+1}$. This will be the final answer to this division ok. So, this is how we can actually do a division of two polynomials ok. Let us remove this and see whether to verify whether we have got the final answer to be correct or not. So, I have removed it, you must have noted the answer.

(Refer Slide Time: 20:21)



And the final answer that we have got here is $x^2 - x + 2 + \frac{2x}{q(x)}; q(x) = x^2 + x + 1$. Yes, so I have got the correct answer. So, here while doing this, we have derived one algorithm which we will emphasize in the next slide.

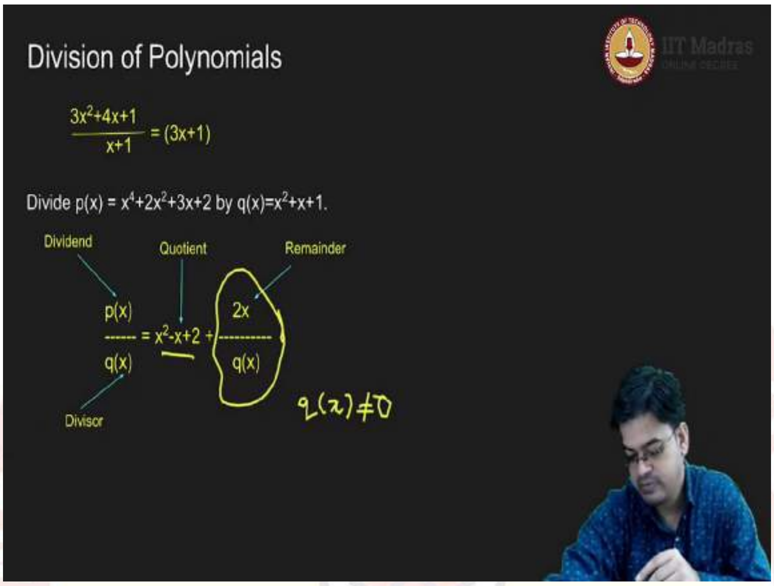
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 35
Division Algorithm

(Refer Slide Time: 00:14)



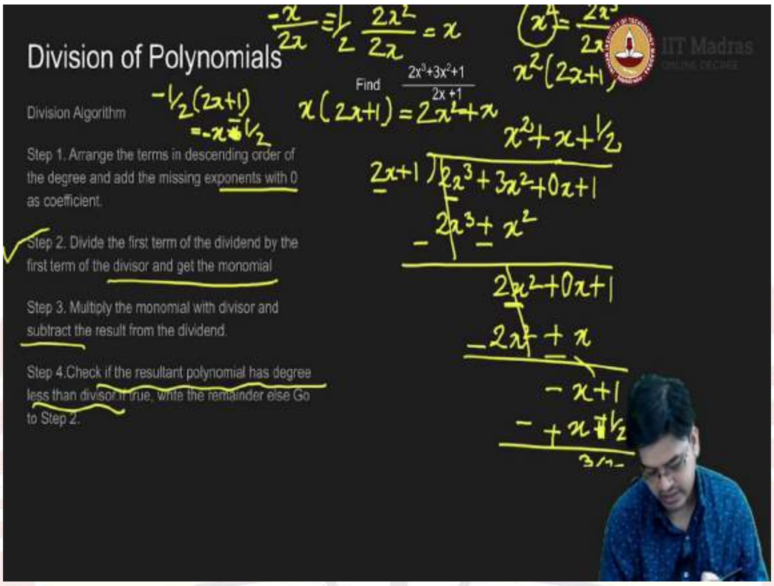
So, let us go to the next slide and emphasize the algorithm that we have derived just now. In order to understand the algorithm, you need some terminology. For example, this $p(x)$ is called the divident; the $q(x)$ is called the divisor. The term that you get over here, here is called the term that you get over here is called the quotient. And the $2x$ that you have got is called the remainder.

Remember you will declare something as a reminder only when the degree of the denominator is higher than the degree of the numerator, this is the strategy that we will follow.

So, now, you are very clear about the terminology, the numerator is the divident, the denominator is the divisor, the term the polynomial term that you get after dividing is called the quotient, and the rational and the remainder is something that where the degree of the numerator is smaller than the degree of the denominator.

This is also called a rational function. If you look at polynomial as a function, then division of two polynomials is a rational function, only condition that we are enforcing is $q(x)$ cannot be equal to 0, this is the condition is which is always in place. Let me eliminate this and let us go and study the algorithm.

(Refer Slide Time: 01:56)



So, for division of polynomials, we will use the following division algorithm which we have derived just now, where in the first step what we will do is we will arrange the terms in the descending order of the degree, and add the missing exponent with 0 as a coefficient.

Then after adding the missing 0, 0 as a coefficient after adding the missing exponents, next what we will do is, we will take the first leading terms or the leading monomials, and we will divide the dividents monomial, the leading monomial of the divident and the leading monomial of the divisor together. And we will get some number which is which we will call as quotient, temporary quotient and that quotient we will multiply with our divident.

Once we multiply with our divident, what we will actually do is we will subtract that from the original expression for the polynomial that is our numerator. Whatever is remaining, we will treat that as the next divident. Once we treat that divident, then we will check if the degree of that new divident is higher than the degree of the denominator or divisor. If yes, then we will continue with the procedure; if no, we will terminate the procedure; this is how we will give the division algorithm.

Let us understand this division algorithm by using one example. So, here is an example. This is the numerator $2x^3+3x^2+1$ divided by $2x+1$, and I want to find the answer to this question. Let us figure out how to find the answer.

So, in the earlier quest, what I did is I have used the standard numerator denominator. Now, there is a popular method for division of the polynomial which is called long division, which works in a similar manner and the same division algorithm works, but you will have a better handle over the terms.

So, in this long division, what you will do is you will put a parenthesis over here, and you will put $2x+1$ outside the parenthesis, and you will put this term that is $2x^3+3x^2$. Now, remember the first step plus $0x+1$ ok. So, this is how we will write. Now, according to our standard terminology, what we will do is we will take the leading terms $2x$ and $2x^{3.}$ So, somewhere in the rough you do that. What is $2x^3$ divided by $2x$? This will give you $x^{2.}$.

So, you write x^2 over here, multiply x^2 with $2x+1$. Once you multiply x^2 with $2x+1$, write that term over here, $2x^3+x^2$. Now, according to our algorithm divide the first term of the dividend by the first term of the divisor and get the monomial that monomial is x^2 over here. Next step, multiply the monomial with the divisor and subtract the result from the divident. So, this is the result from the divident, result from multiplication, and you are subtracting it from the divident.

So, this will cancel off. So, this will give me 0 and $3x^2-x^2$ will give me $2x^2+0x+1$. This is the result ok. So, now, this result, I will check whether the degree of this result this polynomial that I have obtained is greater or smaller than this ok, that is what we will do. Check if the resultant polynomial has a degree less than the divisor that is not true.

So, we will go to step 2. What is the step 2? Which is this, divide the first term, first term of this divident with this that is you will divide $\frac{2x^2}{2x}$. So, what you will get here is

x . So, you will simply add x over here. And then you will multiply that x with $2x+1$. Once you do that, you will get $2x^2+x$. So, you write here $2x^2+x$.

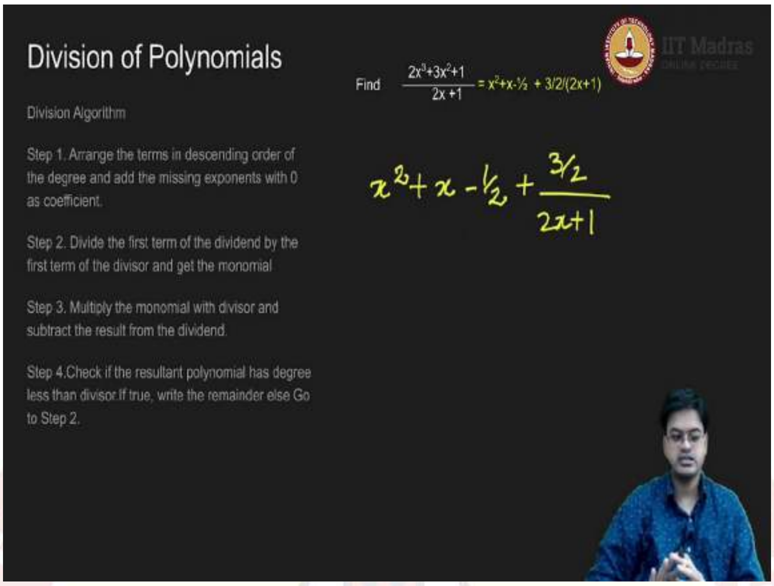
Then what is the next step? You subtract it from the result, so minus, minus $2x^2$ vanishes, this gives me $-x+1$, ok. So, $-x+1$, again I will go to the same step because this degree is same, it is not less than the degree of the denominator.

So, I will again follow the same procedure; $\frac{x}{2x}$ which will give me $\frac{1}{2}$. So, naturally I will add $\frac{1}{2}$ over here. And once I add $\frac{1}{2}$ over here, when I multiply $\frac{1}{2}$ with $2x+1$, what I will get here is $x+\frac{1}{2}$. So, I will write that $x+\frac{1}{2}$.

But remember over here the thing was $-x$. So, I should what I should have done is I should have multiplied -1 to the x that means, $\frac{-x}{2}$. So, the answer is $\frac{-1}{2}$, and you will multiply $\frac{-1}{2}$ over here, so $-x$ this will not be plus this will be $\frac{-1}{2}$, so $-x-\frac{1}{2}$ which will be given a negative sign. So, this will be $x+\frac{1}{2}$. So, I will get the answer to be equal to $\frac{3}{2}$ ok. So, the answer is $\frac{3}{2}$.

So, what is what will be the resultant answer? This should not be plus, 1 minute, let me make it very clear. This cannot be plus; this should be minus, because I have to multiply with $\frac{-1}{2}$. And here it is the remainder is $\frac{3}{2}$. So, what I got here is $x^2+x-\frac{1}{2}$ and the as a quotient, and the remainder is $\frac{3}{2}$.

(Refer Slide Time: 09:23)



So, let me rewrite it again that is I got $x^2+x-\frac{1}{2}+\frac{\frac{3}{2}}{2x+1}$, this is what I got. Let me verify this result. And we have demonstrated the algorithm, yes, $x^2+x-\frac{1}{2}+\frac{\frac{3}{2}}{2x+1}$, this is how we will consider division of polynomials in general.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

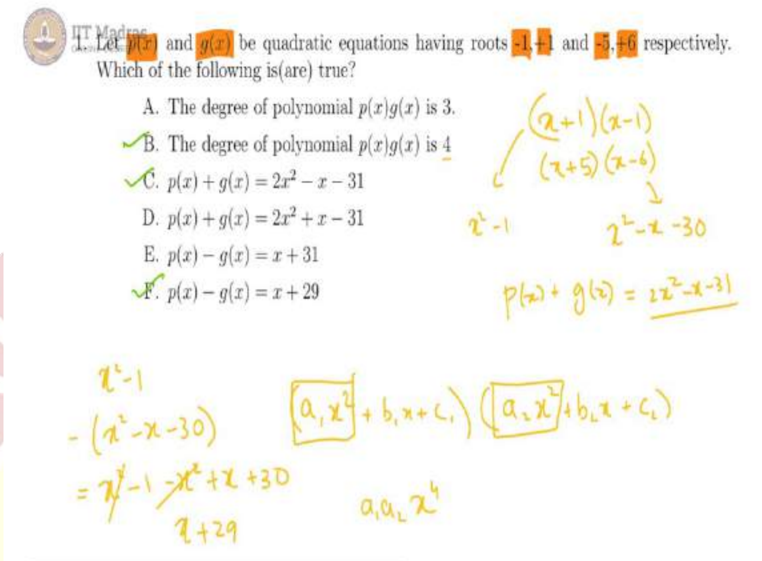
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 06 - Tutorial 01

(Refer Slide Time: 00:16)



Hello, mathematics students. In this week's tutorials, we will look at some questions based on polynomials and the algebra of polynomials. In this question, we have two quadratic equations, which are $p(x)$ and $g(x)$, presumably equal to 0, and they have the roots, $-1 +\ 1, -5+5$ respectively. Then the degree of the polynomial $p(x)\ \times g(x)$ is three, it is not because you have two quadratic equations, and you are multiplying them.

So, the x^2 terms will have to necessarily multiply, so $(a_1x^2 + b_1x + c_1) \times (a_2x^2 + b_2x + c_2)$, when you multiply these, this term, and this term will have to be multiplied and you are going to get $(a_1a_2x^4)$, so the degree has to be 4, which is this. So, B is correct. And then we have the sum, is equal to, so we need to find the respective quadratic equations now for this, so this would be $(x + 1)\ \times (x - 1)$, the other would be $(x + 5)\ \times (x - 6)$.

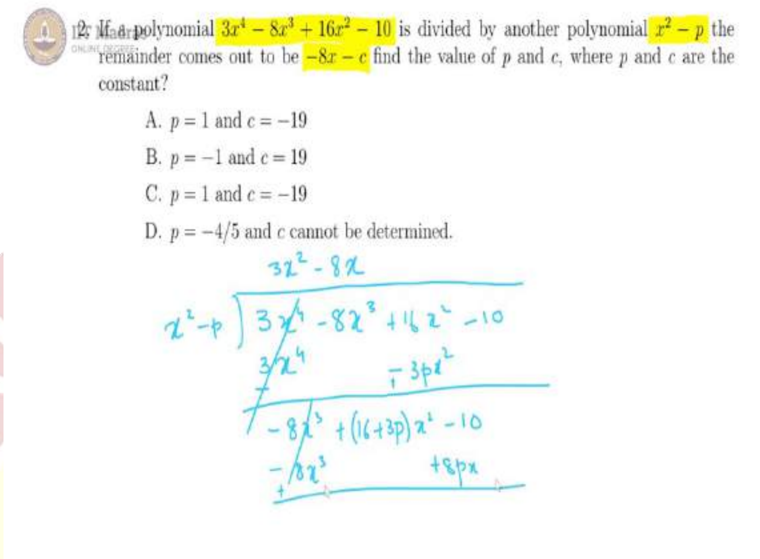
So, this gives us this is, $x^2 - 1$. And this is essentially $x^2 - x - 30$. So, when we add these two, we get $p(x) + g(x) = 2x^2 - x - 31$. So, C is correct, and that would imply D is wrong. And now we are looking at the difference $p(x) - g(x)$ and that would give us $x^2 - 1 - (x^2 - x - 30)$ =x square minus 1 minus of x square minus x minus 30, which is $x^2 - 1 - x^2 + x + 30$. So, $x^2 - x^2$ cancel off and you have $x + 29$. So, E is wrong and F would be correct.

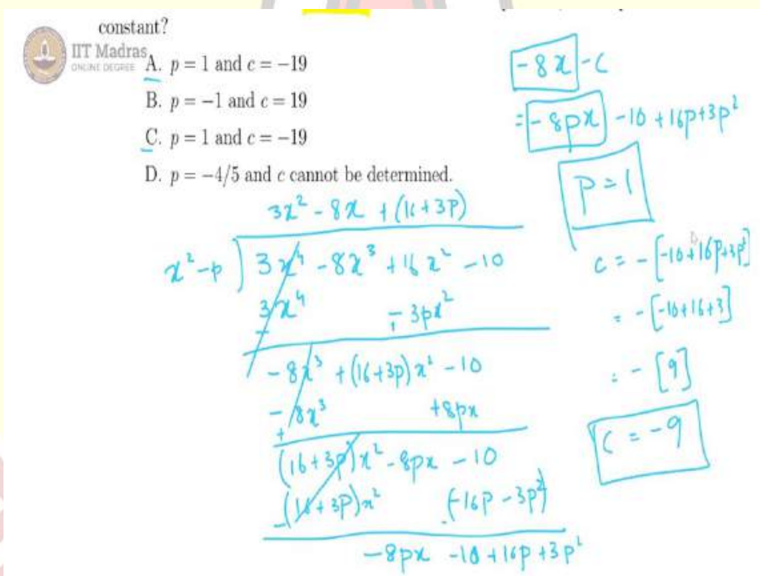
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 06 – Tutorial 02

(Refer Slide Time: 00:16)





In question number two, there is this polynomial, $3x^4 - 8x^3 + 16x^2 - 10$ and is divided by another polynomial $x^2 - p$, then the remainder comes out to be $-8x - c$. They are saying find the value of p and c. So, let us do the division, then, we have $3x^4 - 8x^3 + 16x^2 - 10$ and here we have $x^2 - p$.

So, this gives us $3x^2$ to start with and so this will be $3x^4 - 3px^2$, so we should write it there, $-3px^2$. And this goes off, and we get $-8x^3$ + this becomes +. So, $16 + 3px^2 - 10$. So, now we have $-8x$ coming up here, which gives us $-8x^3 + 8x$, so $+8px$ and this goes off again.

So, we have $16+3px^2-8px-10$. So, we again multiply by 16 plus 3p here, and that gives us $16+3px^2$. And there is no x term, we get minus $16p-3p^2$, then this of course cancelled again. So, we are left with $-8px-10+16p-3p^2$, because this is being subtracted.

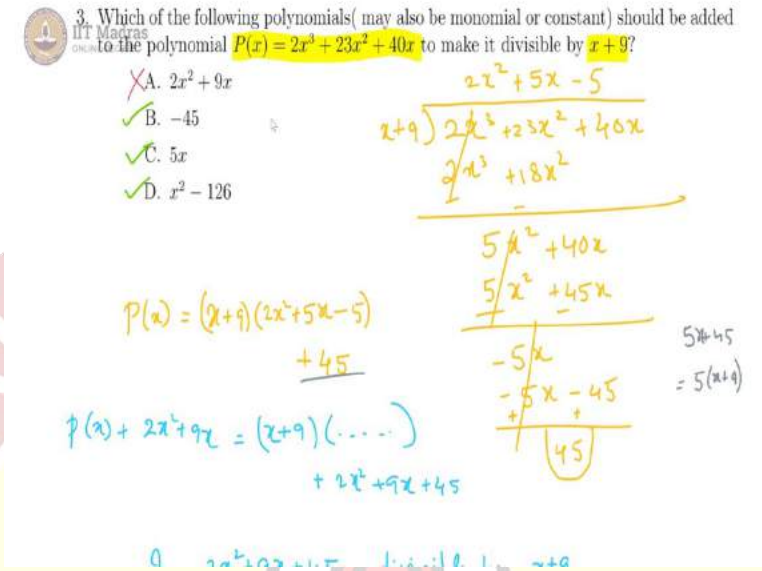
So, they are saying this remainder is $-8x-c$. And that is equal to $-8px-10+16p-3p^2$. So, the x terms have to be the same here, which gives $p=1$. And then c would be the negative of $-10+16p+3p^2$, which is equal to the negative of $-10+16+3$. So, that is the negative of 9, and so we get -9.

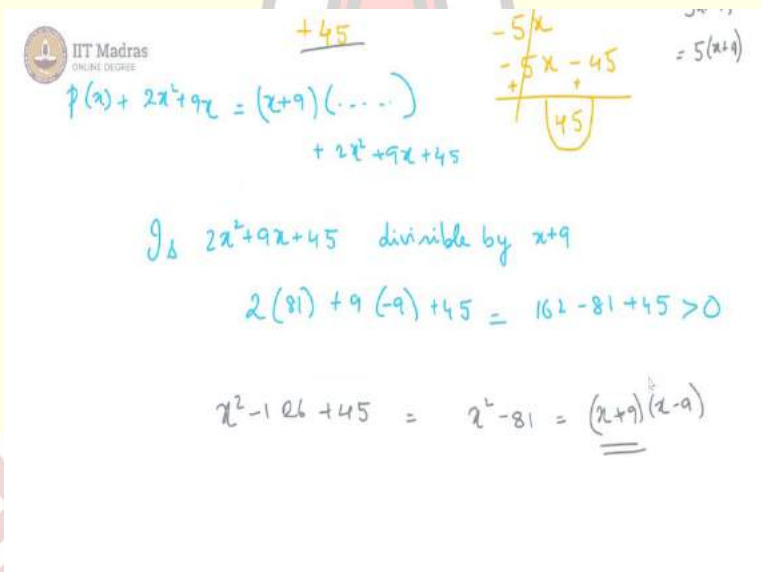
And that will indicate that none of the options are correct. So, this probably was supposed to be -9. We observed that option A and option C are in fact the same thing. So, one of this was probably supposed to be -9. Anyway, so our answer is that $p=1$, and $c=9$.

IIT Madras
ONLINE DEGREE

Mathematics for Data Science 1
Week 06 - Tutorial 03

(Refer Slide Time: 00:16)





Now, we have this problem, which of the following polynomials should be added to the polynomial $p\ (x)$ to make it divisible by $x +\ 9$. So, we need to recognize that it is not necessary that there is only one polynomial that you add, because since it is only divisibility, we can add a number of polynomials to $p(x)$ and make it divisible by $x +\ 9$. So, we have to check for each of these cases.

So let us see, or what we can additionally do is, we can look at the remainder that we get by dividing $p(x)$ with this and then see what to do with that remainder. So, if we did the division, now, we have $2x^3 + 23x^2 + 40x$ and we are dividing it with $x + 9$. So, start with $2x^2$, so we get $2x^3 + 18x^2$. So, this cancels off, this gives us $5x^2 + 40x$.

So, we do $+5x$ additionally, then we get $5x^2 + 45x$, so negative and negative so we are left with $-5x$ and then that gives us a -5 additionally here, so we have $-5x - 45$, therefore these two go off and we are left with 45 as our remainder. So, $p(x)$ is essentially $(x + 9)$ into the quotient +45. So, if we subtracted 45 from $p(x)$, we will get divisibility by $(x + 9)$.

So, B is necessarily correct. Let us look at what happens if we added A, if we added A, $p(x) + 2x^2 + 9x$ is some multiple of some product of $(x + 9)$, and some quadratic plus $2x^2 + 9x + 45$. So, unless $2x^2 + 9x + 45$ is divisible by $(x + 9)$, $p(x)$ would not be divisible by $(x + 9)$.

So, what we should really be checking is $2x^2 + 9x + 45$. Is it divisible by $(x + 9)$? And the direct way to check it is to substitute $x = -9$, so you will get $2 \times 81 + 9 \times -9 + 45 = 162 - 81 + 45$, which is greater than 0, it is not equal to 0. So, no, A does not give us divisibility by $(x + 9)$.

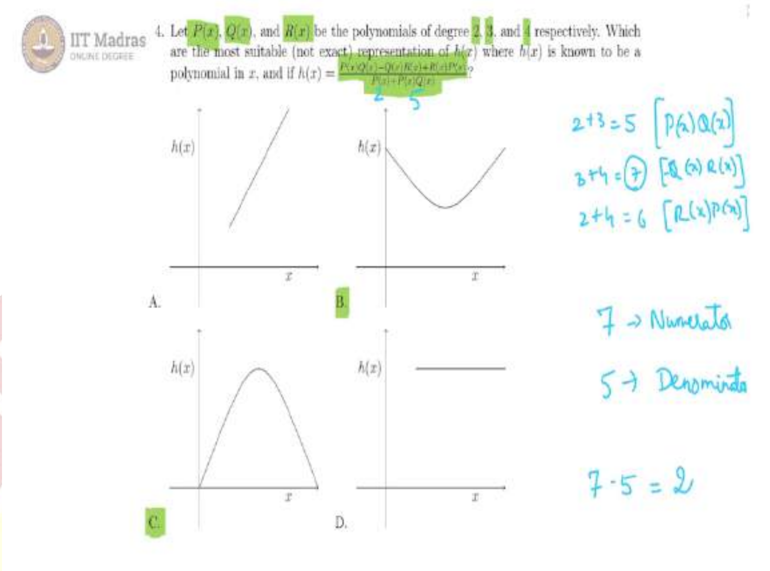
What happens if we added $5x$, we get $5x + 45$. So, we have this 45 remainder, so we are getting $5x + 45$, which is equal to $5(x + 9)$, which is directly divisible by $(x + 9)$. So, this is correct too, c is also correct. And what happens if we added $x^2 - 126$, then we would get $x^2 - 126$+45 as the additional part upside from $(x + 9)$ into that quadratic, so this is equal to $x^2 - 81$, which is equal to $(x + 9)(x - 9)$. So, $(x + 9)$ is dividing this particular polynomial. So, we can add $x^2 - 26x - 126$ also, and get divisibility by $(x + 9)$.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 06 - Tutorial 04

(Refer Slide Time: 0:15)



In this question we have 3 polynomials, $P(x), Q(x)$ and $R(x)$ and their degrees are given to be 2, 3 and 4 respectively. Which are the most suitable, although not necessarily exact representation of $h(x)$ where $h(x)$ is a polynomial in x and it is given as $\frac{P(x)\times Q(x)-Q(x)\times R(x)+R(x)\times P(x)}{P(x)+P(x)Q(x)}$. So, what we need to do here is to identify the degree of the numerator and the denominator.

Numerator degree $P(x)\times Q(x)$ will give $2+\ 3=5$ that would be the degree of $P(x)\times Q(x)$, the degrees will add up and when we look at $-Q(x)\times R(x)$, then again the degrees will add up which will give us $3+\ 4=7$, so this is from $-Q(x)\times R(x)$ and then $R(x)\times P(x)$ gives $2+4=6$. This is $R(x)\times P(x)$ degree.

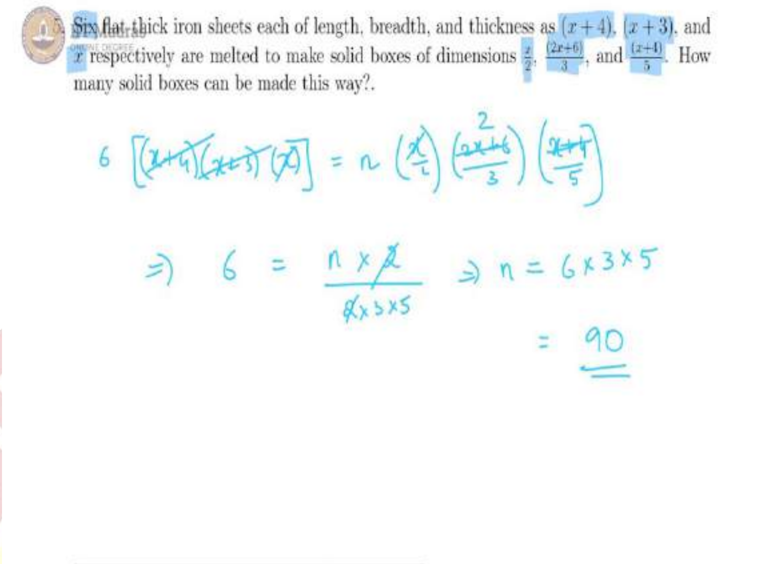
And in the denominator $P(x)$ anyway has degree of 2 and $P(x)\times Q(x)$ we have seen has degree of 5. So, since we are adding all these polynomials together, the degree of the entire numerator is the maximum which is 7. So, we have 7 as a degree of the numerator and 5 as the degree of the denominator. Since it is a division, the powers will have to subtract, so degree of $h(x)=\ 7-5=2$. So, $h(x)$ is a quadratic and that would indicate B and C are possibly the curves because these look like quadratic curves. A and D are definitely straight lines.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 06 - Tutorial 05

(Refer Slide Time: 0:14)



There are 6 flat, 6 of them, thick iron sheets each of length, breath and thickness $x + 4, x + 3$ and x respectively and they are melted to make solid boxes of dimensions $\frac{x}{2}, \frac{2x+6}{3}, \frac{x+4}{5}$. How many solid boxes can be made this way? So, basically the volume will have to be equal. So, first we find the volume of our 6 sheets put together that would be $6\,[(x + 4) \times (x + 3\,) \times x]$ and this would be equal to the volume of the solid boxes.

So, let us say there are n solid boxes and then the volume of each is $\frac{x}{2}, \frac{2x+6}{3}$ and $\frac{x+4}{5}$. So, now this x and this x cancels and this $x + 4$ and this numerator here cancels and $2x + 6$ is $(x + 3) \times 2$ so, this is one time and this is 2 times. So, what we get is $6 = \frac{n\times2}{2\times3\times5}$. So, 2 and 2 also cancels. This implies $n = 6 \times 3 \times 5$ and that is 90. So, you get 90 boxes overall.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 06 - Tutorial 06

(Refer Slide Time: 0:14)

6. Let x be the number of years since 2000 (i.e. $x = 0$ denotes the year 2000). The total amount generated (in Lakhs ₹) by selling a product is given by the function $T(x) = 5x^4 + 3x^3 + x^2 + x$. The different cost for that particular year are given in the table. What will the profit be for the particular year?

Cost type	Cost (in Lakhs ₹)
Purchase	$x^4 + x^3 + x^2$
Transportation	$x^3 + x^2 + x$
Miscellaneous	$0.5x^2 + 0.5x$

$$5x^4 + 3x^3 + \cancel{x^2} + x - (x^4 + x^3 + \cancel{x^2}) - (x^3 + x^2 + x) - (0.5x^2 + 0.5x)$$

$$4x^4 + x^3 - 1.5x^2 - 0.5x$$

In this question, let x be the number of years since the year 2000, so $x = 0$ denotes the year 2000. And the total amount generated in lakhs by selling a product is given by $T(x)$. So, this is a polynomial which has the variable as a number of years since 2000, and the different cost of the particular years are given here. So, purchase cost is this polynomial, transportation cost is this polynomial, miscellaneous cost is this polynomial.

So, we now have to find out the profit for that year. So, that would just be $T(x)$ minus all these cost. So, it is $5x^4 + 3x^3 + x^2 + x - (x^4 + x^3 + x^2) - (x^3 + x^2 + x) - (0.5x^2 + 0.5x)$. So, this would be the total profit and for that we now have to look at the each x power term.

So, x^4, there are 2 terms, $5x^4$ and $-x^4$. So, we get $4x^4$and x^3 there are 3 terms, $3x^3$, and $-x^3$ and $-x^3$ here. So, we get x^3 and x^2x square terms there are 4, there is this x^2 and then there is this $-x^2$ and another $-x^2$ and minus $-0.5x^2$.

So, that will give us minus 1.5 x square because this and this cancels off and then we get $-1.5x^2$. And lastly the x term there is x and $-\ x$ which cancels off and $-0.5\ x$. So, $-0.5\ x$. So, this would be the total profit for that year.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 06 - Tutorial 07

(Refer Slide Time: 0:14)

7. A company is planning to produce a product A through three available processes. Cost of production through 1^{st}, 2^{nd} and 3^{rd} processes are $M_1(x) = 100x^3 + 20x^2 + 10$, $M_2(x) = 20x^4 + 10x^2 - 20$ and $M_3(x) = x^3 + 20$ and the waste management cost for each of the processes are $W_1(x) = 0.01x^2 - 0.008x$, $W_2(x) = 0.01x^4 - 0.001x^3 + 0.001x^2$ and $W_3(x) = 0.01x^2$ respectively, where x is the cost of raw material per kg.

(a) What will be the effective manufacturing cost $E_1(x)$, $E_2(x)$, $E_3(x)$ for each of the processes?

(b) What will be the ratio of effective manufacturing cost of 1^{st} and 3^{rd} process when the cost of raw material per kg is ₹ 1?

(c) Which of the processes among M_1, M_2, and M_3 should the company choose when the cost of raw material per Kg is ₹ 10.

$E_1(x) = M_1(x) + W_1(x) = 100x^3 + 20.01x^2 - 0.008x + 10$

$E_2(x) = M_2(x) + W_2(x) = 20.01x^4 - 0.001x^3 + 10.001x^2 - 20$

$E_3(x) = M_3(x) + W_3(x) = x^3 + 0.01x^2 + 20$

In this question, we have a company which is producing a product A through 3 processes and the cost of production are given as $M_1(x)$, $M_2(x)$ and $M_3(x)$. These are the 3 cost of production. They have given us polynomials of x. So, what is x? x is the cost of raw material per kilo. And now they are also giving us the waste management cost as $W_1(x)$, $W_2(x)$ and $W_3(x)$. What will be the effective manufacturing cost?

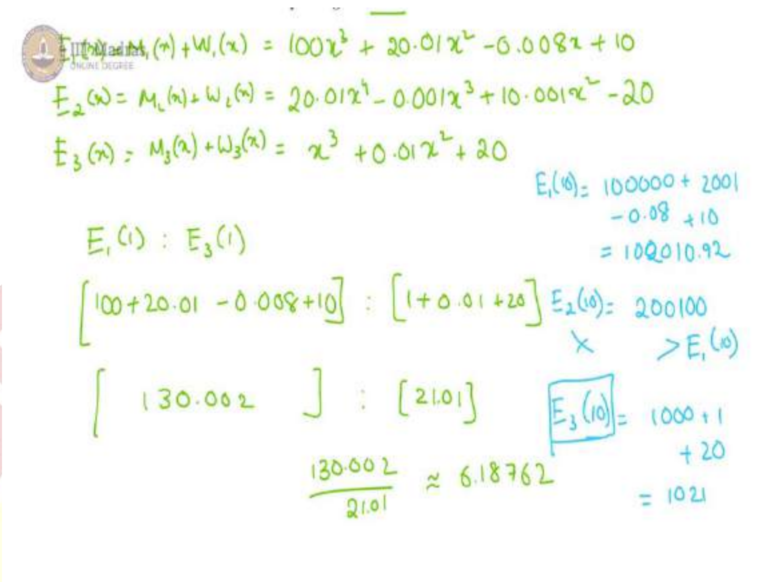
So, effective manufacturing cost simply has to be the sum of these. So, $E_1(x) = M_1(x) + W_1(x)$, so that is going to be so, M_1 is here, W_1 is here. So, M_1 would be $100x^3$ and there is no x^3 term in W_1, so we first write down $100x^3$, then there is an x^2 term here, $20x^2$, there is also an x^2 term here, 0.01. So, their sum will give us $20.01x^2$ and then there is no x term in M_1. There is an x term here, so you get $-0.008\,x$. So, this is also done. And then lastly we have the constant term which is $+10$. So, this is the effective manufacturing cost for process 1.

Likewise, process 2 would be $M_2(x) + W_2(x)$. So, here this is $M_2(x)$, this starts with an x^4 and W_2 also has an x^4 term. So, we have $20.01x^4$ plus there is no x^3 term in M_2, there is an x^3 term here, so there is no, this is not plus, we write $0.001x^3$, then the x^2 term there is a $10x^2$ here and a $0.01x^2$ here. So, $+ 10.001x^2$ and lastly there is a constant term and no constant term over there. So, this is -20. So, this is the E_2.

And then E_3 is going to be $M_3(x) + W_3(x)$ and in M_3 there is just 2 terms, the x^3 terms and a constant. W_3, there is only one term which is the x^2 term. So, we just write all of them together,

$x^3 + 0.01x^2 + 20$. So, this is the effective manufacturing cost for the third process and the three of them are given here.

(Refer Slide Time: 3:35)



Now, what is the ratio of effective manufacturing cost of first and third processes when the cost of raw material per kg is rupees 1? So, basically we are looking for the ratio $E_1(1)$: $E_3(1)$ which is then we just substitute 1 in the E_1 term, so we get $[100 + 20.01 - 0.008 + 10]$: $[1 + 0.01 + 20]$ is to, so we get 130.002 is to 21.01. So, we have to put this down as a number 130.002 divided by 21.01 is roughly 6.18762.

Then we have the third question, third part of this question which says, which asks which of the processes M_1, M_2 and M_3 should the company chose when the cost of raw material per kg is 10? So, the company should chose the cheapest process. So, we have to find out $E_1(10)$, $E_2(10)$ and $E_3(10)$. And then if we looked at that $E_1(10)$ =$100x^3$ is 1 lakh plus $20.01x^2$ is $2001 - 0.08 + 10$. This is then 102010.92.

Moving on then $E_2(10)$ is 200100. So, $x^4 = 0^4$ so, this is what we get and this is 200100, so $x^4 = 10^4$, so this is what we get and this is 200100. We see that the remaining smaller terms, x^3, x^2 and constant term have small coefficients as well, 0.001 and 10.001. So, they will not really impact the value very much. So, we know that this is already larger than $E_1(10)$, so we do not consider it.

Let us look at $E_3(10)$ which is then $1000 + 1 + 20$, so this is just 1021 rupees. So, $E_3(10)$ is the least which is why the company should chose the third process as their process when x is equal to 10 rupees per kilo.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

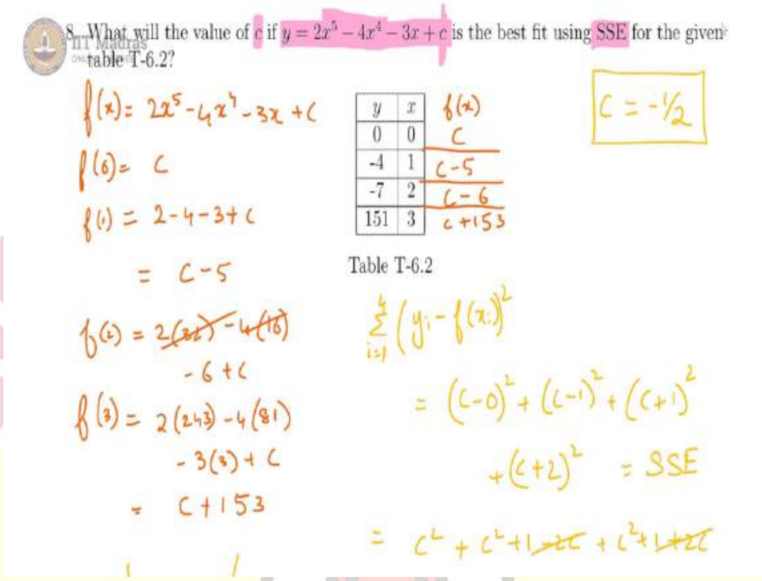
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 06 - Tutorial 08

(Refer Slide Time: 0:14)

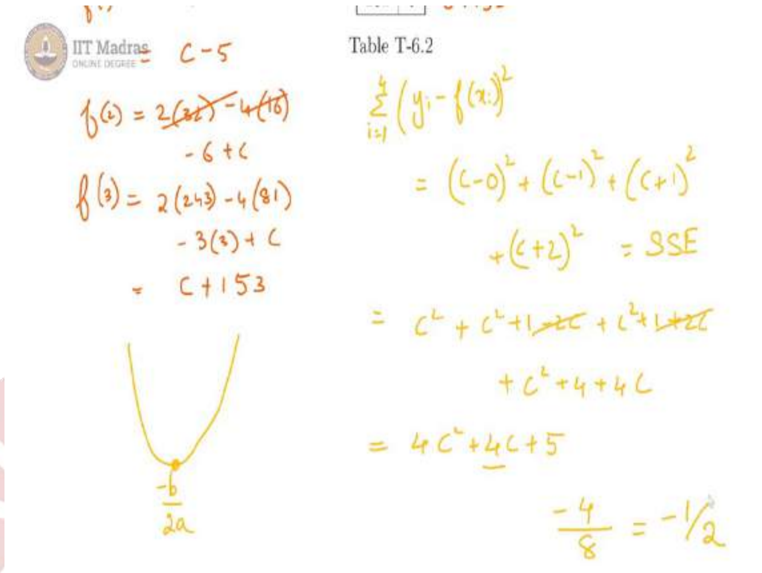


Our last question we are looking at the best fit for some data. So, this is the fit we have obtained a fifth degree polynomial for this data, these 4 points and they are asking what is the value of c, c is the constant term here. What is the value of c if this curve has to be the best fit using sum squared error? So, let us assume this curve is $f(x) = 2x^5 - 4x^4 - 3x + c$. So, we are going to have to also put up the $f(x)$ value, so $f(0)$ is then c because everything else is power of x, so $f(0)$=c, and then we have to look at $f(1)$ which is $2 - 4 - 3 + c$ that is equal to $c - 5$.

So, here this is $c - 5$ and then $f(2)$ is $2 \times 32 - 4 \times 16 - 6 + c$, now 2×32 is 64, 4×16 is 64, so these two cancel off, so you get 6 $c - 6$. And lastly, we have $f(3)$ which is $2 \times 243 - 4 \times 81 - 3 \times 3 + c$ so that gives us $c + 153$, so here it will be $c + 153$. So, for finding SSE we are going to have to do $f\ (x - y)$ or $(y - f(x))^2$.

So, $(y_i - f(x_i))^2$ and we are going to sum it from $i = 1$ to 4 and that gives us $(c - 0)^2 + (c - 1)^2 + (c - 1)^2 + (c + 1)^2 + (c + 2)^2$ So, this is the sum squared error.

(Refer Slide Time: 3:08)



And we get $c^2 + c^2 + 1 - 2c + c^2 + 1 + 2c + c^2 + 4 + 4c$, so this $-2\,c$ and this $+2\,c$ cancels off and we arrive at $4c^2 + 4c + 5$, this is our sum squared error it is a quadratic in c and for minimum and this is also an upward facing quadratic because the coefficient of $c^2 > 0$, so it will be a parabola like this and the minimum occurs at this point which is the vertex of the parabola and that we know is $\frac{-b}{2a}$, here $-b = -4$ and $a = 4$ so $2a = 8$ so you get $\frac{-1}{2}$, so for $c = \frac{-1}{2}$, we get the minimum sum squared error. Thank you.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1

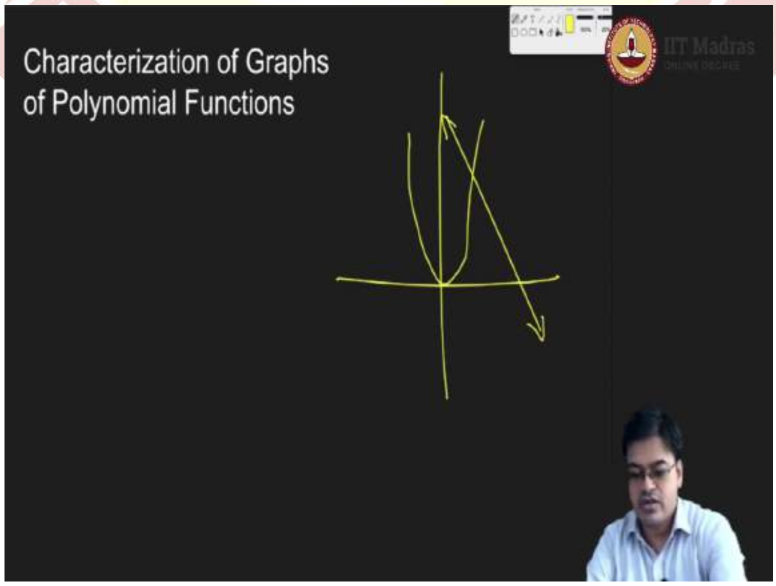
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 36
Graphs of Polynomials: Identification and Characterization

Hello friends, in this video, we will take up our next mission of about understanding the polynomials. This mission is given a graph of a function, whether we can identify the given function is a polynomial or not. If you have been given a polynomial equation, how will you put it on a graph paper?

So, the mission is twofold. First, If you have been given a graph of a function, you will identify whether this function is a polynomial or not. If yes, the, we will answer the second question that is can I derive the algebraic equation of this polynomial? The second part of the mission is we want to identify how the graph looks like if I have the equation of the polynomial. So, let us begin our journey about understanding the Graphs of the Polynomials.

(Refer Slide Time: 01:17)

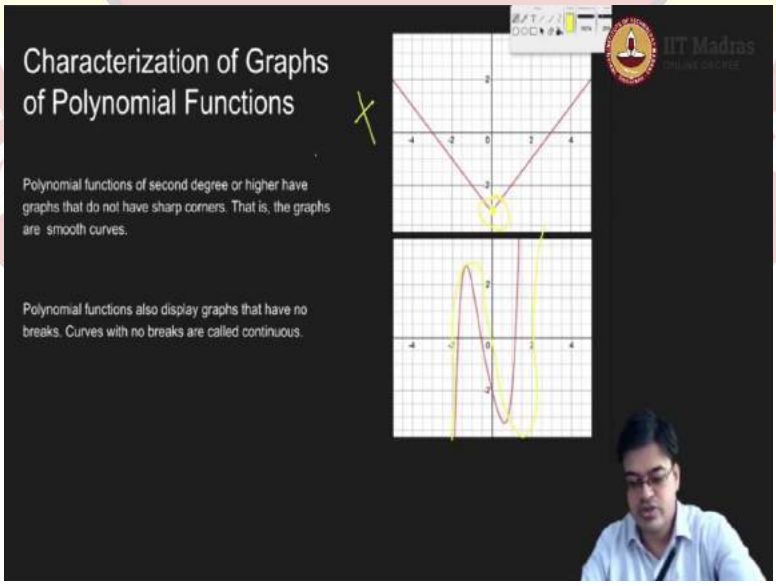


So, first of all let us recollect from our earlier experience that is linear functions and quadratic functions. If I am as linear functions and quadratic functions themselves are graphs of the functions. So, when I am plotting these two functions or when I am putting them on the graph paper, what happens? There you will never feel any abrupt jerk while drawing these functions. If you are trying to draw, you, for example, if you are trying to draw a line, then what you will do is you will simply draw a line, and then on graph paper. And there would not be any jerk for drawing the line.

In a similar manner if you are asked to plot a quadratic curve, you will find a axis of symmetry, and around the axis of symmetry you will do something like this, this is let us say this is the graph right, that means, the curve that you are trying to draw is has always been smooth. So, that one feature we can record in our mind. And say that the, if I have been given a polynomial function, the polynomial function must be smooth that means, I should be able to join the points effortlessly without having any jerk.

If there is any corner or edge in the graph then it better not be a polynomial function. Another thing is you can draw these graphs without lifting your pen; you can draw these graphs without lifting your pen that means, these graphs always are continuous.

(Refer Slide Time: 03:04)



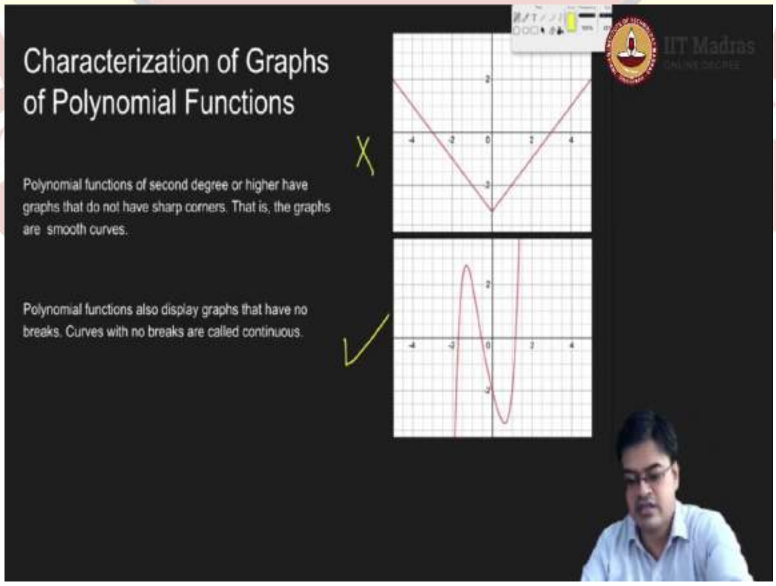
So, let us try to list these properties; first if you have polynomial of second degree or higher even a linear degree, the graphs do not have sharp corners that is the graphs are always smooth curves, this is a first feature that we will notice if I have been presented with graph of a function.

Second thing is polynomial functions always display graphs that have no breaks that is what I meant; the graph that I am drawing is always going to be continuous curve. Or you can say in better words, it is curves with no breaks are called continuous, and therefore the function itself will be continuous.

So, let us identify through two graphs. Let us have this graph. Now, is this a polynomial function? Does it satisfy the first criteria? That is it should be a smooth curve, is it a smooth curve? Yes, it is a smooth curve. But if you look at this point, then it had some sharp corner over here, this corner is very sharp. And therefore, I cannot qualify this as polynomial function; this is not a polynomial function.

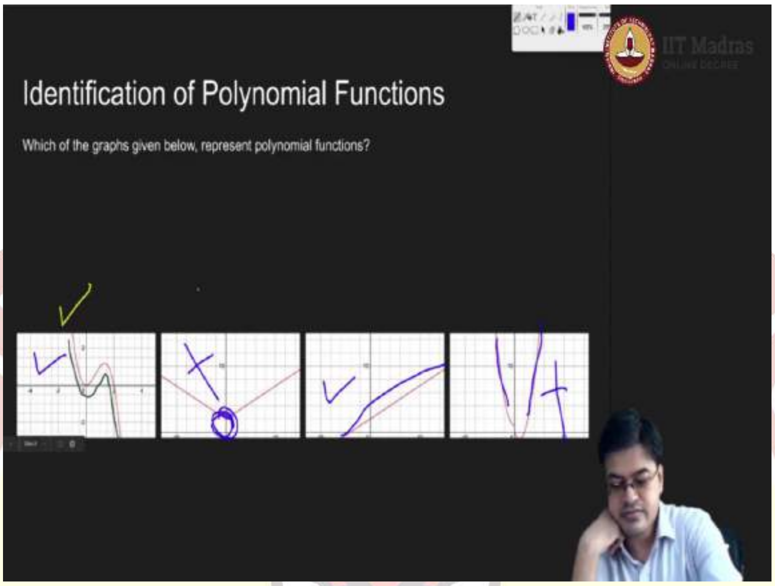
Let us have a look at the next graph which is this. Now, here I can use my free hand skill to draw a curve and I can actually find out how I can draw better curve. For example, if I start drawing this curve, then I can easily pass through this. So, you will all the transitions are very smooth, because the transition is very smooth I can easily identify this to be a polynomial function.

(Refer Slide Time: 05:06)



Therefore, this qualifies to be a polynomial function, whereas this does not qualified to be a polynomial function. Let us take a quick look at some other graphs and see whether they will qualify as polynomial function or not.

(Refer Slide Time: 05:22)



So, let us go ahead. And pose a question, which of the graphs given below represent polynomial functions? So, one by one I will unfold the graph, and we will argue for whether they are polynomial functions or not. This is the first graph. As I mentioned earlier, this also qualifies to be a polynomial function.

For example, if I want to draw a curve across this, I can easily draw a curve without lifting my pen and therefore, it qualifies to be continuous, and it has no breaks in between. So, therefore, it is continuous and it does not have any sharp edges, and the graph seems to be free hand. Therefore, it qualifies to be a polynomial function. So, my answer to this question is yes, this is a polynomial function ok.

Let us go ahead with the next graph ok. Now, what about this graph? Of course, we have argued for say similar graph in the earlier page that this graph does not seem to be a very neat graph, and it has a corner over here. This is the corner point of the graph. And therefore, this disqualifies to be a polynomial function; this is not a polynomial function. Again let me reiterate this was a valid polynomial function.

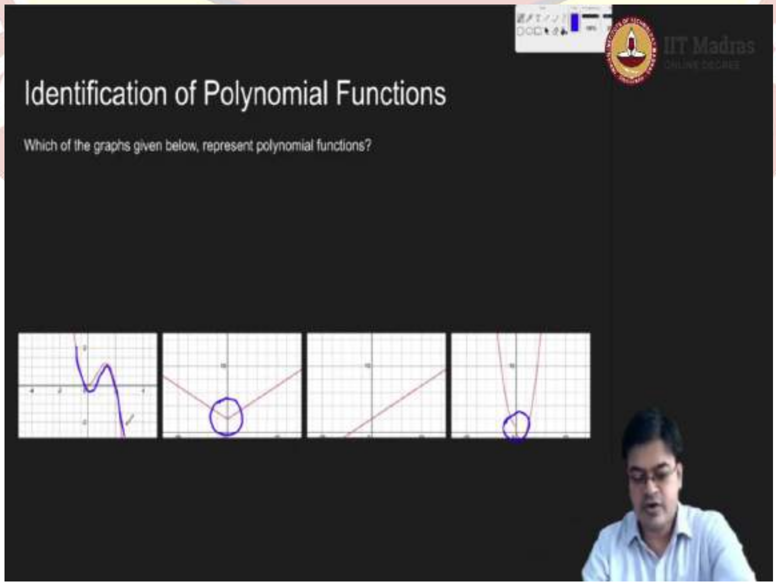
Next graph, let us look, let us try to see the next graph. This graph more or less seems to be a graph of a line, because it is a graph of a line. You can see this is also smooth. The transition is very smooth. So, again this will qualify as a polynomial function at least as far as the graph is visible. This is a graph of a line and it is a linear polynomial. So, it qualifies to be a polynomial function.

Let us go to the next graph ok. So, this graph is actually smooth. I can draw a curve over here, but at this point let me erase this graph; actually, you do it here. At this point, at this juncture, there is some problem. What is the problem? Over here, if I am drawing a curve over here, then I have to lift my pen come to a point 0 and then start drawing it.

So, this defeats the criteria that the graph should be continuous. Though the curves are very smooth, but at this point, this in this juncture, the problem is you cannot have a drawing without lifting your pen. Therefore, this will disqualify to be a polynomial function. This is not a polynomial function.

So, we have identified what is a polynomial function and what is not a polynomial function. Generally, whenever you have several ups and downs in the functions, we will estimate them or we will guess them to be a polynomial function if there is no corner as given in this graph, second graph to be precise.

(Refer Slide Time: 09:15)



And if there is no there should not be any corner of this kind and there should not be any discontinuity of this kind, ok. Then we can easily safely say that the given function is a polynomial function ok.

And if you are looking at the ups and downs, these up and down of a function, those are the typical features of polynomial functions. With this knowledge, we are ready to handle polynomial functions because now, if you have been given a function on a graph paper you can identify whether the given function is a polynomial function or not. The next important criteria for about polynomial functions like quadratic equations is our identity our ability to identify zeros; zeros of the function.

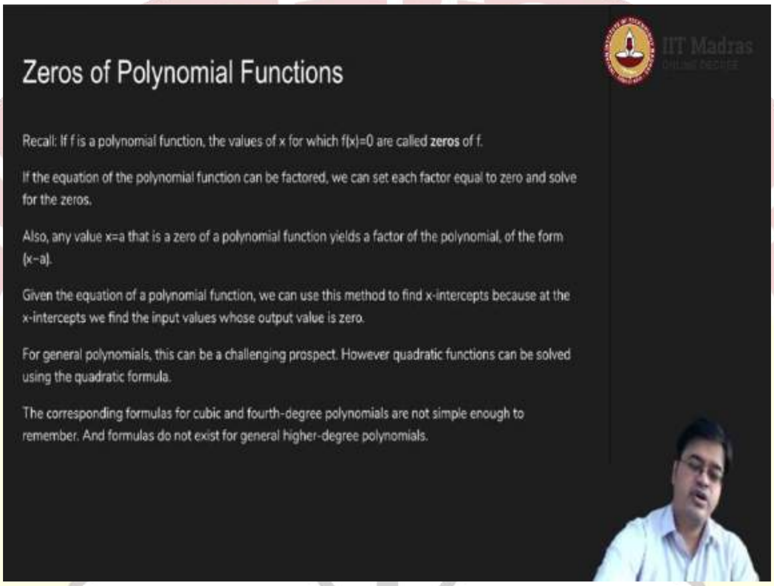
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 37
Zeroes of Polynomial Functions

(Refer Slide Time: 00:15)



So, let us focus on Zeros of Polynomial Functions. So, for clarity, let us recall what is zero of a polynomial function. If f is a polynomial function, then the values of x for which $f(x) = 0$ is called zero of f. A value x of for which $f(x) = 0$ is called zero of f.

Now, when we studied quadratic functions, we had several methods of identifying the zeros of the quadratic functions. For example, we actually tried to graph the quadratic function because we knew some techniques, we actually plotted set of ordered pairs on a graph paper and join the curve smoothly, then we identified it is crucial to identify axis of symmetry and around axis of symmetry you can plot and wherever it intersects x axis, we will call that as a zero of a function. This is how we identified quadratic zeros of quadratic functions.

Another way that we used which will be helpful here is; factoring the quadratic function into factors given a quadratic function identify the factors and write the polynomial into intercept form. If you are able to do that, then you have again identified zeros of the polynomial because when you said that quadratic function to be equal to zero and if it is in a factored form, all the coefficients corresponding to that factor will be all the numbers corresponding to that factor will be zeros of the polynomial function.

So, now, we will focus on the factoring component of polynomial functions. So, if the equation of the polynomial function can be factored, then we can set that each factor to be equal to 0 and solve for zeros; this is an important step. But it as we have seen in quadratic functions, this is not always possible.

In such case, if you put some random values, if you throw in some random values in the function and you get something like $x = a$, you will get the value to be 0 that is also helpful.

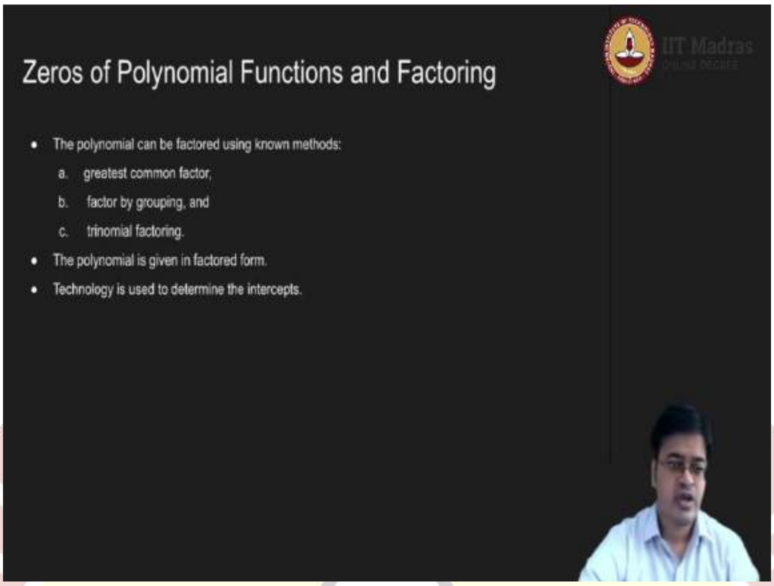
Then, you can guess that is a factor and you can use the previous video to divide the polynomial by $(x - a)$ which will give you the remainder term and that remainder, you can actually figure out whether you can; you can consider factoring for that remainder or not all these things are possible or the other factor it is not remainder sorry it is the other factor. So, these are some possible ways.

Up to quadratic equations, we had some easy ways out easy way out like given the equation of a quadratic function, we can use this method to find x intercept because x for x intercepts, we get zeros that is what I explained earlier also. So, you can find x intercepts and you will easily get this. You can use the similar technique of finding x intercepts for a general polynomial function also, but it is very difficult to plot a polynomial function ok.

Given a graph of a polynomial function and you have identified based on our previous criteria, you have identified that this is a polynomial function, you can guess what are the zeros of the polynomial function that way this statement helps. But, if you go for higher order polynomials that is general polynomials, this can become messy, it can be really challenging.

Quadratic equations can be easily solved using quadratic formula we have a solution for quadratic equations. But, the cubic and four-degree polynomials have some formulae which you may study in your tutorials, but they are not easy enough to remember. And, for higher degree polynomials, you do not have any idea of how to approach finding zeros of the polynomial functions, you have to go by trial and error method and whatever knowledge you have about square, quadratic, linear and cubic polynomials.

(Refer Slide Time: 05:00)



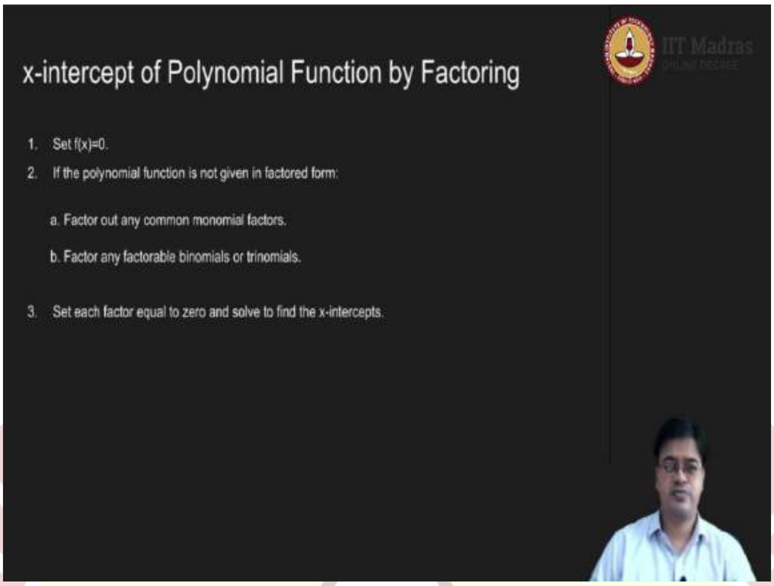
So, let us summarize what we have; what we have discussed just now. If I want to identify zeros of polynomial functions, the factoring technique is a crucial technique. So, what you can do is you can look at the polynomial and if you look at the polynomial, there is one easy way out that if you can identify the greatest common factor that is the greatest monomial that can be taken out common you can use that technique.

Once, if there is no such technique, if once that is available, the polynomial is more or less manageable, then you can use the technique of factor by grouping. So, you can create groups in that and see whether a anything is coming out common that is another technique. Another thing is you can instead of handling groups, you can decide to handle three terms at a time so, that is a trinomial factoring. This will be helpful when you have very high degree polynomial. So, these are the common methods for factoring the polynomials.

Once you can factor the polynomials each of them can be equated to 0 by writing a polynomial in a factored form. And then finally, if you are not very sure, then you can use some graphical tools which are available these days on computer or on the net one such tool is; Desmos which we are using in our presentations.

So, you can use those tools to determine the intercepts. In these tools basically, you will give of equation of a function and it will be graphed they will give the they will project the graph of a function right. So, this is our zeros of the polynomials and factoring play a crucial role.

(Refer Slide Time: 06:53)

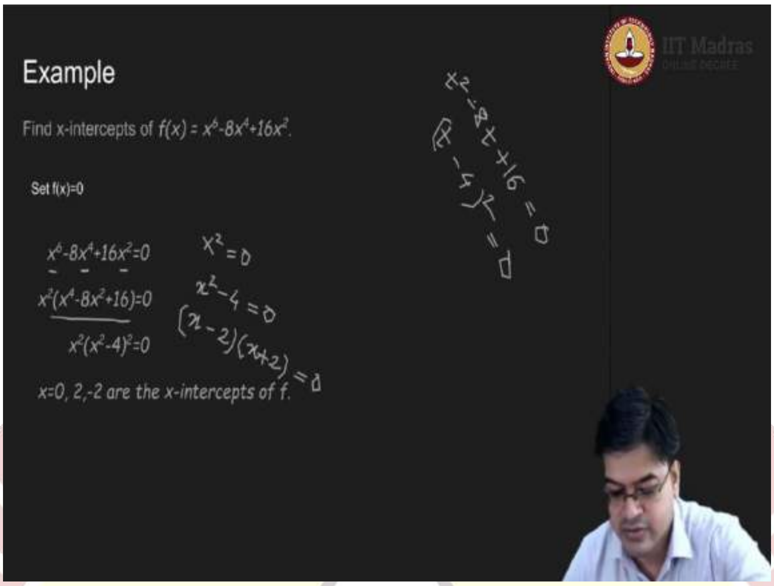


To understand this, let us see how to find x-intercept of a polynomial function by factoring. So, what we have discussed just now is we have set the equation that is $f(x) = 0$ in order to facilitate factoring $f(x) = 0$, then if the polynomial is given in factor form; factored form then equate each of them to be equal to 0 which we have seen for quadratic case also.

If it is not given in factor factored form, first in that you will look for is you take out some common monomial that is available in all the terms if that is that is there and you have taken out or if that is not there still you can go to the second step that is whatever at the rest of the terms you can factor them into factorable binomials or trinomials, you look for try to look for combinations which we have done successfully for quadratic equations well doing the factoring. So, you can do a similar thing over here.

And then finally, set each factor equal to zero that will give you the x-intercept. This is the; this is the strategy that we will follow for finding x-intercept of polynomial function by the method of factoring.

(Refer Slide Time: 08:21)



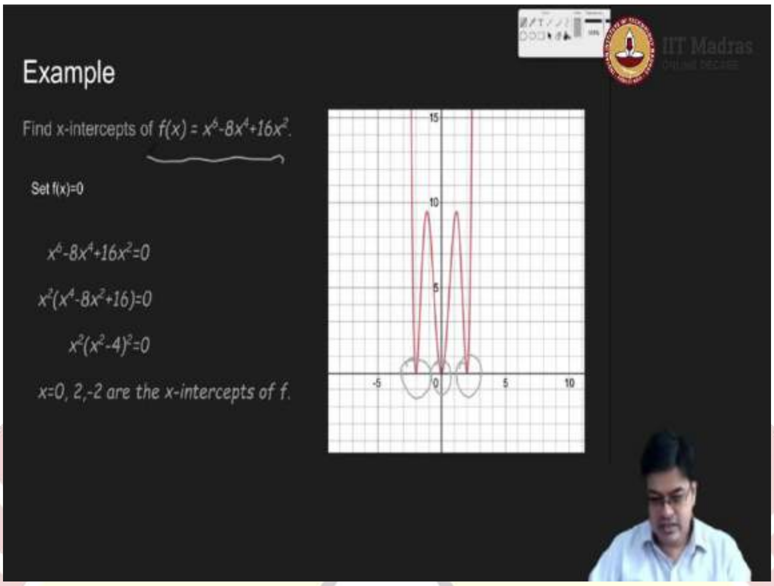
So, let us look at this example where we will follow the steps of the algorithm. So, the a question says; find x-intercepts of a function $x^6 - 8x^4 + 16x^2$. So, as per our algorithm or as per the steps given in the previous slide, I will set $f(x) = 0$ that is; $x^6 - 8x^4 + 16x^2 = 0$.

Now, you look at greatest common factor, a monomial that is common in all these terms that is x^2. So, what I will do is I will separate out this x^2, I have taken out this x^2 and now, you look at the other factor that is $x^4 - 8x^2 + 16$.

Now, this factor can be related to our quadratic equation of the form $t^2 - 8t + 16$. Can I factor this quadratic equation because there is no term corresponding to x^1 and there is no there are no odd terms essentially. So, I can use this and I can leverage the skill of quadratic equations to solve this equation and from quadratic equation point of view, I know this is $(t-4)^2 = 0$. So, instead of t here, it is x^2. So, that will give me $x^2 \times (x^2 - 4)^2 = 0$.

Now everything is looks in the form of x^2. So, what are the values of x? What are the feasible values of x? Those will be the x-intercepts. So, you can put x^2 is so, this will give me $x^2 = 0$ or $x^2 - 4 = 0$. So, $x^2 - 4$ can further be factored into $(x-2)(x+2) = 0$. And with this understanding, I can write $x = 0, 2, -2$ are the intercepts of f x-intercepts of f ok.

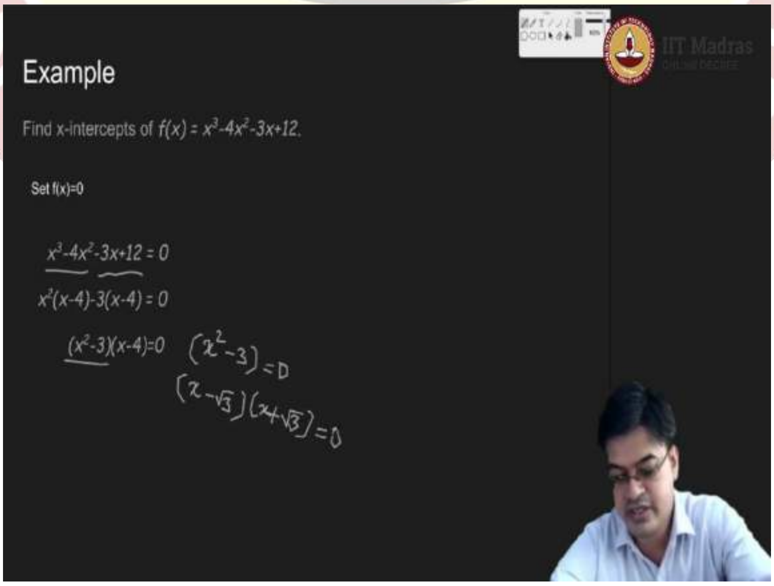
(Refer Slide Time: 11:09)



Now, as per the last step in the algorithm, you want to verify this result. How will you verify this result? Using the technology; so using Desmos, I have drawn this graph and you can verify that $x = -2$ which is here, $x = 0$ which is here and $x = 2$ which is here are all x-intercepts of a polynomial function given by these $f(x)$ ok. So, this is how we will identify x-intercepts.

Let us understand this strategy by looking at one more example.

(Refer Slide Time: 11:49)



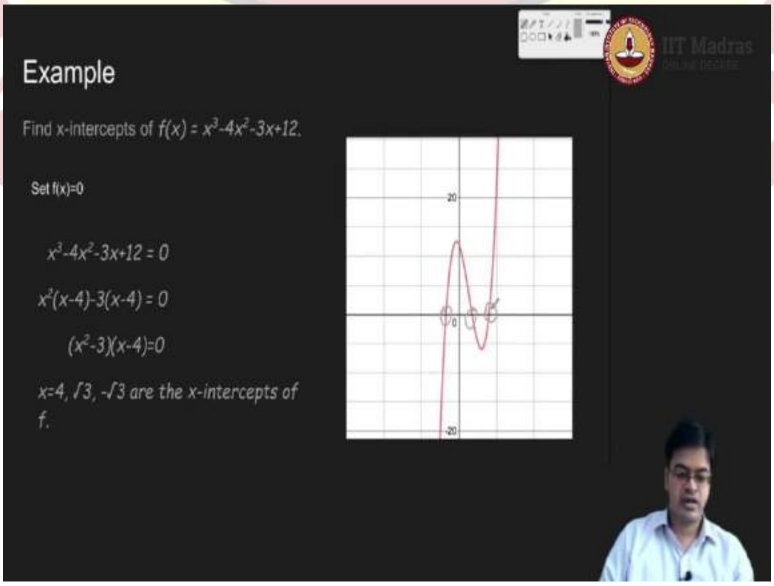
So, now here, we have been asked to find x-intercept of a polynomial function which is a cubic polynomial function $x^3 - 4x^2 - 3x + 12$ fine. So, as per our set up, this first step is set $f(x) = 0$. So, you have set $f(x) = 0$ that essentially gives me $x^3 - 4x^2 - 3x + 12 = 0$.

Then, the second a step if you have any common monomial, there is no common monomial because the last term is a constant term so, you cannot figure out a common monomial. Then is there any pattern? Can you look at two-two terms each binomials or trinomials because there are four terms, it is better to look at binomial terms.

So, if you look at the first two terms, you can see that you can throw out x^2 as a common thing, if you throw out x^2 as a common thing, then you will be stayed with $(x - 4)$ as a term as a one factor. And if you look at these two terms, then again if you take out 3 common - 3 common, then you will get $(x - 4)$. So, using the technique of binomial, binomials in this case, I am able to see this kind of factoring possible.

Good, that essentially means I can rewrite this expression as $(x^2 - 3)(x - 4) = 0$. Then, I want to solve this $(x^2 - 3)$ that is all is remaining which is a quadratic equation. So, you can easily solve using quadratic formula or a factoring, but here in this case, I know the factors so, that will be $(x - \sqrt{3})$ and $(x + \sqrt{3}) = 0$.

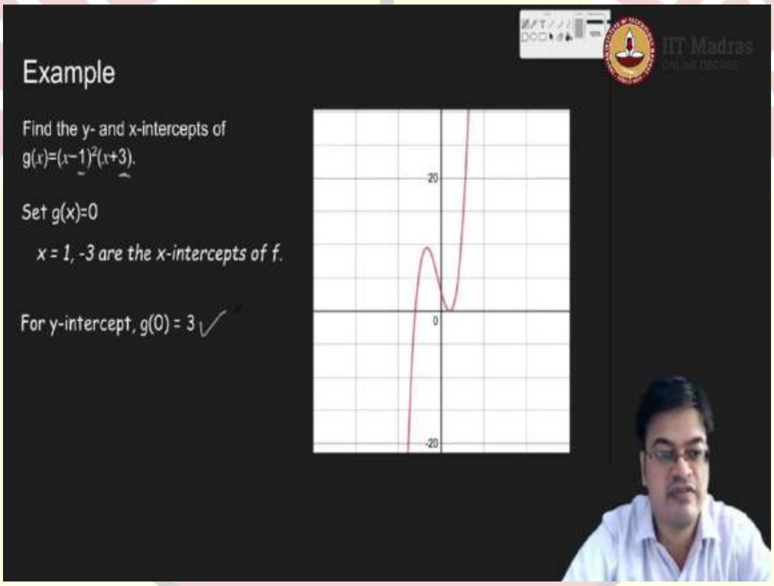
(Refer Slide Time: 14:14)



And therefore; therefore, the solution of this quadratic equation is well known that is $x = 4, +\sqrt{3}, -\sqrt{3}$ are the x-intercepts of the function. The final step is I want to verify using some technology or a graphing tool. This is the graph of a function.

So, in this case, you can easily verify there are three roots: first root this one which is a occurs, it occurs at , $-\sqrt{3}$, this one this is , $\sqrt{3}$, this one is 4. So, these are the four these are the three roots of a cubic polynomial. Roots or x-intercepts or a zeros of a cubic polynomial ok.

(Refer Slide Time: 15:05)



So, let us go ahead and see; let me remove this blocks. Another example where I want, I am interested in finding x-intercepts as well as y-intercepts of a polynomial function which is given in a factored form.

So, the polynomial is given in factored form. So, visually you will be able to guess the roots. So, as a standard set up, we will set $g(x) = 0$. Once you said $g(x) = 0$, it is very clear that $x = 1$ and $x = -3$ are the x-intercepts of f.

What about y-intercept? What is the y-intercept at all? So, y-intercept is where x is given to be 0. So, simply substitute $x = 0$ in the expression of $g(x)$, you will get $g(0)$, which is $0 \times (-1)$; the whole square that is 1+0+3 that is 3 so, 1×3 is 3 so, your $g(0) = 3$.

So, this is how you will figure out x-intercepts and y-intercepts of the function and this is the graph of that function. Using technology, I have verified.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

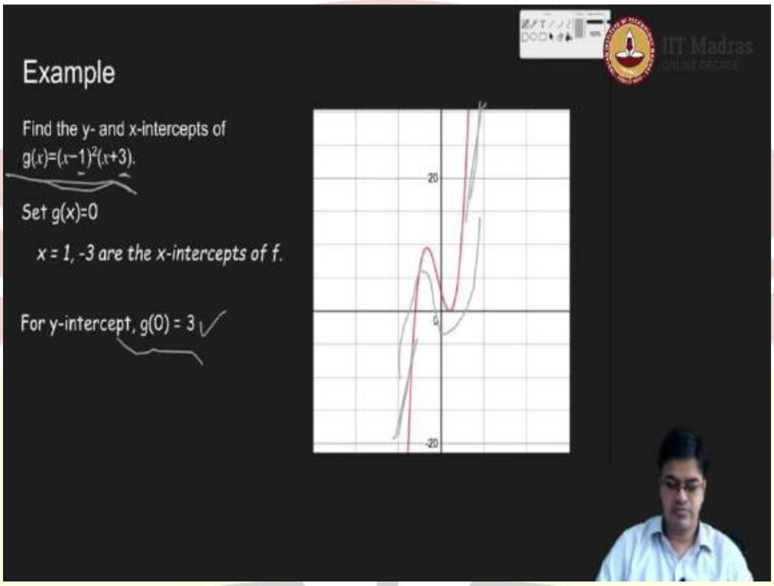
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 38
Graphs of Polynomials: Multiplicities

(Refer Slide Time: 00:14)



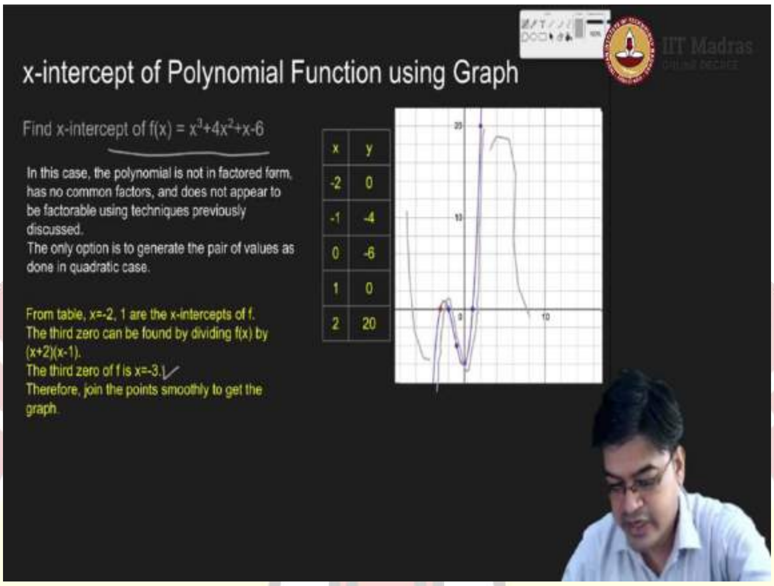
So, now, so far, we have mastered two skills; given a graph of a polynomial function, I know whether the given function is I can identify whether the given function is a polynomial function or not, ok. The second thing that we have seen is from algebraic expression of polynomial function whether it is in factored form or non-factored form, I have some set of rules or algorithm which will help me to identify, the roots of the polynomial or the zeros of the polynomial.

So, with this knowledge, can I explicitly write a polynomial function, or do I need to know something more about it? That is what the question that is troubling us. For example, the knowledge about x-intercepts in this case, and the knowledge of y-intercept, is this helping us to understand how the polynomial will look like?

For example, how will I decide the polynomial is going down from here, polynomial is going up from here, and it will stay going up forever, or when will this kind of shape come, the curve when will the rise and fall will happen, I do not know anything about this right now. What I know is simply the function should be smooth, yes, this is a smooth function.

But, from this graph can I write this equation? Seems to be difficult right now, but we will get the handle over it in due course of time.

(Refer Slide Time: 02:11)



So, let us now look at the x-intercepts or identifying the x-intercepts using graph. So, you have been given a polynomial function. You have used a technology to identify the graph, but still you are not convinced.

And, you want to try it by your hand. So, how will you do it? That is the question. So, if I want to find the x-intercept, the given polynomial is a cubic polynomial, fine. So, this polynomial is not given in a factored form; I cannot find greatest common factor that is not possible.

Then, if I want to find something which is like binomial thing that is $(x + 4)$, but the rest the other term is $(x - 6)$. So, I am not actually getting these two tricks done. So, there is no way in which, I can factor this polynomial. So, one crude way, if you do not know how to go about, is to plot the pair of values as we have done in quadratic case.

So, simply find out what are the function values at some points. So, these are some standard points, I have drawn them symmetrically 0, 1, 2, -1, and - 2. When I considered these two points, because the function is very nice, I accidentally came across two zeroes that are - 2 and 1, good.

So, -2 and 1 are the x-intercepts of f which is clear from the table. Now, can I use this knowledge to find the third zero? The answer is yes. And, we know the long division. So, what you do is you consider $(x + 2)$ as one factor and $(x - 1)$ as another factor. You multiply $(x + 2)(x - 1)$ and treat that as a divisor, and take $f(x)$ as a dividend, and do the long division.

If you do the long division; you will get, you may pause the video and try for yourself; otherwise you will get the third 0 to be equal to $x = -3$; that is $x + 3$ is another factor. And, this is a cubic polynomial, so it cannot have any other factor, can I have at most 3 roots. So, you got this $(x + 3)$, $(x + 2)$, and $(x - 1)$ as the factored form of this equation. Then, it is easy to plot the equation along with this table of values.

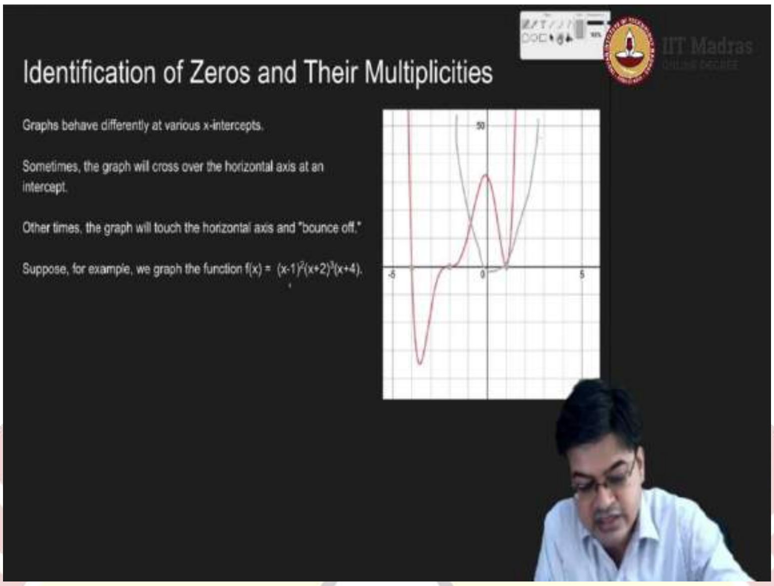
So, what you will do is, you will simply put up the points, you will simply put up the point. So, over here I know something and I know I have figured out the third root to be equal to $x = -3$; therefore, I can put that point as well.

So, this is $x = -3$ and, how to draw a line passing through; so, the next step is joining the line. So, you can draw join a smooth line passing through these points, then at this point it will turn up; at this point it will turn up, but to connect to this point, it has to go down, and then I do not have any idea. So, right now I can draw only up to this, right.

We will analyze further and see the cubic polynomial cannot turn more than two times, so that we will that we will come later. But, right now I can draw it this way. So, let us see what our graphical tool gives us. Yes, where the previous image and this image are slightly perturb but they are exactly matching. So, this is how the behaviour of the function will be.

So, now we have slightly better edge over drawing the graph of a function, when we have been given a equation of this form, but still I do not know why this is not turning up or why this is not coming down, I need to understand these things in a better way by using some analytical tools. For a moment, we got the correct graph.

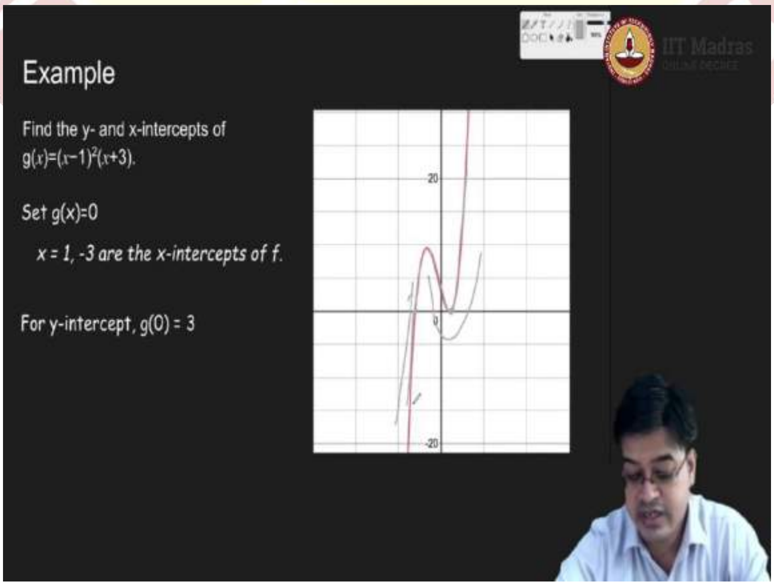
(Refer Slide Time: 06:52)



So, let us move ahead and try to see the behaviour of the graphs around the intercepts. For that, it is important to know the multiplicities of factor. So, in particular, graphs behave differently for at various x-intercepts.

We can go back to the previous case, where everything is of linear order that is $x+\ 3$, $x+2$ and $x+1$, the graph was behaving like this. If you go further back; why would this graph behave like this? Over here, when $x=-3$ is a factor, the graph the graph was actually like a straight line and over here it turned around.

(Refer Slide Time: 07:33)



So, why should it turn around the factor? So, for example, $x+\ 3$ is one factor that is -3, and here $x-1$ is one factor. But, for this factor it turned around; and for this factor it crossed, it cross the x-axis. So, why is this happening? So, I need to have a deep understanding of this. For that, we will discuss the next that is what we will discuss in the next slide. So, is it related to the function being appearing the factor appearing multiple times? That is what we will try to see.

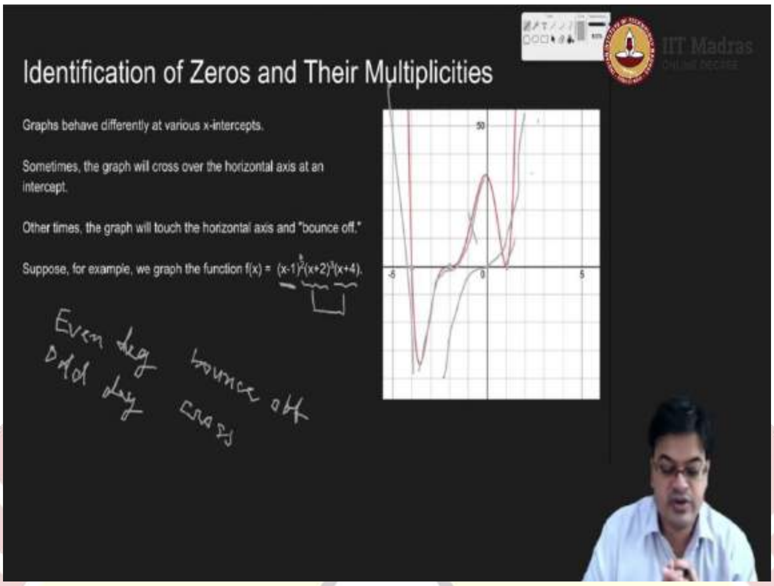
So, as mentioned as shown in the earlier slide, that the graph can cross over the horizontal axis or it may bounce off; that means, it will touch and go up that is tangential to that axis. So, why is this happening at x-intercept? That, so in that case let for that making the understanding clear we will write a polynomial in a factored form which is $(x-1)^2(x+2)^3(x+4)$, right. And, let us draw that polynomial using technology or graphing tool, ok.

So, now some crucial things, let us identify the factors first; $x-\ 1$ that is $x\ =\ 1$, this is the factor that we are talking about. Then, $x=-\ 2$, this is the factor that we are talking about and $x+4$ this is the factor that we are talking about -4.

Now, at these points, what is happening, what exactly is happening at these points? So, when I consider the factor $(x-1)^2$, because it is quadratic and if I recollect the graph of a quadratic function, it behaved some it is not to the scale, but it behaved something like this, right. It will never cross x-axis.

So, a similar feature is visible over here, when I consider the graph of this function. So, if I consider a graph of this function, because $x-\ 1$ is coming twice; it is square $(x-1)^2$ is there; so, what I am getting is the behaviour is of the quadratic nature, ok.

(Refer Slide Time: 10:39)

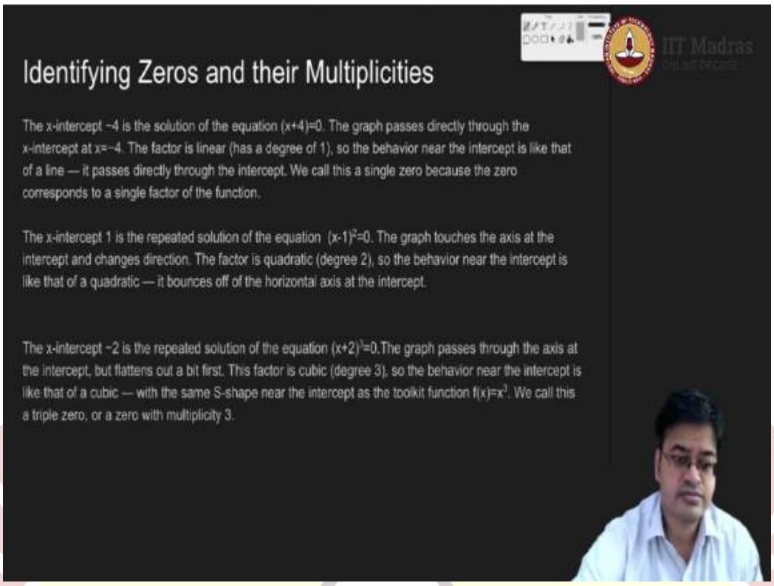


Now, let us look at that $(x+2)^3$. What is a graph of $(x)^3$? A graph of $(x)^3$ is somewhat like this; it crosses x-axis y is equal to x cube, it crosses x-axis. So, now that behaviour is evident when I consider that instead of x, I consider $(x+2)^3$ that behaviour is evident over here. It actually cuts and crosses x-axis.

And, if you look at the third factor that is $x+\ 4$ which is $x=-4$, it is behaving like a straight line that is also. So, what is happening here? I have two things; one and this one. So, in these both cases we have odd degree polynomials and, the odd degree polynomials as we know actually cross x-axis. And, in this case I have an even degree polynomial which is actually bouncing off the x-axis, this is a typical feature.

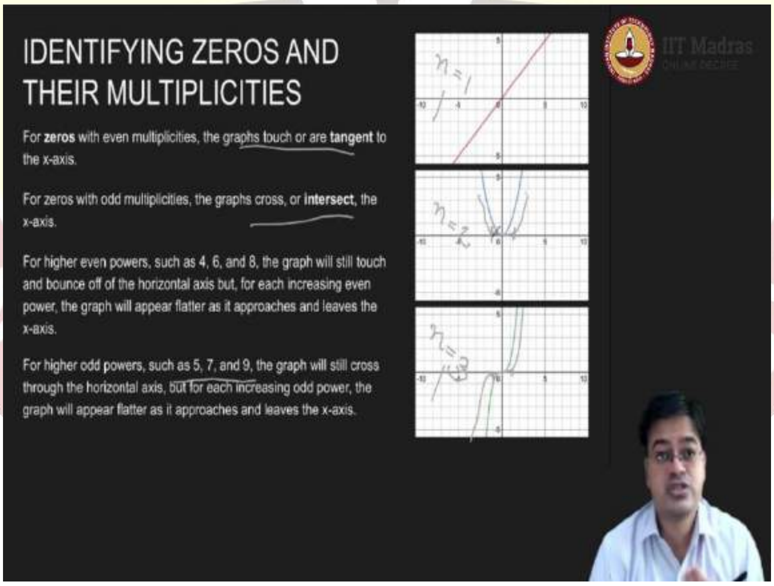
So, if I have even degree what we are saying is, if the polynomial is a, the factor is of even degree, then it will bounce off; that means, it will not cross x-axis but, if the polynomial as odd degree, then it will actually cross x-axis. So, these are the two typical features that we will employ while for plotting the functions which are of polynomial nature once we understand the factors.

(Refer Slide Time: 12:25)



So, in the next slide, I have given a general description of these factors, you can go through these slides later, but it is essentially the same that I have said just now, ok.

(Refer Slide Time: 12:38)



So, now for identifying zeros and their multiplicity. What do I mean by multiplicity? How often that factor is appearing. In the previous case the factor 1 was appearing twice, factor minus 2 was appearing thrice, and other factor was appearing only ones. So, if I want to identify the zeros and their multiplicities, I should look at the shapes of the curves. For example, if you look at the first graph, here the degree of the polynomial $n = 1$, here $n = 2$, $y = x^2$ this is, and here $n = 3$.

As mentioned earlier, it is more; it is more evident now, that when I have odd degrees, the curve actually passes through x-axis, when I have even degrees we can draw $y = x^4$, but this will be slightly broad and it will cut the x-axis, it will be somewhat like this. Let us not get into that. But, for odd degrees you will get something of this form, or even degrees you will get the bouncing off pattern and for odd degrees you will get a pattern which is actually crossing, one minute.

Let me reiterate this that; this is very as this is very important. If you have an odd degree polynomial, then you are almost sure you are sure to cross x-axis. If you have even degree polynomial, you will never cross x-axis at that point, you will simply bounce off from x-axis, ok. So, that gives us some more clarity. So, if the zeros of the polynomial or the factor has even multiplicities, the graph will touch or is tangent to x-axis or zeros with odd multiplicities, the graphs cross or intersect the x-axis.

Now, if you look at the even powers which are 4, 6 and 8, how will you guess, what is the strength of the power? So, in that case the graph will still touch x-axis it will bounce off, but which with each increasing even power it will appear to be flatter and flatter while approaching the zero and leaving from the zero. For example, the base will broaden; in this case, it will be something like this if it is x^4; x^6 further flattening.

If in a similar manner when you have odd powers like 5, 7 and then the graph will appear to be more flat over here, and while leaving also it will leave slowly and then it will decay at very fast rate. So, this is the typical feature from the bulge at these intervals you can actually guess the multiplicity of a polynomial. That is the importance of this slide.

So, now we have added one more weapon in our arsenal that is we will identify the multiplicities of the zeros. First we identify zeros. So, at the step zero is we identified given a function where whether it is a polynomial function or not. Then, we identified the x-intercept of the function that is zeros of the functions or roots of the functions.

After identifying roots of the functions, roots how many times repeated that is what, we have identified here in this by using the graphical tools. This is quite powerful. And, you can use it more often to understand the polynomial function. When you will actually solve some problems on identifying the polynomial functions, you will get a better hold of it.

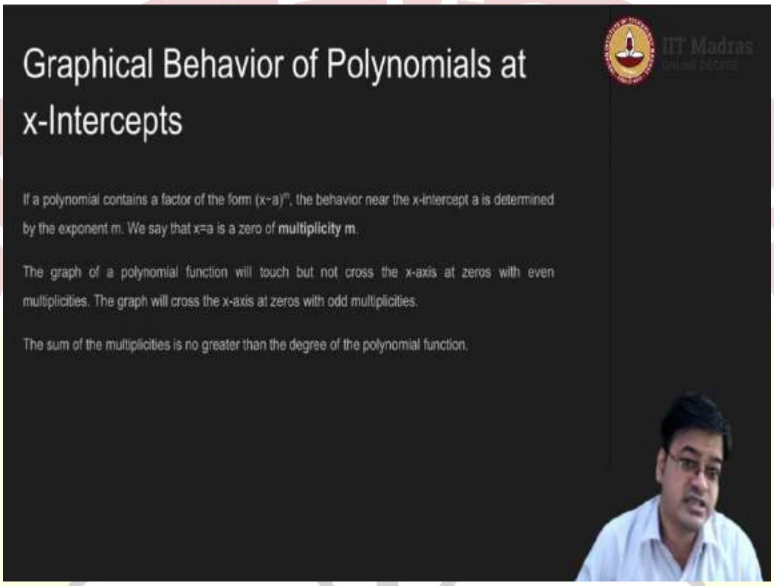
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 39
Graphs of Polynomials: Behavior at X-intercepts

(Refer Slide Time: 00:14)



So, let us go ahead and look at the Graphical Behavior of Polynomials at X-Intercepts. So, in particular if the given polynomial has a factor of the form $(x-a)^m$, this m is called the multiplicity of the polynomial. And you will say $x = a$, is a zero of a polynomial f with multiplicity m. This is to fix the terminology.

Now, the graph of a polynomial function will touch, but not cross x axis at zeros with even multiplicities and the graph will cross x axis at zeros with odd multiplicity. We have iterated it enough number of times.

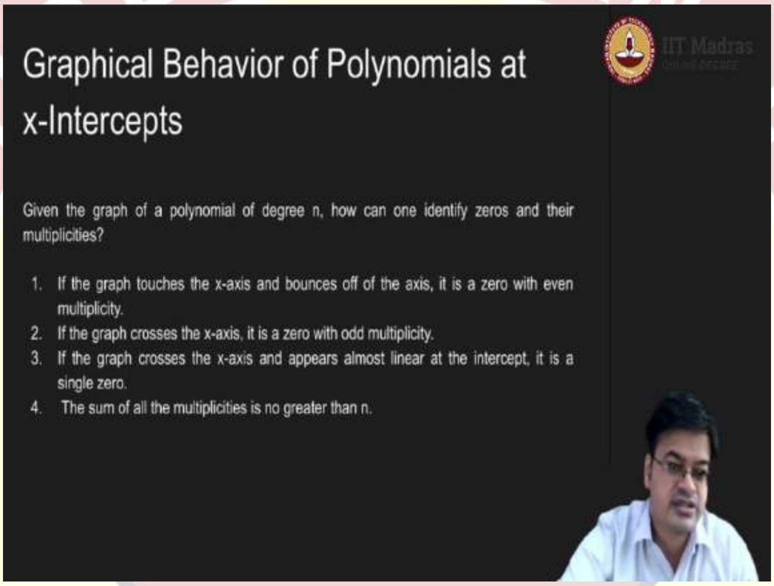
Also, one important thing is the degree of the polynomial cannot exceed the sum of the multiplicities, or the sum of the multiplicities is always less than or equal to the degree of the polynomial function, this is quite common sense right. If it exceeds, if it actually exceeds the polynomial degree then it is a polynomial of higher degree.

So, then now why it is not equal to, can be one question. The sum of multiplicities you will say will always be equal to the degree of the polynomial function. For that we need to understand that all the roots all the zeros of the polynomials or the only way we are identifying zeros of the polynomials is by identifying the x intercepts.

So, all x intercepts are real roots of the polynomials, but as in the quadratic case we have seen that some of the polynomials, some of the quadratic equations do not have real roots. In such cases the x intercepts are not visible.

I will demonstrate it further through some examples.

(Refer Slide Time: 02:35)

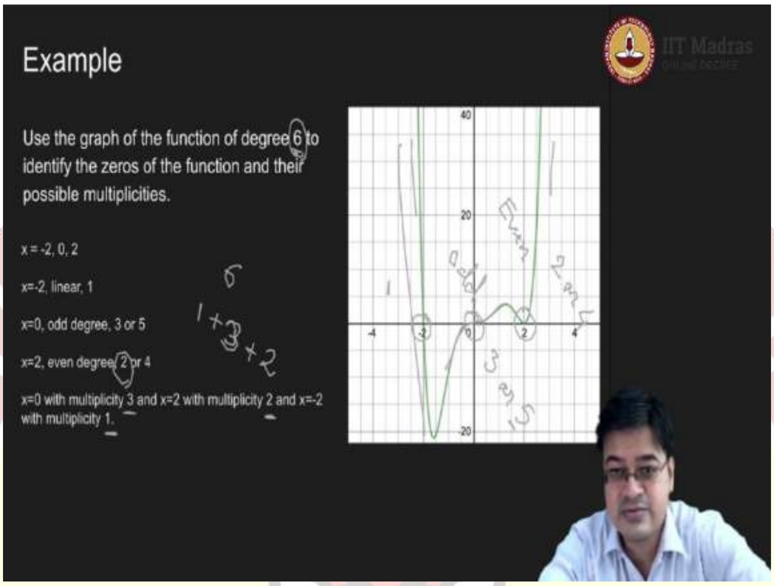


So, let us try to see. So, given a polynomial of degree n, that is a graph of a polynomial of a degree n. We want to identify zeros and their multiplicities; this is our goal. So, you are you have been told that this polynomial is of degree n and this is the graph of the polynomial, how, what will you do about it?

So, in that you will look at all the coordinates, where the graph touches x-axis, you take them. If the graph touches x-axis and bounces off the x-axis then, it is a zero with even multiplicity. If the graph actually crosses x axis, it is a zero with odd multiplicity. And finally, when you will conclude you have to take care that because the polynomial is of degree n, the sum of the multiplicities should never exceed the actual degree of the polynomial.

Another thing if the graph crosses x axis and appears almost linear at the intercept, then it is a single order; that means, it appeared only once; it is a linear function. And finally, that is what I explained the sum of the multiplicities is no greater than this fine.

(Refer Slide Time: 04:10)



So, let us go ahead and see some examples and let us see whether we can apply these principles in action. So, use the graph of a function of degree 6. So, it is a degree 6 polynomial, which is given to you and this is the graph, wonderful.

So, now you can easily see -2, 0 and 2. So, x intercepts are -2, 0 and 2 fine. Then if you, let us start from left; so, at x is equal to minus 2, how is the behavior? It is more like a straight line, it is more like a straight line. So, at minus 2, I feel the behavior is linear or it is a onetime event.

At 0, what is happening is; it is having this S shape, somewhat twisted S shape. And that is indicative of odd degree that is indicative of odd degree and degree can be 3 or 5 or 7 or 9 I do not know right, but I have been given that the polynomial is of order 6. So, at most it can have a degree 5 right, the multiplicity 5.

Now, look at this particular junction, it actually bounces off the x axis. So, this is a typical trait of even degree polynomial. So, what can be the degree? It can be 2 or 4 right; so it can be 2 or 4 ok. Now, we need to collate this information. So, x intercept $x = -2$ is linear, there is no doubt about it. $x = 0$, you have odd degree 3 or 5 is the degree and $x = 2$ it is even degree.

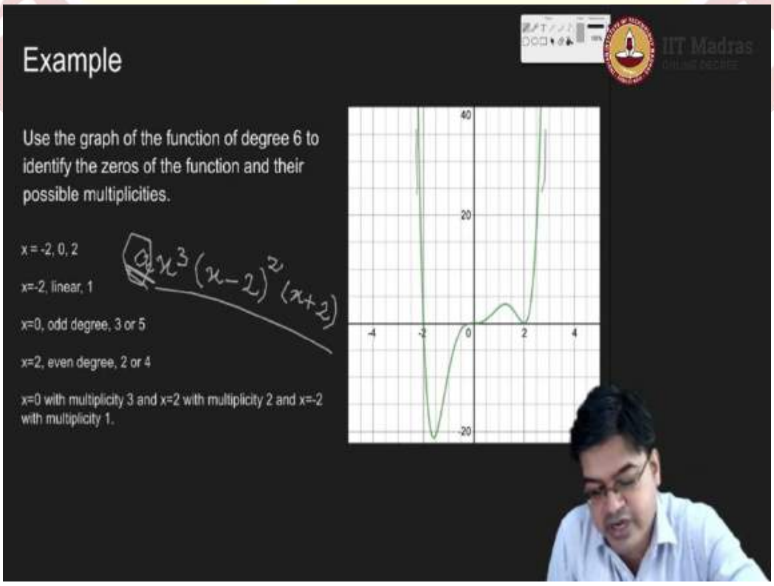
Now, together sum of the degree should be equal to 6, of which this 1 is fixed. So, now, I can assign 3 or 5. If I assign 5 then $1 + 5 = 6$ and if $1 + 5 = 6$ that essentially means; there is no root of the form $x = 2$ that is not possible. So, I have to assign 3 over here. Once I assign 3 over here, then I do not have any other choice, but to choose 2 over here.

So, therefore, I have identified the multiplicities of the factors like $x = -2$ will have a multiplicity of 1, $x = 2$ has multiplicity of 2 and $x = 0$ has multiplicity of 3. So, this is how we will identify the factors, and identify the zeros of the functions, and identify their multiplicities. This is much better for drawing the graph of a function.

Still, I have not answered a question that why this should go to infinity and why this also should go to infinity; those things are not clear, but we will come to them later. Right now, we have much better understanding about zeros of the polynomial functions and their multiplicities and how we can use them to understand the function.

Now, another interesting thing that you can ask yourself is ok, I have seen this function and I know the multiplicities of zeros and everything. Can I use this knowledge to actually tell what is the equation of the function or not. We can try our hand on it, but we may not be able to because $x = 0$ is with multiplicities 3.

(Refer Slide Time: 08:39)



That gives me x^3. $x = 2$ with multiplicity 2 that gives me $(x - 2)$ the whole square and $x = -2$ with multiplicity 1 which will give me $(x + 2)$ ok.

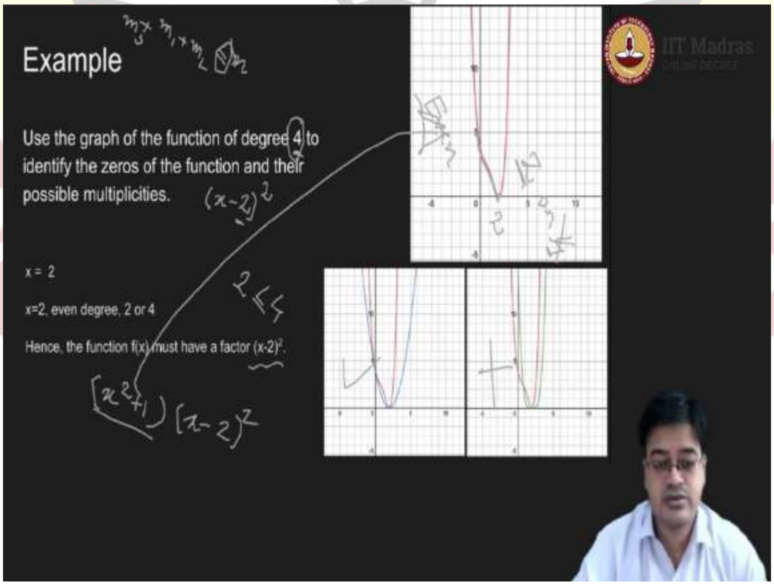
So, now I can say that this is the polynomial function that is graphed here, but how will I verify? So, for that I need something which is non-zero. So, you can choose some point and check whether this is there, but there is a catch over here.

It need not match the values that are given here. So, what we will do is though the factors are correct the polynomial is of degree 6 here the degree is 6, we will put some unknown a over here and we will determine this a by putting the actual values.

We will come to it later when we have better understanding about the behavior of this kind, but right now you keep this in this point in mind; that we do not know this a. So, we cannot actually give the exact equation of the polynomial with reference to these points.

Though we know the form of the polynomial, but we do not have accuracy up to the exact matching on the coordinate plane with numbers. Just remember this point and this juncture and let us go ahead and do some other problems.

(Refer Slide Time: 10:16)



So, in the next problem is of similar nature. I have a polynomial of degree 4 and I have been projected with a graph of this polynomial function.

Now, this is an interesting example because, you have a polynomial of degree 4. In earlier case we had a polynomial of degree 6 and all 6 roots were actually visible. In this case I have a polynomial of degree 4, but there is only one root that is visible or one 0 that is visible.

So, what is that number? It is 2 ok. And based on our understanding of the algorithms what we know is the graph actually bounces off; the graph actually bounces off the x axis which is 2. So, this is a typical trait of an even degree polynomial, or even degree even multiplicity. So, in this case I can say it can have a multiplicity of 2 or 4. I cannot exceed beyond 4 because the given polynomial has degree 4 only ok.

So, now because the given polynomial has degree 4, it is safe to assume that this polynomial is of degree 4 right. But if this polynomial is of degree 4 you can see there is a perturbation of the shape over here. It is not the shape of a polynomial of degree 4, for example; y=x^4 will not be in this form. So, I can rule out the degree 4 constraint.

Therefore, I do not have any other choice, but to say that the polynomial is of degree 2, the multiplicity of this particular factor 0 is 2; that means, I have a factor of the form $(x-2)^2$ the whole square that is all I can say in this case. Yeah, $x = 2$, is of even degree 2 or 4 and hence the function based on the reasoning that I have given it must have a factor of $(x-2)^4$.

Let us understand this graphically as well. This is a function, the blue line over here, is actually $(x-2)^2$. As you can see it passes very closely to through the function and the other line is $(x-2)^4$, other graph the green line is $(x-2)^4$.

Now, if you look at graph of $(x-2)^4$ closely there is no possibility of changing the shape. You can scale with that unknown a, but you cannot change the shape of the function. So, this graph is actually ruled out. So, graphically we have understood why we are ruling out and this graph is not ruled out. So, this is somewhat familiar. So, it will have some factor of this form.

Now, one exercise for you is this graph actually though it is a polynomial of degree 4, the way we have constructed is we have multiplied (x^2+1) with $(x-2)^2$.

And if you look at this particular factor x square plus 1 because you, now what you can do is you can actually consider this $(x-2)^2$ as one factor. And you can see whether this x^2 you will get the same graph by multiplying this. The beauty of this example is that this (x^2+1) has no real roots.

Therefore, the degree though the degree of the polynomial is 4, I was not able to find the two missing roots; those are not in the real domain. They are in the complex domain. So, those things are there that is why the question that the sum of the multiplicities will always be less than or equal to the degree of the polynomial.

So, let us say there are three factors m_1, m_2, m_3 are the multiplicities. Then $m_1 + m_2 + m_3 \leq n$; that answers this question. So, in this case I was actually able to find for these real roots only 2. So, 2 is less than or equal to 4. So, that is why the; this is the example, why I cannot say that the sum of the multiplicities is always equal to the degree of the polynomial.

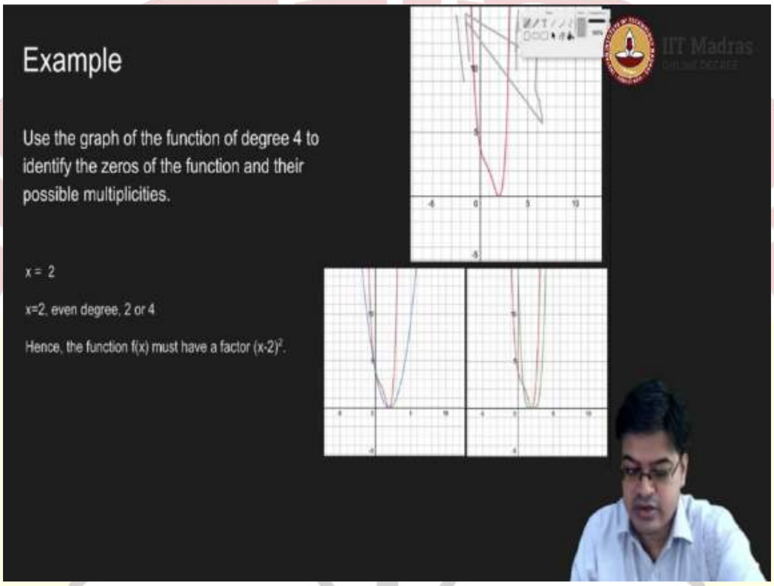
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 46
Graphs of Polynomials: End behavior

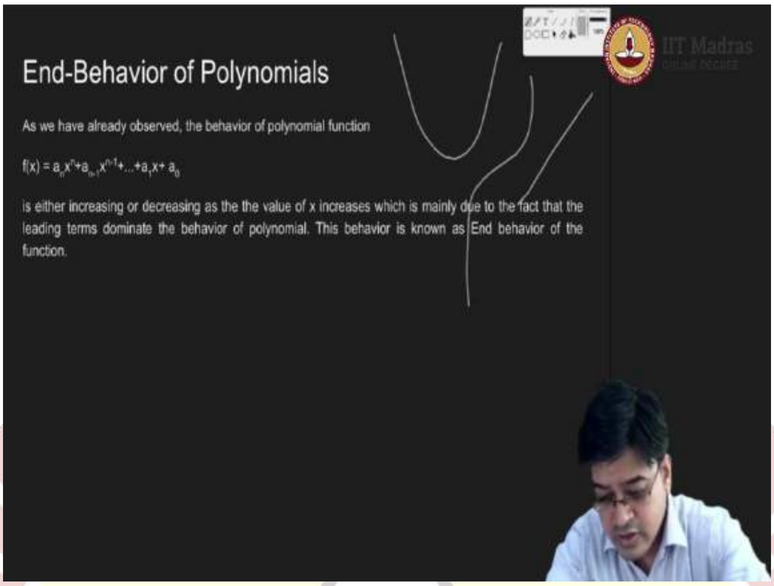
(Refer Slide Time: 00:15)



So, now we have understood how multiplicities affect the polynomial and how we are able to find the multiplicities of the polynomial functions with some factors, correct? Still we do not have an answer to a question that what why what is deciding this behavior that this function will go to infinity, this function will go up as usual, how this behavior is decided, we do not have any answer for that.

Let us try to understand that through end through what is called end behavior of the polynomials. So, the next slide is actually the end behavior of the polynomials ok. So, let us go to the next slide, it is end behavior of the polynomials.

(Refer Slide Time: 01:09)



So, what is an end behavior of the polynomial? In order to understand end behavior, let us define an end behavior properly based on our understanding of quadratic equations. So, when we studied quadratic functions, we looked at the term of the form $a_2x^2 + a_1x + a_0$ right and then, we talked about a_2x^2 whether $a_2 > 0$ or $a_2 < 0$, then we decided the behavior of the function.

If $a_2 > 0$, then we said yes, if $a_2 > 0$, the function will take its minimum and therefore, it will go from both sides to infinity. If $a_2 < 0$, then the function will take its maximum and from both sides, it will go down and it will be unbounded. Now, from the graphs that you have seen in the earlier lectures as well as in this lecture, it is very clear; it is very clear that these functions, polynomial functions are either increasing or decreasing based on the way they wish right.

So, for example, it can be like this also. So, or it can go like this also or it can be a straight line as well, if it is linear or it can move like this. All these are polynomial functions. So, now, we want to have a better understanding. So, what is the behavior of a function after it has passed through all the roots is the question right, that is the term that was troubling us a lot.

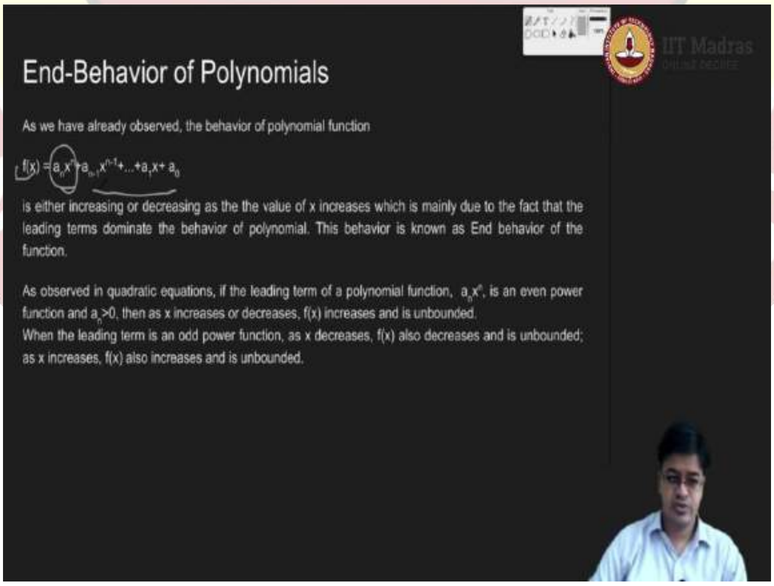
So, for quadratic equations, we have decided that it is basically based on a 2 because quadratic equation the highest degree is 2. So, a 2 it is. So, now, in a similar manner, if I want to consider a polynomial function $f\ (x)$=$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$. Then, the behavior should be decided by this term.

Why should I make this claim? Because if you look at x^n, what we are looking for is as the value of x increases or as the value of x decreases. Now, it is not in that zone, where it is passing through many roots. So, it has passed through all its possible roots and now, after that how the function will behave? There is no determining factor right.

So, in such case, the only determining factor is the term $a_n x^n$; why? Because for large values of x this term x^n will dominate all other terms corresponding to x; x^n raised to n will dominate x^{n-1}, and so on that is when x is becoming large. When x is becoming small that is x is tending to $-\infty$, the term x will be the small x^n will be the smallest possible term or if we that n is of even degree, still it will be the largest possible term.

In any case, the behavior of $a_n x^n$ will play a dominant role in identifying the behavior beyond roots of the polynomials or beyond zeros of the polynomials. This behavior we will call as end behavior of a function and for polynomial functions, it is determined by the leading terms that is $a_n x^n$.

(Refer Slide Time: 05:17)

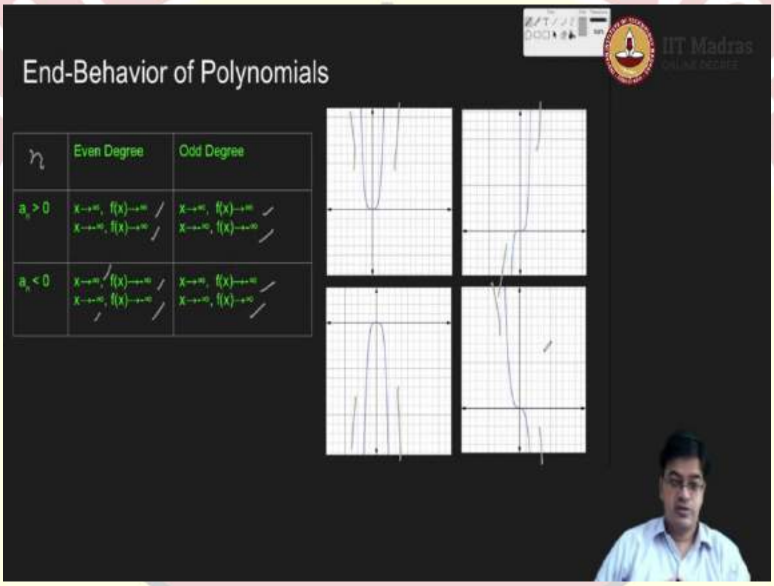


If this $a_n > 0$, and x^n that n is a even power exponent is even, then as x increases or decreases, it is very similar to quadratic. As x increases or decreases, $f(x)$ will always go to infinity. If $a_n < 0$ n is an even exponent, then whether x increases or decreases, $f\ (x)$

will go to $-\infty$. It will go on decreasing. Good. Then, what if $a_n x^n$ that n is the exponent which is of odd power or exponent is odd. What happens?

If $a_n > 0$, then as the function increases, $f(x)$ also increases. If $a_n > 0$ and it is of odd power as x increases, $f(x)$ also increases; as x decreases, $f(x)$ also decreases and both are going to infinity; one is going to ∞, another one is going to $-\infty$. They both are unbounded. Similar thing can be argued for $a_n < 0$. So, in order to improve our understanding, I have tabulated this zone; 1 minute, let me remove this part ok.

(Refer Slide Time: 07:07)



So, this is the better understanding. So, now, you look at the leading term $a_n x^n$ So, this is referring to n, n is of even degree, n is of odd degree. So, if n is of even degree and $a_n > 0$, x tending to ∞, x becoming larger and larger, $f(x)$ will become ∞; $f(x)$ will also increase. $a_n > 0$, x tending to $-\infty$; that means, x is becoming smaller and smaller and smaller; but because the polynomial is of even degree, it will again go to ∞.

In a similar manner, if $a_n > 0$ and the polynomial the leading exponent is of odd degree, then as x tends to ∞, $f(x)$ tends to ∞. You can imagine a function of the form x^3. Similarly, if $a_n > 0$, x tends to $-\infty$, $f(x)$ will also go to $-\infty$because $f(x)$ will also keep decreasing.

Remember polynomials of odd degree crossover x axis, if you link that point to this, then naturally it is very easy to and visualize the behavior of the polynomials. I will demonstrate these two graphs again, once again to reiterate the point. If $a_n < 0$, now $a_n < 0$; that means, x becoming larger, $f(x)$ the term, the leading term of $f(x)$ will be negative more and more negative.

So, $f(x)$ will tend to $-\infty$; but if x is becoming smaller and smaller, the exponent is of even degree, still $f(x)$ will again go to $-\infty$ because $a_n < 0$. Come back to odd degree, here the exact replica of what we have done for odd degree when $a_n > 0$ will happen.

So, in this case when $a_n > 0$, x tending to ∞, we will make bring this $f(x)$ to go to ∞; but in this case, it will bring it to $-\infty$ and similar case is true for the other part that is x tending to $-\infty$, $f(x)$ will tend to ∞. Let us visualize it through graphs.

Let us take this first block even degree $a_n > 0$. Imagine a function of the form x^2 or x^4 as x tends to ∞; both of them are going up. Just remember this figure that will clear this understanding.

Let us go to odd degree with $a_n > 0$, as x tends to ∞, here this is going up. This is going down right. Just imagine a figure of x^3 for the convenience. When $a_n > 0$, just imagine a figure of $-x^2$ or $-x^4$, both of them should naturally go down. That is what is written here as well. In a similar manner, just consider $-x^3$.

So, whatever was going down, will go up and whatever was going up, will go down that is what I meant when I said this. So, now we have much better hold over end behavior of polynomials. Now, you can look at the graph of a polynomial function and you by looking at the end behavior, you can say whether the polynomial, the leading term of the polynomial is of odd degree or even degree.

That is one more understanding, one more level of understanding that we have achieved through understanding this end behavior. But that is not over. We further need better understanding of the functions.

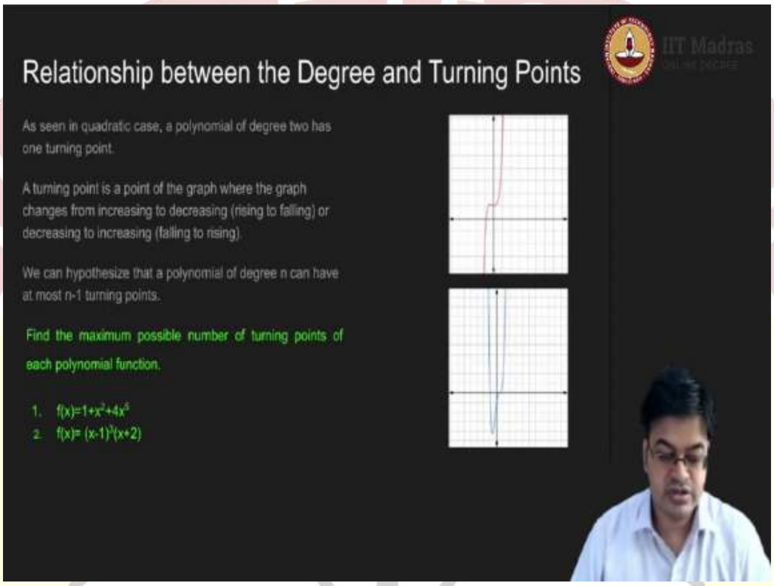
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

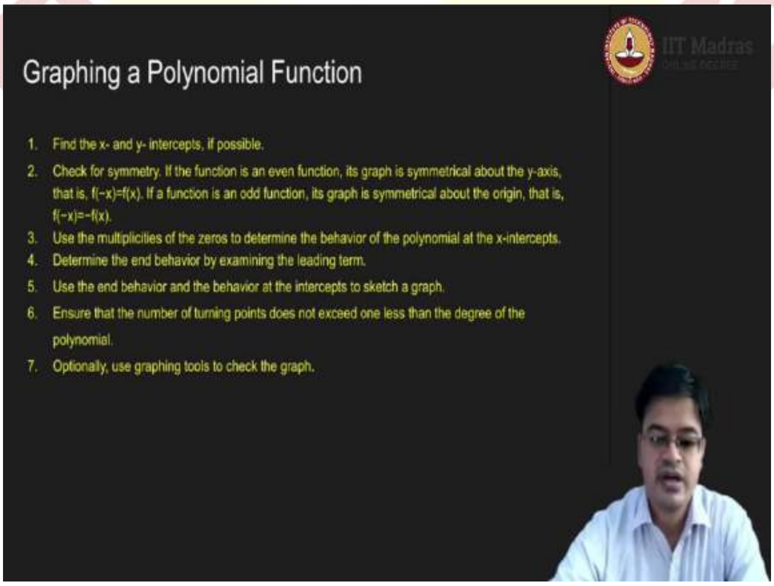
Lecture – 42
Graphs of Polynomials: Graphing and Polynomial creation

(Refer Slide Time: 00:14)



So, now I want to use all this knowledge to plot a Polynomial Function or graph a Polynomial Function.

(Refer Slide Time: 00:19)



So, let us reiterate what are the things that we have seen. For graphing the polynomial function, one way is to find the tabular form and try to graph it as in a crude manner. More knowledgeable way is, you follow this algorithm that is find x intercept, y intercept if possible because it may happen that they do not have any real roots and you may not be able to get x-intercept, all the x-intercepts right.

Then for graphing it is helpful to check the symmetry that is; if $f(x)$ and $f(-x)$ are same if it is an even degree polynomial; that means, you have symmetry about y axis. If it is an odd function you can check whether they are symmetric about origin that is $f(-x) = -f(x)$. Typical case is the first symmetry is $y = x^2$, it is an even degree polynomial and it is symmetric. So, once you have drawn here for $-x$ you have to just keep the mirror image.

That is how it helps in graphing. In a similar manner a $y = x^3$ is a odd degree polynomial and $f(-x) = -f(x)$. Therefore, whatever you got about origin if you reflect about origin then you will be able to retain the same shape; you do not have to compute explicitly. This is the way this checking of symmetry helps.

Next identify the zeros; x intercepts we have already identified. So, you have identified the zeros. Then you identify their multiplicities. If you identify the multiplicities of the polynomials you know the behavior of the polynomials at x intercept. You just recollect; multiplicity the sum of the multiplicities of all zeros cannot exceed the degree of the polynomial that you have to keep in mind. After identifying the multiplicity you know the behavior at the zeros of the polynomial function.

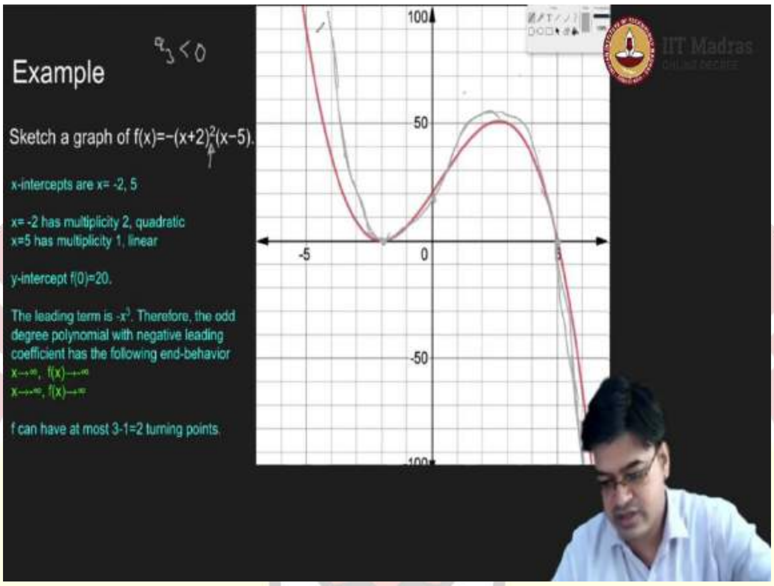
Now, you want to know the behavior beyond zeros of the polynomial function that is; the end behavior. So, end behavior you can use the leading term and you can identify the behavior. Remember the table that we have shown for identifying the end behavior.

And finally, you use the end behavior the behavior at intercepts to sketch the graph. Turning points - the number of turning points can be identified we may not be able to locate exactly where the turning point is. For that, you need the tools of calculus to identify the exact location of a turning point.

And when you identify those when you roughly estimate the turning points; kindly ensure that the number of turning points do not exceed one less than the degree of the polynomial.

So, if the degree of the polynomial is n, the number of turning point should not exceed n $- 1$ ok. And finally, you can use technology to sketch the graph. So, use graphing tools like Desmos or some other tools for graphing the function ok.

(Refer Slide Time: 04:07)



So, let us see this in action. So, here is an example I want to sketch a graph of this polynomial function; $-(x+2)^2(x-5)$. Obviously, I have figured out oh it is a $-(x+2)^2$. So, the first thing that we; so, I want to graph this function. So, the first thing that I want to find is the x intercept, because it is given in factored form it is no brainer; $x = -2$ which is this point $x = -2$ and then $(x - 5)$. So, $x = 5$ which is this point.

These two are the zeros of the polynomial functions; $x = -2$ has multiplicity 2 and it is an even degree polynomial. So, over here the behavior of function I am trying to sketch it will come from here, it will go from here. So, I know the behavior of the function is of this form, it will just pass through the axis.

And $x = -5$ I do not know the exact values, but roughly it will be more of the linear form and it will pass through the point -5, then it will come down over here and then it will pass through this point. So, up to this I am ok. Now, you look at this polynomial if you look at this polynomial, then the polynomial will be a cubic polynomial; it will have a negative term.

So, essentially $a_3 < 0$ ok. So, the end behavior of this polynomial because $a_3 < 0$ as $x \to \infty$ this function will tend to infinity. Yes, and as $x \to -\infty$ the function will naturally go up like this.

So, this is the vague understanding of the behavior. If I want to get more precise on what values this is roughly the shape of the function. If I want to get more precise on what values the function takes, I can consider the y intercept as well that is I will put $y;\ x = 0$. So, it will be 2^2 4 $2^2 4$ yes, and into -5 that will give me $-20 + 20$. So, this intercept that I have drawn is wrong. It should be somewhere here + 20. So, let me erase this and redraw the function again.

Let us take the eraser. So, it may not go this high as well. So, over here the font behavior of the function ok; so, let me again go back to the marker. And the function may cut here itself pass through this point and join this point. Yes, so, this bulge will not be there because this function is linear over here. It may be of this form. So, let me again erase this part yeah.

So, let us see. So, we have identified the end behavior ok, final check that number of turning points. The function is cubic, so it can have at most two turning points, there are only two turning points: one is here, one is here fine. So, let us see whether whatever we have said is correct or not. So, let me hide this first.

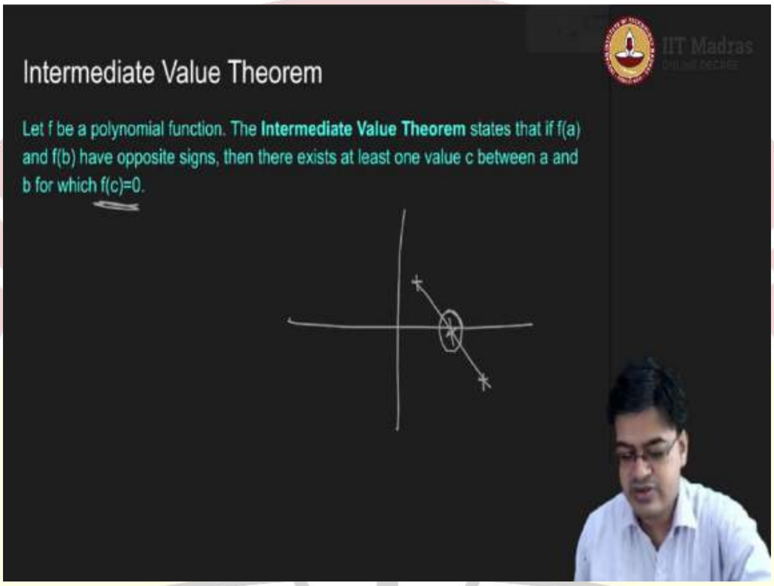
So, x intercept is -2 and 5, no problem. $x = -2$ has multiplicity 2. So, the quadratic behavior should be plotted there. Yes, $x = 5$ has multiplicity 1. So, a linear behavior is plotted here assume this is a line. So, linear behavior is a plot must be plotted here. And, then $f(y)$ intercept $f(0)$ is 20 which we corrected we were not correct in the initial stages and the leading term is $-x^3$.

So, therefore, the odd degree polynomial with negative leading coefficient has the following end behavior; as $x \to \infty\ f(x) \to -\infty$, this is the behavior that we have plotted; $x \to \pm\infty\ f(x) \to \infty$, this is also correct. And f can have at most $3 - 1 = 2$ turning points, this is the behavior right.

So, now, I was roughly ok in drawing the graph of a function. This is because I do not exactly know the behavior of the turning points. So, I will be roughly ok in drawing the graph of a function, but not exactly. If you want to be more precise you can actually tabulate the values around some critical points and then you can figure out. This is when the formula is given to you.

Now, the question can be asked that what if the formula is not given to you, but you have been given only a function. And from the graph you need to identify the polynomial.

(Refer Slide Time: 10:17)



In such cases one theorem which will help you a lot, I will not use this theorem in a rigorous manner. But it will help you a lot, is intermediate value theorem because we are dealing with continuous functions. This intermediate value theorem is valid for all continuous functions.

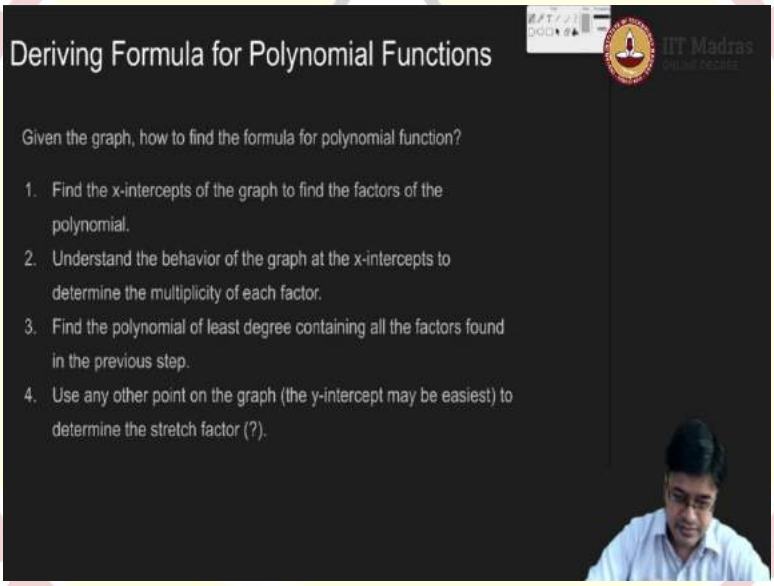
What this theorem says is, a polynomial function is a continuous function. So, let f be a polynomial function, then the intermediate value theorem states that; if $f(a)\ and\ f(b)$ have opposite signs; that means. So, let us say $f(a) > 0$, and $f(b) < 0$ and $a > b$, then there exist at least one c between a and b such that $f(c) = 0$; that is essentially the meaning.

For example, I have this coordinate plane my value of $f(a)$ is here, and $f(b)$ is here, and the function that is given to me is a continuous function right. So, finally, it has to pass through the x axis to reach the value here right. So, in such cases we will say that; this is the 0 of the polynomial that is what we are calling as c, $f(c)$.

So, you using this you when you are actually having trouble in finding the zeros of the function, you can actually evaluate two values any two values of opposite signs. And if you evaluate any two values of opposite signs, then you know that there is some root some 0 in between that will improve that you will gain a confidence by doing these things.

So, this is an important theorem in mathematics, intermediate value theorem. You can use this to find the roots of the polynomial when you are having difficulty in identifying the roots of the polynomial. So, you simply put $f(a)$ and $f(b)$ if they have opposite sign, then there is at least one root in between; that is the meaning. You can use this theorem to your advantage.

(Refer Slide Time: 12:37)



So, using this theorem we can actually derive a formula for polynomial function. You use this theorem to identify the zeros; rest of the methodology is similar. So, how to derive a formula for polynomial functions? So, given a graph of a polynomial, how to find with in coordinate axis? You have all the numbers attached to it, then the question can be asked as to how to find the polynomial function the algebraic expression of a polynomial function?

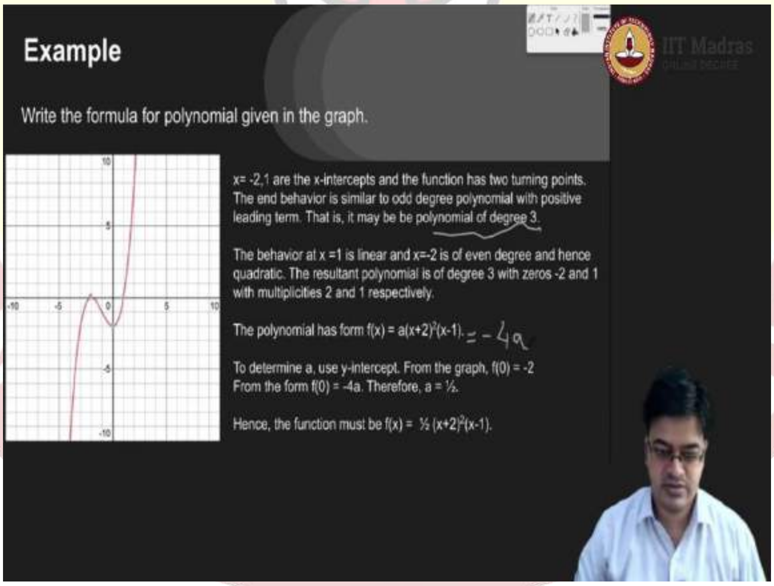
So, in that case our modus operandi is similar to what we have done. Find the x-intercepts from the graph. Find the factors of the polynomial, this we already know. Understand the behavior of x intercepts around x intercepts to get more understanding of the x intercepts that is zeros of the polynomial about their multiplicities.

So, you will find multiplicity of each factor. Once you have gained understanding identify the end behavior that also you have to do. Next, after doing that you find the least degree polynomial containing these factors. What are the factors? Those are x intercepts that you have figured out. You have also seen the end behavior, so the least degree polynomial which will give you that particular function behavior.

Once the least degree polynomial is figured out you use any point on the graph that is why the coordinate axis is important, the numbers are important. You use any point on the graph, in particular y intercept is the easiest and in that case you can determine the stretch factor.

The stretch factor over here is the unknown a that I have told you while figuring out the factors in one of the examples. So, that is the stretch factor. It will be more clear when we will solve the examples ok. So, this is our recipe for attacking the problem of deriving the formula given a graph.

(Refer Slide Time: 14:58)



So, let us try to apply this recipe to one example. So, write the formula of a polynomial given in the graph, the graph is here ok. So, I will go around and try to find the x intercepts of this graph. So, one x intercept is here which I think is $x = 1$ and other x intercept is -2. So, -2 and 1 are the x intercepts; y intercept over here is-2, $0-2$ is the y intercept. So, we have identified x and y intercepts.

The graph actually seem to have two turning points. So, the least degree if it has two turning points the least degree polynomial will be because $n - 1 = 2$. So, the least degree polynomial should be cube degree 3 polynomial right ok. And since it is crossing over this end from end behavior also you will have some understanding that it is yeah, it should be an odd degree polynomial. So, therefore, the polynomial may be of degree 3, correct.

It should be an odd degree polynomial; it has only two turning points. So, the least degree of the polynomial is 3. Now what you will do next? Next I want to identify the multiplicities, that is $x = 1$ it the function more or less seems to be linear and at $x = -2$, the function more or less seems to be quadratic.

So, it is very easy in this case because, $x = -2$ is a even degree behavior, x is equal to because it is bouncing off. So, it is a even degree behavior and $x = 1$ is linear behavior and the polynomial is of degree 3 or more, but odd degree. So, the first instance is you guess the function to be of the form $(x + 2)^2(x - 1)$. So, now, I have not yet used the information that the intercept the y intercept is happening at -2, correct.

So, that information I have to use now because that is the function value that I have. These are the based-on factors we are basically equating to 0 right. So, the a may be missed out. So, where the non-zero value comes you should be able to figure out. You can you are free to choose any value, but for me it is better to choose y intercept.

So, y intercept is -2, we have already seen that, but if you put what is y intercept? It is $f(0)$. So, if you put this in the function form the value of 0 in the function form over here you will get actually this is to be equal to $-4a$. So, if you are getting this to be equal to $-4a$, then $-4a$ must be equal to -2; that means, $a = \frac{1}{2}$ great.

So, if $a = \frac{1}{2}$you substitute this value into the function. So, $f(x) = \frac{1}{2}(x + 2)^2(x - 1)$, fantastic. So, you have you got an algebraic expression. Now, to match this algebraic expression, you use the technology that is graphing tool to plot the function and you can verify the result for yourself that yes, this is the function that we have actually plotted ok.

So, this is the complete understanding of two-step mission that is; given an algebraic expression how to graph the polynomial function. Given the graph of a polynomial function, how to write an algebraic expression of a polynomial function. This ends our topic on polynomial functions.

Thank you.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

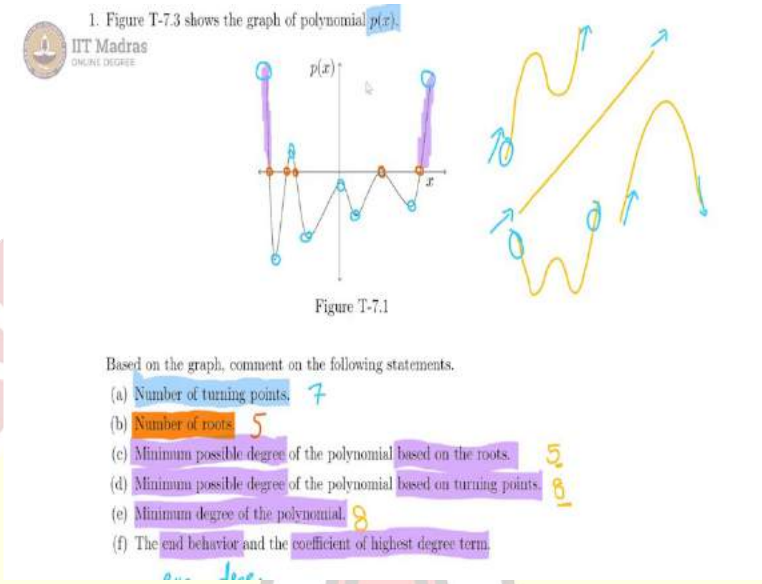
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 07 - Tutorial 01

(Refer Slide Time: 0:14)



Hello mathematics students. In this tutorial we are going to look at questions based on graphs of polynomials. So, in this question there is this polynomial $p(x)$ whose graph is given here and we are supposed to comment on the following statements, the number of turning points, so that is easy so there is a turn here, 1, 2, 3, 4, 5, 6 and 7. So, there are 7 turning points. And then we are asked the number of roots, so roots would be where the polynomial touches or cuts the x axis so that is 1, 2, 3 and 4 and 5, so there are 5 distinct roots.

Now, what is the minimum possible degree? Minimum possible degree of this polynomial based on the number of roots. So, the minimum possible degree would be the same as the number of roots so if there are n roots to a polynomial then it should have a degree of at least n, so 5 is the minimum possible degree of this polynomial based on the roots. But now they are asking what is the minimum possible degree based on the turning points.

So, here we see this thing a straight line has no turning points, a quadratic equation has 1 turning point and a cubic would have 2 turning points at most likewise a quartic that is a fourth degree polynomial would have 3 turning points at most. So, if you have n turning points, then the minimum possible degree of the polynomial would be $n+1$. So, here that is 8.

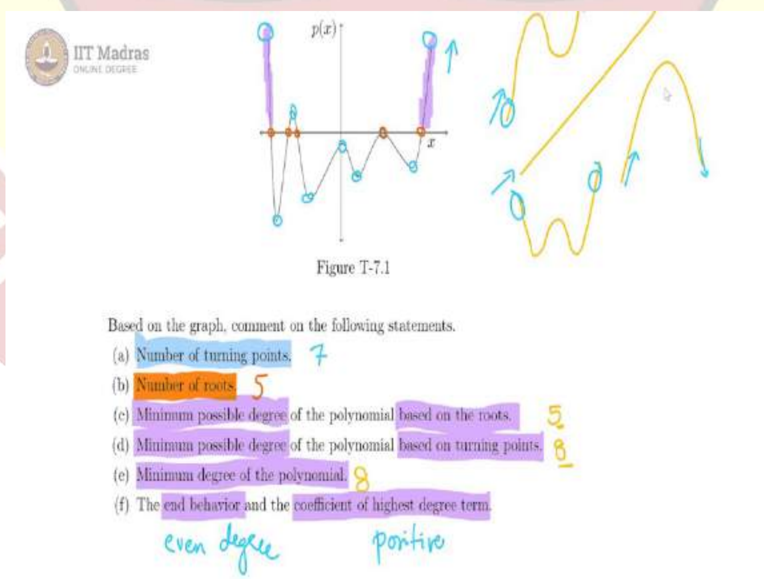
Now, what would be the minimum degree of the polynomial given all the information we have? Then you know that it has to be at least 8 the five which is on the basis of the roots is a lesser number than it and we know already that it has to be at least 8, so the minimum degree of the polynomial should be the greater of these two which is 8 because 6 and 7 and 5 are not allowed on the basis of turning points.

And then we are being asked what is the end behavior and the coefficient of the highest degree term. So, the end behavior shows that the polynomial is coming from ∞ and going to ∞ which means the degree of the polynomial is definitely even. So, we can say that it is an even degree polynomial.

So, as you can see we have just drawn these basic raw curves for the linear and quadratic and cubic and quartic polynomial. So, linear which is an odd degree polynomial it comes from -∞ and it goes to +∞ whereas quadratic a parabola here it is coming from -∞ and it is going to -∞ .

So, when it is even degree you see that the ends of the curves are in the same directions. Similarly, for quartic here this is coming from ∞ and going to ∞, whereas for a cubic this is coming from -∞ and going to +∞. So, here this is coming from ∞ and going to infinity.

(Refer Slide Time: 4:17)



Therefore, this is a even degree polynomial and the coefficient of the highest degree term. So, the coefficient of the highest degree term determines whether the behavior of the polynomial as x increases whether it is going to $+\infty$ or $-\infty$, if the coefficient of the highest degree term is positive, it goes to $+\infty$. So, if this is going to $+\infty$ so this has to be positive coefficient for the highest degree term.

IIT Madras

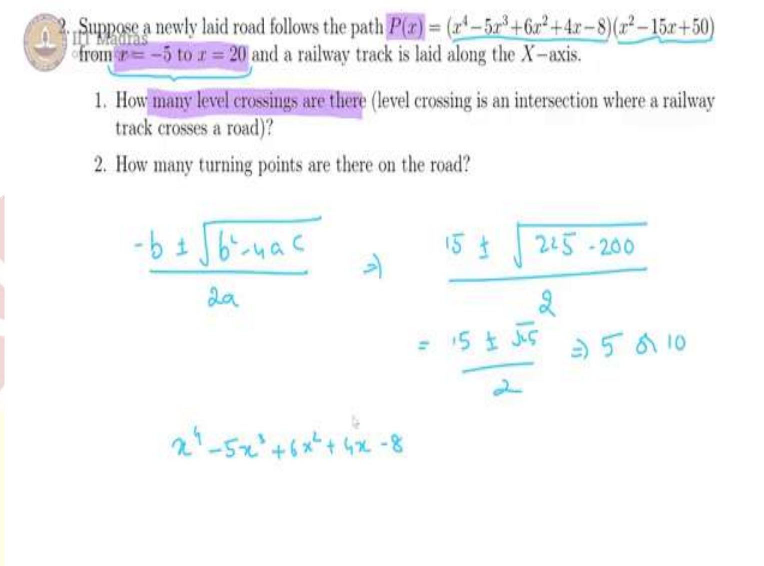
ONLINE DEGREE

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 07 Tutorial 02

(Refer Slide Time: 00:15)

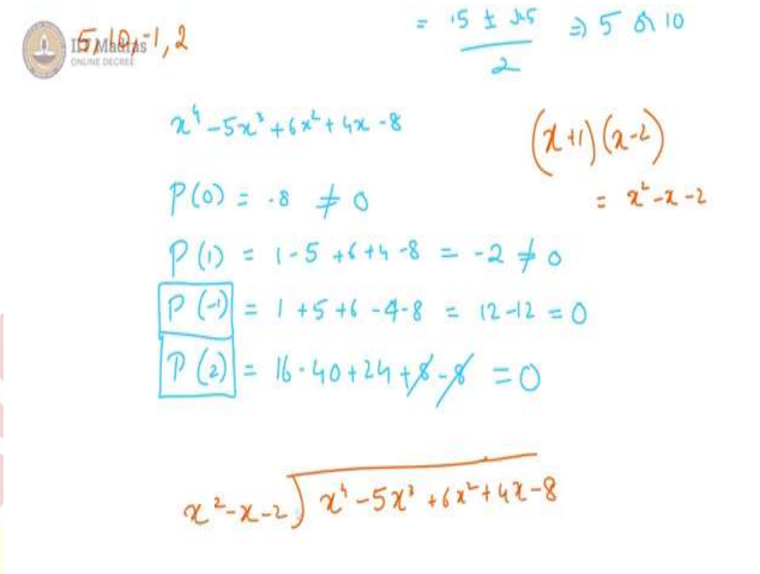


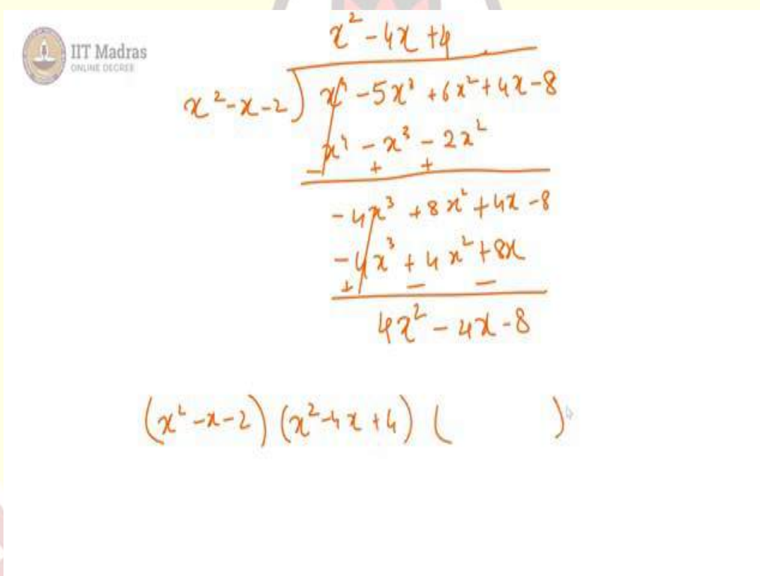
Now second question there is newly laid road which follows the path of this polynomial about some coordinate system, from $x = -5$ to $x = 20$. And railway track is laid along the x axis. So how many level crossings are there? So what we are interested in is; how many times does the x axis cut this polynomial? And for that we have to find the roots of this polynomial because roots give when the polynomial is touching or cutting the x axis.

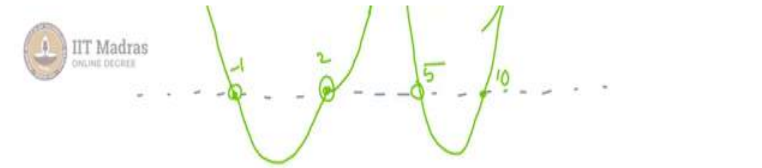
Now this is of quartic forth degree polynomial multiplied with the quadratic polynomial, so the degree is 6, so at best we could have 6 roots but let us find out what these roots are. The easy way to start is to first find the roots of the quadratic, so that would be using $\frac{-b\pm\sqrt{b^2-4ac}}{2a}$. We get $\frac{-15\pm\sqrt{225-200}}{2}$, which is $\frac{-15\pm\sqrt{25}}{2}$ that is essentially 5 or 10.

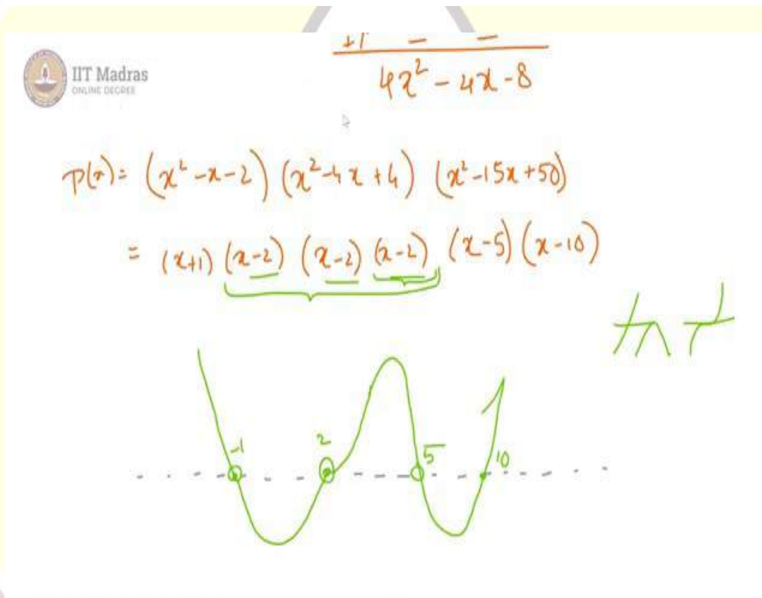
So you get $\frac{10}{2}$ or $\frac{20}{2}$ so 5 or 10 those are the two roots and they are both within the given range. Anyway now we look at the other part, the quartic part. So here we have $x^4 - 5x^3 + 6x^2 + 4x - 8$. In this situations what is typically suggested is that we do a little bit of trial an error, we try out with the basic small integers and we see if we can find any roots at all.

(Refer Slide Time: 02:26)









So let us start with $p(0)$, $p(0)$ is -8 which is clearly $\neq 0$, so 0 is not a root then we have $p\ (1)$ which is $1 - 5 + 6 + 4 - 8 = -2$ which is again $\neq 0$, so not a root. Then we try $p(-1)$ and we get $1 + 5 + 6 - 4 - 8$, so this is equal to $12 - 12 = 0$. So yes $p(-1)$ gives you 0 which means we have another root that is - 1.

So let us note down our roots that we have found here, roots we have found so far are 5, 10 and -1. Now going back to our trial and error let us try $p\ (2)$ and $p\ (2)$ gives us $16 - 8 \times 5 = 40 + 6 \times 4 = 24 + 8 - 8$. So we get $16 + 24 = 40$, $40 - 40 = 0$, so this is 0. So we have another root that we have found. So we now have two roots for our quartic and those two roots give us another quadratic which is $(x + 1)(x - 2)$ that is $x^2 - x - 2$.

So if we divide our quartic with quadratic we will get the other quadratic within it. So here we have $x^4 - 5x^3 + 6x^2 + 4x - 8$ and we divide it with $x^2 - x - 2$ so here go x^2 so $x^4 - x^3 + (m-2)x^2$ -,+ and + cancel this of you get $-4x^3 + 8x^2 + 4x - 8$.

And then we do $-4x$ times this, $-4x^3 + 4x^2 + 8x$, so + - and - cancel this and here we have $4x^2 - 4x - 8$. And that is just 4 times this, so + 4. So our quartic, so is basically x$(x^2 - x - 2)(x^2 - 4x + 4)$ and this gives the quartic and additionally we have to also multiply for our p of x we have to multiply the other quadratic which is $x^2 - 15x + 50$, this one.

So this is p of x totally, an if we further separate it out into all its roots we get this one as we know is x + 1 into x - 2 and this is if you notice x - 2 to the whole square, so $(x-2)(x-2)$ and then here this we have found the roots already which is $x - 5$ into what was the other root; the other root was 10, $x - 10$.

So these are our roots and the coefficient of x power 6 will be positive clearly. So therefore this is an even degree polynomial and thus if we have to sketch the graph it look something like this, it comes from infinity and what is the least lowest root here, the lowest root is - 1. So at - 1 if we draw this as the x axis at - 1 you have one root it crosses the x axis and then it goes around and it comes to 2.

But 2 is a triple root, so what happens with a root if it is a single root it crosses the x axis but if it is a double root it will just touch the x axis and come back but since it is a triple root it actually crosses the x axis. So here we do have a crossing and then afterwards at 5 and 10 we will have, so this will be for two this will be for - 1, this will be for 5 and this will be for 10. This is just a rough plotting of the graph.

The question was how many times does it intersect the x axis, so we have to draw this basic sketch and we find that the intersection are 4. If x - 2 was not a triple root, if it were a double root or a quadruple root like if it is there are 2 times or 4 times then the graph would be very different. It would be x - 1 would still be the same but at 2 you would not actually see a intersection, you will just see a touching. It would not be a cut.

So therefore we have to check how many times the $\sqrt{2}$ occurs. Since it is an odd number of times we can say it is actually cutting x axis and that gives us a number of level crossings is 4. And how many turning points are there? Now we can look at our graph and quickly tell; 1, 2 and 3; 3 turning points.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 07 - Tutorial 03

In this question Saraswati bought an 8 gram gold chain for rupees 40000, so we can presume that 1 gram is 5000 rupees on first November. And after 10 months that is August 2021, she sold the chain and bought a new 10 gram gold chain by paying an additional 10000 rupees. Suppose, the rate of the gold per gram is denoted by $G(t)$ and it is a function of time $G(t)$ is given to be this cubic polynomial here and we are taking t is to be 0 at the time when Saraswati bought her first gold chain. So, t is a number of months since her buying her first gold chain.

Now, what is the when $G(t)$ is a polynomial of the rate for both used and new good. So, all gold has the same rate as what we are considering and what is the rate of gold per gram when she sold her first chain. So, after 10 months at $t = 10$ is what we are really looking for. So, that means we are looking for $G(10)$ and that gives us $0.07 \times 1000 - 1.4 \times 100 + 7 \times 10 + 5$ and this is $70 - 140 + 70 + 5$. So, that is actually 5. So, the rate is back to 5000 per gram. So, it is again rupees 5000 per gram.

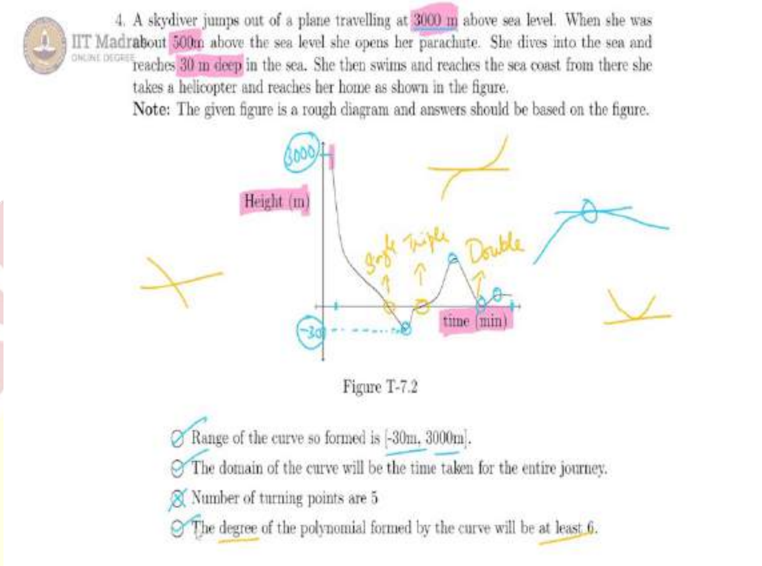
Now, if she had sold the first gold chain after 6 months how much extra would she have paid for buying the 10 grams gold chain? So, after 6 months we have to find the price, the rate, so that would be $G(6)$ and that is 0.07×6^3 is $216 - 1.4 \times 36 + 7 \times 6 + 5$. And then we get this is $15.12 - 1.4 \times 36 = 50.4 + 42 + 5 = 42 + 5 = 47 + 15.12 = 62.12$ that is $62.12 - 50.4 = 11.72$ that would give us then the rate is 11720 rupee per gram. And Saraswati is selling 8 grams at this price that would mean she basically has to pay for the additional 2 gram and that would be 2×11720 which is equal to rupees 23440, this is how much she pays extra for her 10 gram gold chain.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

IIT Madras
ONLINE DEGREE

Mathematics for Data Science 1
Week 07- Tutorial 04

(Refer Slide Time: 0:14)



A skydiver jumps out of a plane travelling at 3000 meter above sea level. And when she was about 500 meter above the sea level, she opens a parachute. And then she dives into the sea and reaches 30-meter-deep into the sea. Then she swims and reaches a sea coast from there she takes a helicopter and reaches her home as shown in the figure.

So this figure is between the height and time there is no x coordinate in this. This is only about the y coordinate taking sea level to be 0 and the time. So initially she is way above sea level, so this point is going to be our 3000 because as where she is jumping from and then she is dropping quite quickly and then she slowly dropping after she opens a parachute. And she reaches under the sea so here it is negative till she goes to the point where it is - 30 meter below sea level.

And from there she swims out and she takes a helicopter and she goes. So we are supposed to see which of these option are correct and the range of the curve so formed is - 30 to 3000 which is true - 30 to 3000 is her total y coordinate range. And the domain of the curve will be the time taken for the entire journey that is true so your curve starts here when she jumps till the point she reaches home.

So this is correct this is also correct. Number of turning points are 5 so I see only 1, and 2, 3 turning points. So this is not maybe this is a turning point I am not able to say because it appears to go a little like this and then bend down. So probably this is a turning point. But either way there are not 5 they are less than 5 so this is wrong. And then the degree of the polynomial formed by the curve will be at least 6.

Now we have to look at the roots here. So let us take this root this is a single root it just cuts the $x$axis like this, whereas this is a more, if the x axis like this, it is kind of touching it this way and that only happens if your root is a triple root at least. So it cannot happen for single root and it does not happen for any even powered root because for root which occurs even number of times you would not cut the x axis.

So this has to be at least a triple root. So this is a single root the first one is the single root this one we are assuming it is a triple root because they are asking for minimum degree at least so we are looking for what it what the number of roots is in the minimum. And here there is a root which occurs an even number of times because it is touching the x axis and turning around it is not actually crossing the x axis.

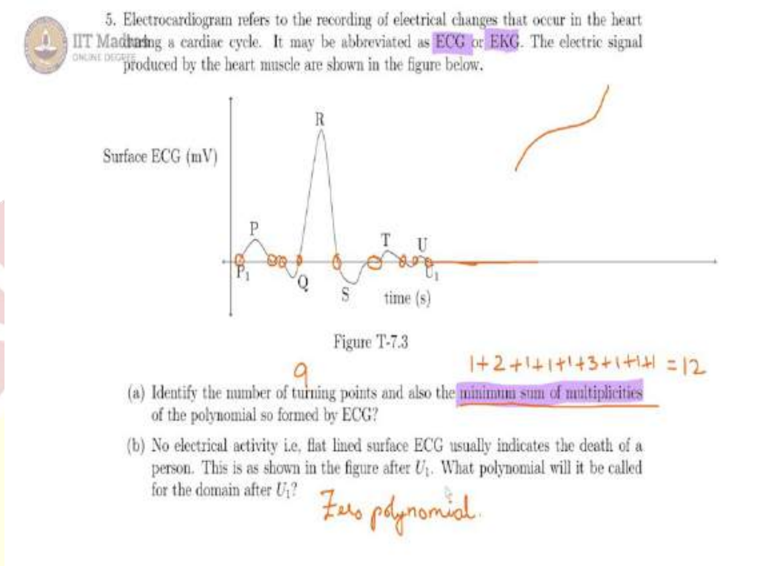
So this is at least a double root so we have one +3 +2. So we have at least 6 roots therefore the degree also has to be at least 6. So this is also correct.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Week 07- Tutorial 05

(Refer Slide Time: 0:14)



In this question, we are looking at an Electrocardiogram which is often called ECG or EKG. This is a recording of electrical changes that occur in your heart during a cardiac cycle. So here we have some ECG shown to us as a polynomial. And they are asking identify the number of turning points. Ok? So that will be 1, this is 2, this is 3, 4, 5, 6, and here this is not a turning point, it is flattening out like this and rising. Therefore it is not a turning point. We already have 1,2,3,4,5,6 so this is 7,8,9. That means we have 9 turning points. And then they are asking for the minimum sum of multiplicities for that we look at the roots and so this one is directly cutting through this root is directly cutting through the axis.

So the multiplicity of this is 1 and here this it is touching and coming back so it has to have an even multiplicity. So the minimum is 2 and then here again this and this are both 1 each $+1$, this 1 also should be 1. And here we see this flat lined situation which occurs when you have an odd multiplicity but not 1. So the minimum there would be 3 and then this is a 1 this is a 1 and this also has to be 1.

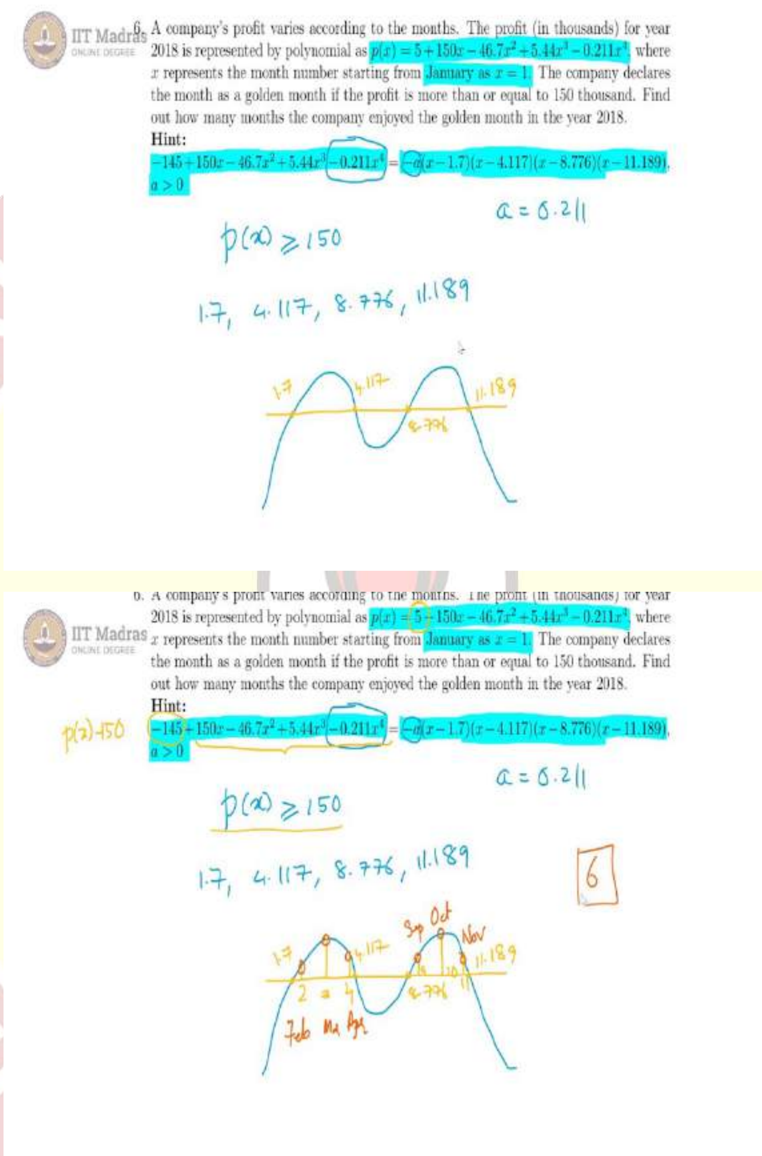
So plus $1 + 1 + 1$ which gives us all put together 3, 4, 5, 6, 9, 10, 11, 12, 12. So the minimum sum of multiplicities is 12. No electric activity that is flat lined surface ECG usually indicates the death of a person. This is as shown in the figure after U_1 so after U_1 we presume that it is a, it is basically along the x axis what polynomial will it be called for the domain after U_1. Clearly a 0 polynomial. It is simply y is equal to a constant. So it has no degree and it is a 0 polynomial.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

IIT Madras
ONLINE DEGREE

Mathematics for Data Science 1
Week 07- Tutorial 06

(Refer Slide Time: 0:14)



So here we have a company's profit varies according to the months. So they are going to have a profit versus time along in the profit in thousands for year 2018 is represented by this polynomial $p(x)$ is equal to this quartic polynomial fourth power polynomial where x represents the month number starting from January as $x = 1$. So January is x=1. The company declares the month as a golden month if the profit is ≥150 thousand.

So golden month is when $p(x) \geq 150$. Find out how many months the company enjoyed the golden month in the year 2018. So how many times does this happen? Alright, and there is a hint also given to us where this cortex is apparently equal to. So they have basically given us the roots of the polynomial and a does not matter, $a > 0$.

And so we can also tell what a is a has to be minus 0.211 because that is the coefficient of x^4 here and minus a will be the coefficient of x^4 and the RHS therefore a has to be equal to 0.211, $-a = -0.211$. Therefore, $a = 0.211$. Anyway so now given that we already have the roots, the roots are essentially 1.7, 4.117, 8.776, and 11.189. So these are the roots and the coefficient of the highest power of x is negative.

It also even so our curve is going to be something like this. Where the x axis cutting it here and that would mean this point is 1.7, this is 4.117, this is 8.776, and this is 11.189. So all of this is given to us but what we are supposed to find is to related to $p(x) \geq 150$. So given that $p(x)$ is all this stuff plus 5 and here we have the same terms of $x - 145$, we can see that this particular polynomial given to us is simply $p(x) - 150$.

So whenever this polynomial that we have drawn here is greater than it. We have a golden month so that would be month 2, 3, and even 4, then 5, 6, 7, and 8 do not come in and here we would have month 9, and 10, and also 11. So as you can see this, this, this 3 and again 3 here. So it is overall 6 months during which we can also tell which ones is Feb, March, and April, and here this is September, October, and November.

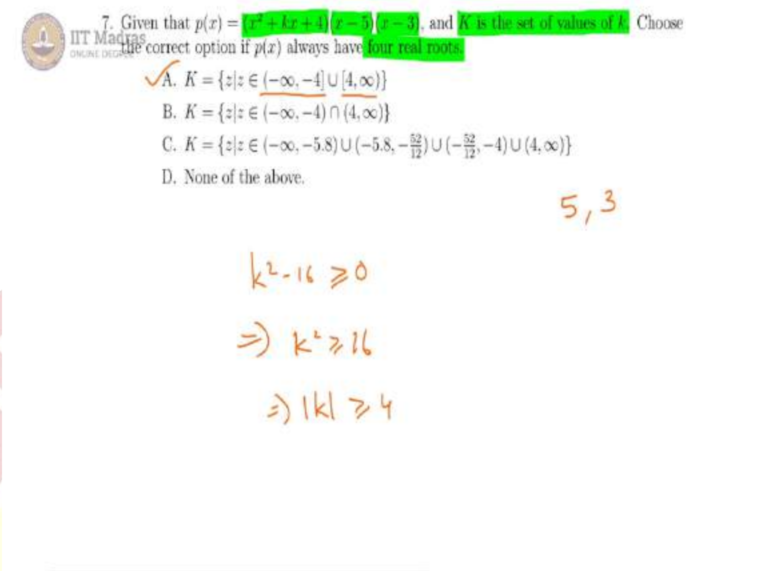
So 6 months are the golden months in that year for this company.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science -1
Week 07 - Tutorial 07

(Refer Slide Time: 0:14)



In this question, we are given a polynomial $p(x)$ which is a product of a quadratic with a monomial and another monomial. And the quadratic has some variable k in it, capital K is the set of values of this small k, choose a correct option if $p(x)$ always has 4 real roots but they need not be distinct and we already know that 5 and 3 are roots because of these two monomials. So, what is remaining is that our quadratic equation also should have roots.

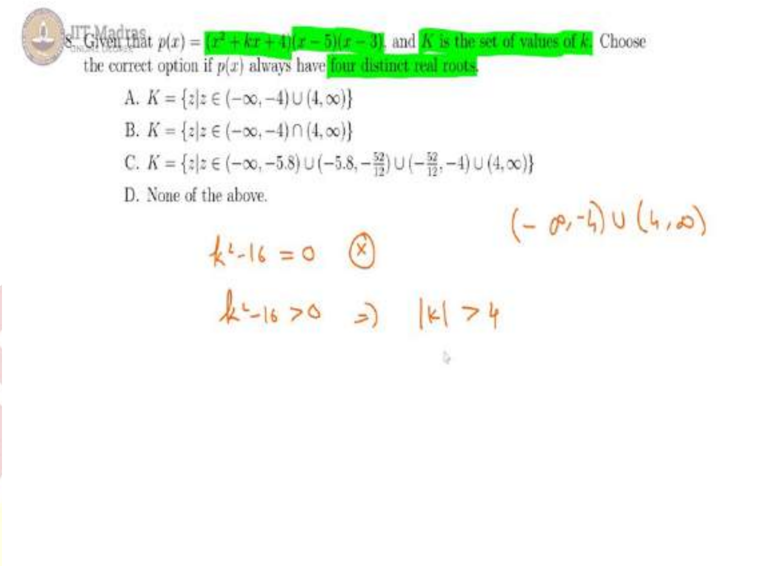
And for that the discriminant which is $k^2 - 16$ should be ≥0. That would indicate $k^2 \geq 16$, thus k, the magnitude of $k \geq 4$. If $k \geq 4$ you get a repeated root you get the same root twice, so what corresponds which option corresponds to this is a because you go from $-\infty$ to -4 and then 4 to $+\infty$ and their union and 4 and $-$ 4 are with closed intervals therefore, they are included.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science -1
Weel 07-Tutorial 08

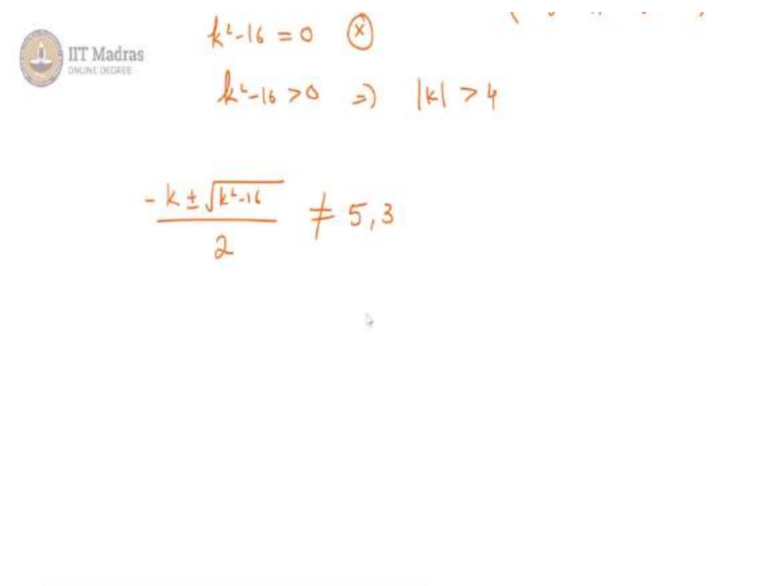
(Refer Slide Time: 0:14)



Question 8 is very closely related to question 7, we again have the same, quadratic into monomial into monomial is the same polynomial and again the same set is given to us. Now we have to see the correct option for $p(x)$ to have four distinct real roots, that means a root should not be equal to each other and that is a catch.

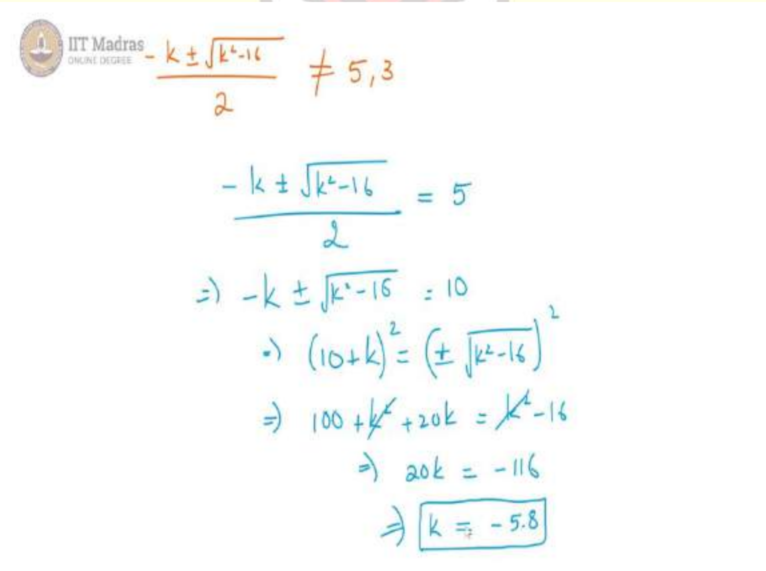
So, we have already seen that $k^2 - 16$ the discriminant being equal to 0 will give us equal roots. So, this case is not done, this time the discriminant has to be greater than 0, so that would indicate $|k| > 4$, so you will have $(-\infty, -4) \cup (4, \infty)$ for the quadratic condition. The other condition here is that the roots for the quadratic should not be equal to 5 or 3.

(Refer Slide Time: 1:24)



So, these routes which are $\frac{-k\pm\sqrt{k^2-16}}{2}$ this should not be equal to 5 or 3.

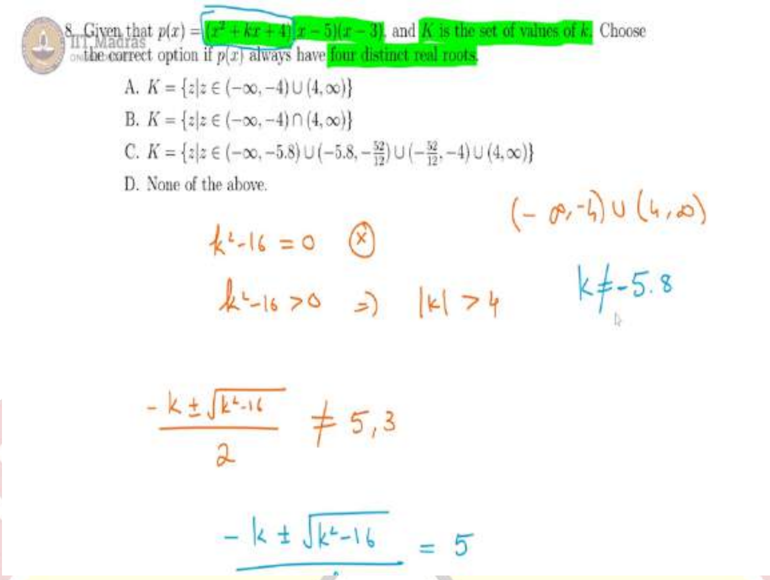
(Refer Slide Time: 1:46)



So, for finding that condition let us start with $\frac{-k\pm\sqrt{k^2-16}}{2} = 5$ let us start with this and you get $-k \pm \sqrt{k^2-16} = 10$ and that would mean $10 + k = \pm\sqrt{k^2-16}1$.

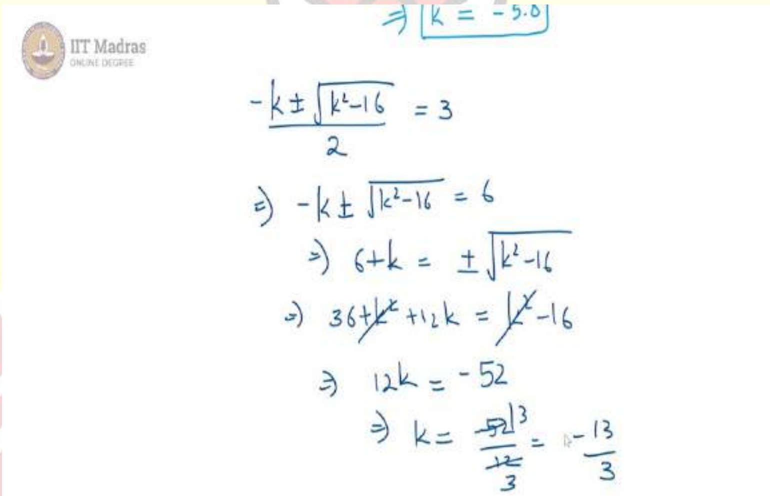
Now if you square this we do not need to worry about the plus or minus, so let us square it and we will reach $100 + k^2 - 16$, k^2 and k^2 goes away. So we get $20k = -116$ and that would imply $k = -5.8$. So when $k = -5.8$ the root of the quadratic part will be equal to 5.

(Refer Slide Time: 2:52)



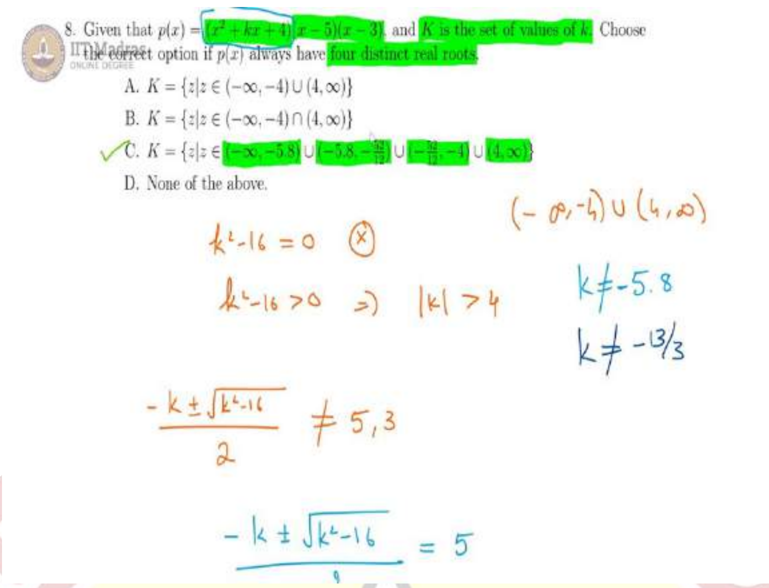
So the root of this part will be equal to 5 and that is not allowed, so we should somehow eliminate 5.8 from this set, -5.8 from this set.

(Refer Slide Time: 3:08)



And further let us check for three case where $\frac{-k \pm \sqrt{k^2-16}}{2} \neq 3$ so we first check when is it equal to 3 and that gives us $-k \pm \sqrt{k^2 - 16} = 6$ that gives us $\pm\sqrt{k^2 - 16} = 6 + k$ and that further gives us $36 + k^2 + 12k = k^2 - 16$. So k^2 and k^2 canceled off and that gives us $12k = -52$ this implies $k = \frac{-52}{12}$ which is essentially for 3 and 4 13, so $\frac{-13}{3}$.

(Refer Slide Time: 4:09)



So, k should not be also be equal to $\frac{-13}{3}$, so which of these options does that is here we see option c is goes from $(-\infty, -5.8) \cup (-5.8, \frac{-13}{3})$ and keeping it open interval we are basically exploding -5.8 and similarly the open interval on the $\frac{-13}{3}$ side on in this and this is essentially excluding $\frac{-13}{3}$ and lastly we are doing the union with 4, ∞. So, this is correct, we are excluding all values from -4 and 4 and also excluding -5.8 and also excluding $\frac{-13}{3}$.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

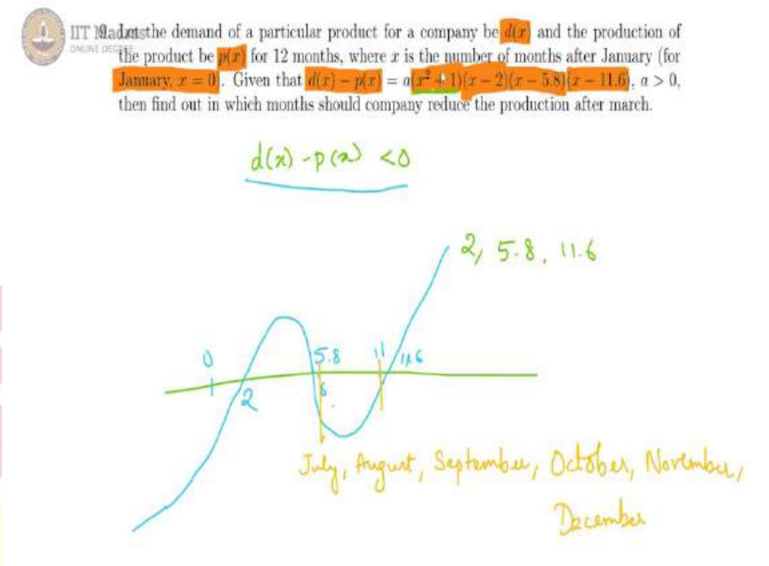
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science -1
Week 07-Tutorial 09

(Refer Slide Time: 0:14)



For our last question, there is a company and they are making a particular product and the demand of the particular product is us $d(x)$, the production of the same product is $p(x)$ for 12 months, where x is the number of months after January and for January we are taking x is equal to 0. And then they have given us $d(x) - p(x)$, as a polynomial and this is essentially a quadratic multiplied by a monomial by another monomial and another monomial.

So, we have a fifth degree polynomial here $d(x) - p(x)$, then find out which months should company reduce production after March. So, reduced production would mean $p(x)$ is greater, that means $d(x) - p(x) < 0$ and we are interested in those situations where this curve is less than 0 and so we just try to graph this curve and $x^2 + 1$ has no real roots.

So, the only roots here are 2 and 5.8 and 11.6, so our curve is going to look something like and since a is positive, then since the coefficient of the highest power is positive we will have this how polynomial go to ∞ as x goes to ∞ and then we have a situation like this because this odd polynomial it goes to $-\infty$ here and there are only three roots, there is 2 and 5.8 and 11.6.

So, we are looking for when is it negative and that we have to do only in the 0 to 12 range because we are only looking for months of 1 year, actually even 12 is not correct because we are starting from 0 we are only going till 11. So, if the root is 11.6 then 11 is somewhere here and this is a 5.8, 6 is somewhere here and so these are the months where you get negative. Now six is not June it is actually July because Jan is taken to be 0.

So, we have July and then August is 7, then September would be 8 and October is 9, November is 10 and December is 11 and all these months you have the function being, the polynomial being lesser than 0. So, these are the months where they should reduce production and they are mentioning after March, if we looked at it before March, then you would have also have to consider January and 0 and 1 which is January and February but since they are specifically saying after March this is the set of months where the company should reduce production, thank you.

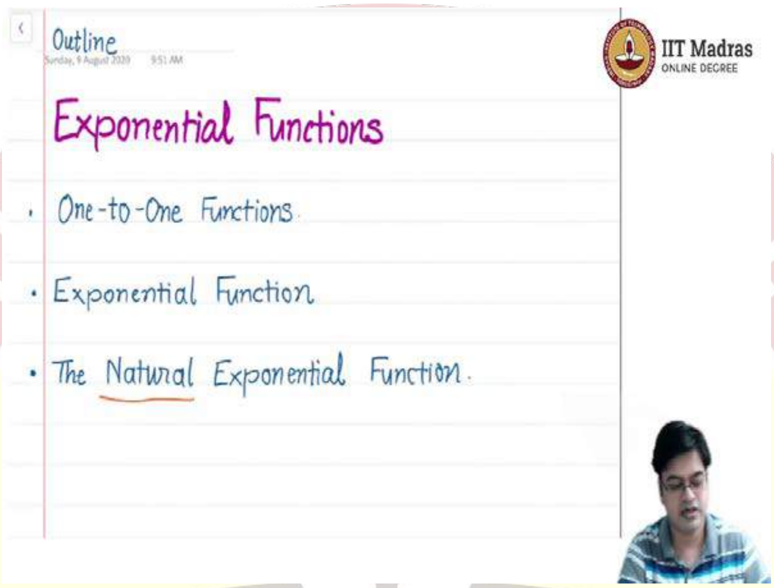
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 8.1
One-to-One Function: Definition & Tests

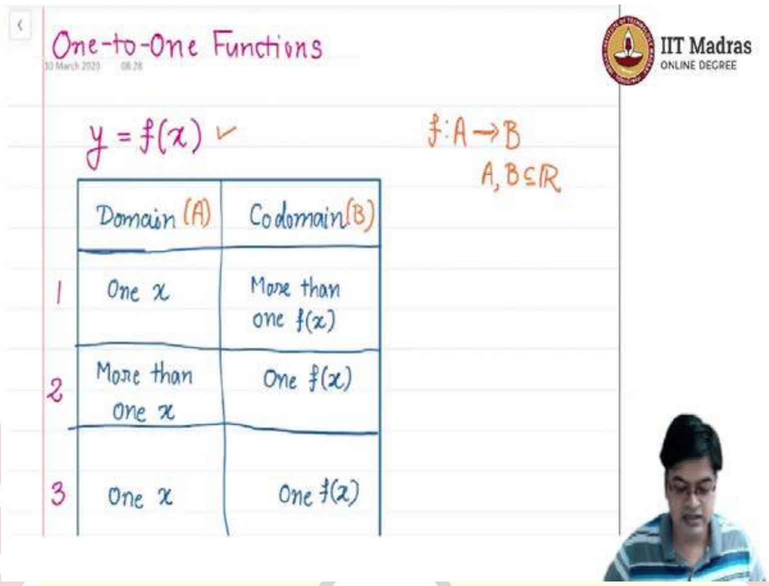
(Refer Slide Time: 00:14)



Hello students. Today, we are going to start a new unit called exponential functions and logarithmic functions. In that our goal today is to understand exponential functions. For studying exponential functions, a concept of a One-to-One Function is very much relevant.

So, we will first study that concept. Then, we will go to a definition of exponential function and properties of exponential functions as we studied for polynomial and straight line case. Then, we will define one interesting function which is called the natural exponential function, and I will justify why it is called natural and what are the properties that it shares with the mathematics world.

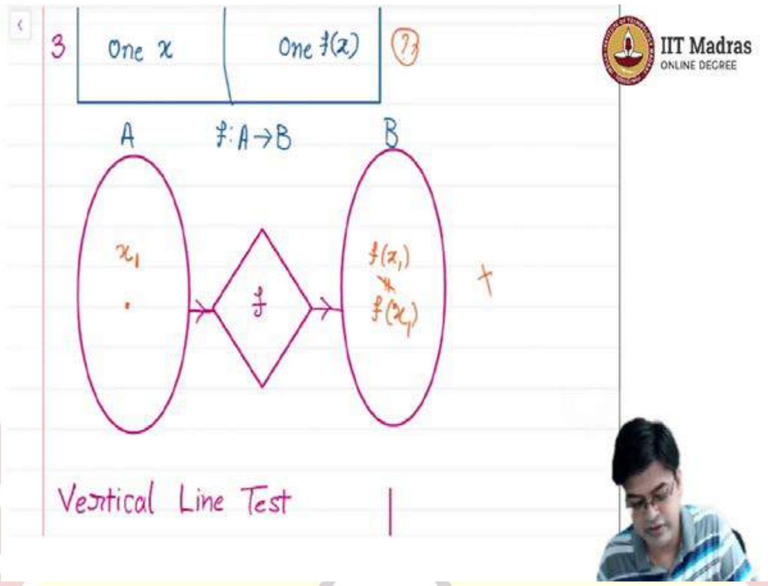
(Refer Slide Time: 01:11)



So, let us start and I think it is time to start and let us start. So, first let us go to a concept of one-to-one function. In order to define the concept of one-to-one function, let us quickly recall, what is a concept of a function? So, whenever I talk about function, I will talk about $y = f(x)$. This f when I say I need some domain let us say A and I need some co-domain B.

In general function can be defined on any two sets, but for us let us consider to simplify our understanding. So, that we will understand in terms of coordinate plane A and B to be subsets of real line, ok. So, now my domain according to this definition is A, and co-domain is B. So, now, what is the function? Function is a relation between one set to the other set. So, it is a mapping that assigns values from one set to the values of other set.

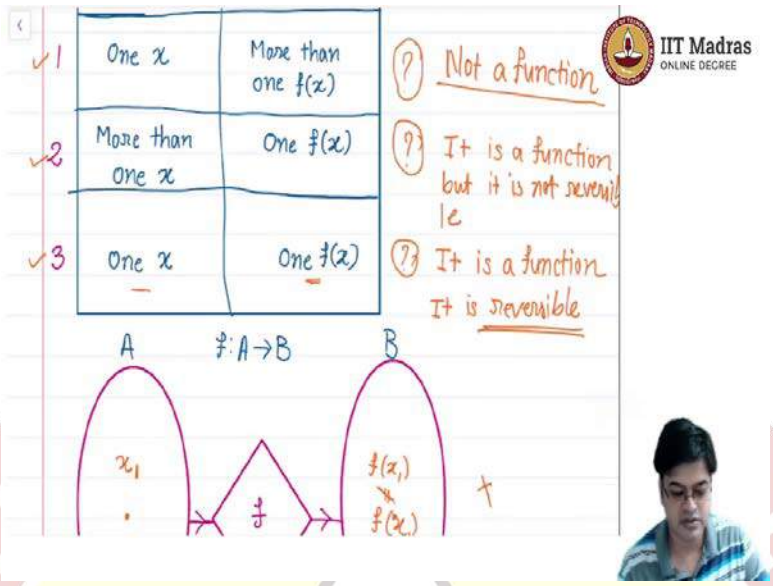
(Refer Slide Time: 02:16)



Let us describe this function as something like this, ok. So, let us see this is the set A which we will call as domain, and this is the; this is the set B that we will call as co-domain. This is A this is B. What does the function do? If we feed some value of x from this set, it will process the value and spit out or give us $f(x)$.

It is like a popcorn machine; you are feeding in the corn and getting out the popcorn. So, this is how the function works. Now, let us see, what are the cases that can happen? Suppose, for same value we are getting different outputs that, is the case, when you consider one x more than one $f(x)$, ok.

(Refer Slide Time: 03:10)

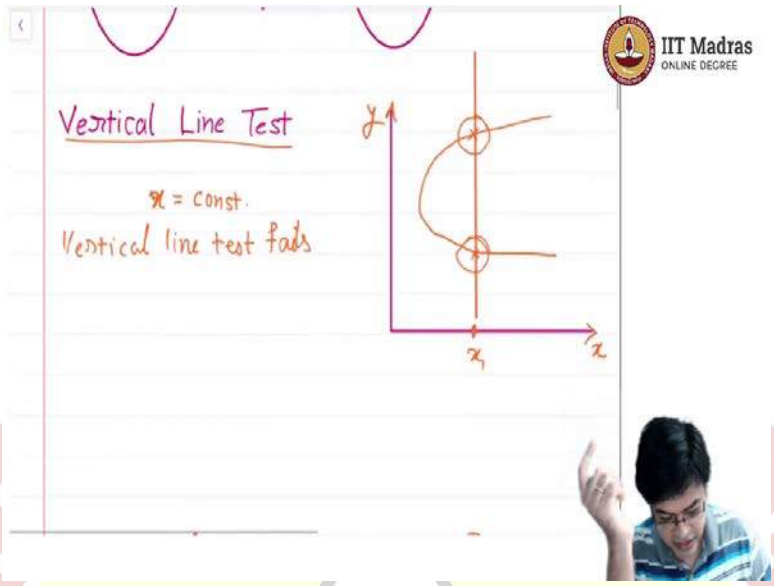


So, now, is this a function is a question mark. We will soon answer the question. Then, it can so happen that you have fed two different values and you are getting the same output, ok. Is this a function? We will answer the question. And, for every single output that you produce, there is a unique output that is produced by f, that is a one $f(x)$ only.

So, let us analyze these three again is this a function this. So, let us analyze all these three things together. So, let us start with the first case that is, one x more than one $f(x)$. The, what do I mean by one x more than one $f(x)$. Suppose, I have put x_1. If I put x_1 as a value if ones gave me $f(x_1)$, other time it gave me $f(x_1)$ which are not equal.

Is this a function? Is this a well-behaved function? It is not. So, I will say no, no this is not a function. Therefore, the first thing the first case I will say is not a function. Because, we are dealing with coordinate plane, let us understand this function with the coordinate plane.

(Refer Slide Time: 04:37)



So, what happens here is suppose, I have been I am taking the value one value x_1 on x −axis. Let us say this is x −axis, let us say this is y −axis. And now, I am taking one value of x_1. So, one I once I fed in I got something which is $f(x_1)$ here. And, other time I fed in I got something which is $f(x_1)$ here, ok. Now, what is this? This is actually while if at all some curve I have to draw, it can be like this.

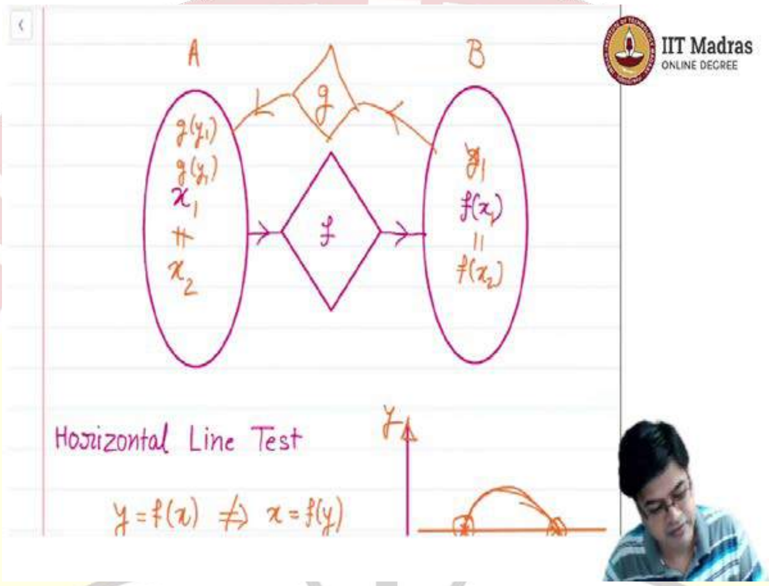
And, what happens here is for same value of x you are getting two different values. Then, your function is not well behaved, because I do not know which value will come when I fed in x_1. So, in that case how will you identify whether something is a function or not? Sometimes, if you have a graph of a function the way it is given; it is very easy to see; what is not a function.

For example, if you take a line which is $x =$ constant. If you take your line which is $x =$ constant and if I draw that line vertically. For example, let us say here this is the line $x =$ constant; $x = x_1$ in fact I have drawn. So, if I draw this line vertically, then I can see that there are two points at which, this line intersects the graph of a function.

When such a thing happens, we say a vertical line test that is, this is the vertical line right it is a parallel to y −axis. Therefore, this line passes through two points; that means, there is something fishy about the function and we will use this as a vertical line test. And say that, because vertical line test fails, this particular function is not a function sorry, vertical line test fails and hence, this cannot be a function.

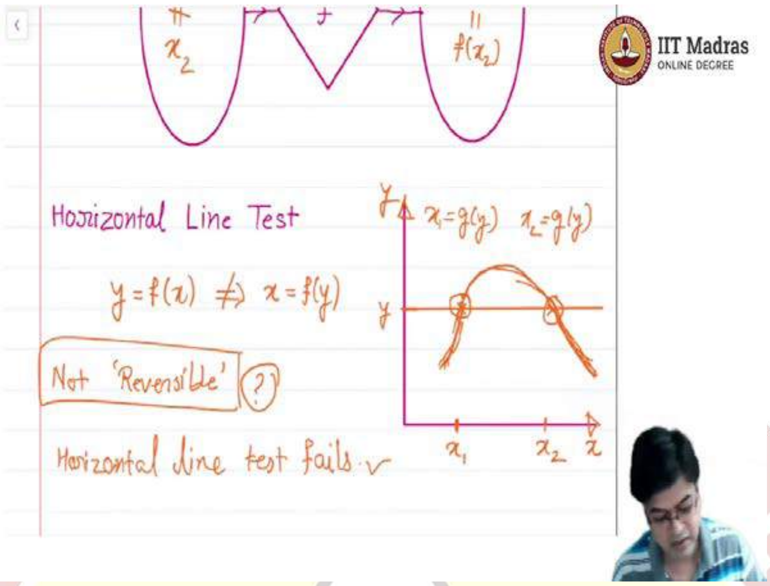
Can you imagine such a function where can you imagine such a relation where this is not a function. For example, ok. Let us not imagine right now. Let us come to the next case and generalize it to the other set up. So, now let us see here. So, first thing is not a function that is very clear. Now, let us take the second case which is more than one x and one $f(x)$, ok.

(Refer Slide Time: 07:41)



So, let us understand it on a paper. So, they are more than one x, x_1 and x_2, both are not equal, but somehow, they give the output which is equal, ok. In such case, what happens? So, as usual general this is domain set A, this is a co-domain B; f is processing unit and I have processed I have given I have fed in two different values of exercise, but the output produced is same.

(Refer Slide Time: 08:27)



So, is this a function or not? The answer is this is a function. And, in this case let us see, how it happens? So, let us have some understanding about x_1 and x_2. So, naturally $x_1 \neq x_2$ and I got $f(x_1)$ which is this and I got $f(x_2)$which is this, ok. Both are same.

And, then how will the function look like? I can join a curve like this, where the function actually passes through these two points and I have a curve like this. So, if this is the curve then these two points are same. And, do you call this as a function? Yes, we call this as a function. Based on our understanding of the function lecture we call this as a function. In this case, something interesting has happened. Let me analyze it in a more thorough manner.

For example, when I considered the first case let me go to the first case. I have drawn a vertical line and I said that, because of this vertical line I can say this is not a function; I can say this is not a function. Now, the similar graph has appeared over here, but now if I draw a horizontal line I have a horizontal line, which passes through two points and I am saying it is a function, correct?

So, if I rotate this graph by say 90 degrees and flip it over then, what I am getting is a graph similar to this function. So, this actually helps me in understanding that, if I want to write y as a function of x, I am able to write it. But, if I want to write x as a function of y that is simply just flip this by rotate this by 90 degrees and flip the y axis. That will give you the exact understanding of the picture. And, from this to this I cannot go, ok. And therefore, the horizontal line if I draw it passes through more than one points.

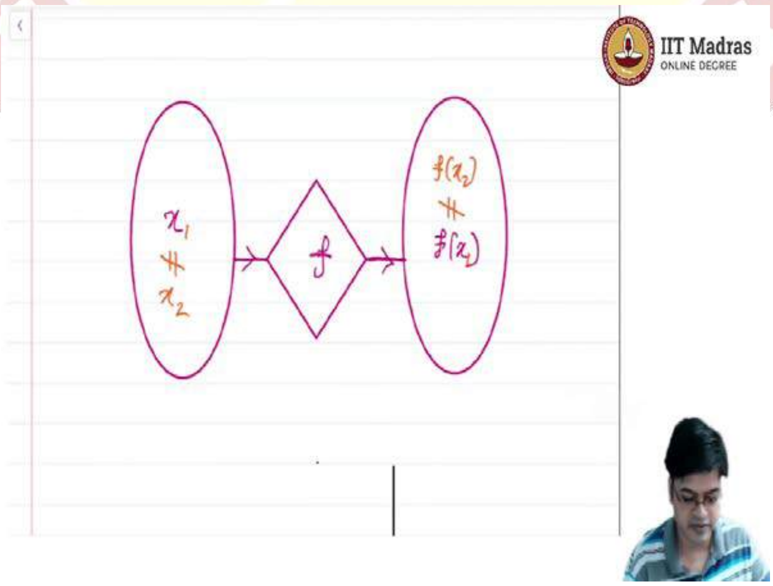
So, what will happen is; suppose, if I take; if I take x_1 here, if I take a point in this domain y_1. And, if I try to setup another processing unit let us say g, and if I feed that $y_1$into g, what I get if I use this f, I will not get something which is similar that is I will get something called $g(y_1)$ once. And, if I feed in again y_1, I will get something else as $g(y_1)$. That is what is happening.

For example, here if you locate this point, this point it is say y. Then, once you feed y into g then, $g(y)$ you will get as x_1 and if you feed it other time $g(y)$, you will get as x_2. Something interesting happens. What I am trying to do is, I am trying to reverse the function and I an easily say that this function is not reversible.

If this is a interesting point which will help us in gaining more understanding of exponential functions. So, if this function is not reversible and a horizontal line test actually detects whether a function is reversible or not; so, this horizontal line test fails; horizontal line test actually fails in this case.

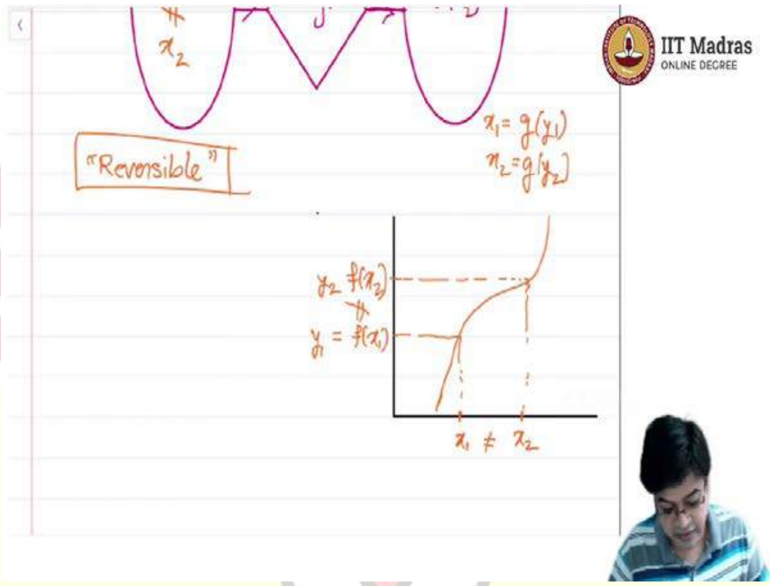
So, what happens here? Here, when you applied vertical line test that is case 1, it was not at all a function. Here it is a function, but our conclusion is it is not reversible. Let us look at this 3rd case, where only one x is there and one $f(x)$ is there. So, here is x, $f(x)$.

(Refer Slide Time: 13:42)



So, if I substitute x_1 and x_2 as two different values, then I will get through this processing unit, I will get $f(x_1)$and I will get $f(x_2)$ and both of them will not be equal to each other for $x_1 \neq x_2$. That is the only way it can happen right; one x, one x to one $f(x)$. So, if I take different $f(x)$ I will get different values, ok.

(Refer Slide Time: 14:12)



So, what will be a typical behavior of such functions? Let us try to figure out. Let us say this is the; this is one function. So, is this function one-to-one? The answer is yes, because if I take x_1 x_2 here this is $f(x_1)$, this is $f(x_2)$. So, for every one, if $x_1 \neq x_2$, $f(x_1)$ is not going to be equal to $f(x_2)$. Such function is called one-to-one function. And is it a function? It is a properly defined function, yes.

So, now, I can summarize using this as, this is a function. In earlier case, we actually characterized whether it is reversible or not. What can you say about this new function that you have defined? This function; so, now because for $x_1 \neq x_2, f(x_1) \neq f(x_2)$. So, if I start this as y_1 and if I start with this as y_2, I can easily retrace back x_1 x_2 that is g there will be some function g such that, $x_1 = g(y_1)$ and $x_2 = g(y_2)$.

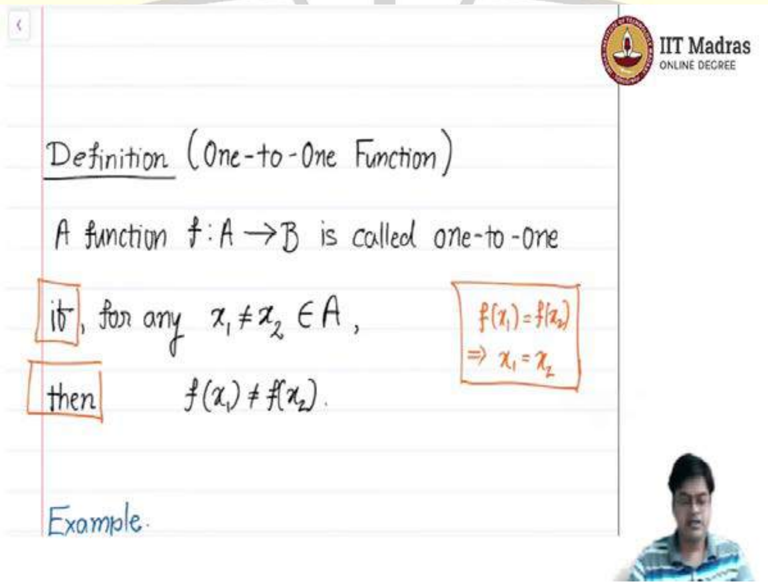
So, this function in some sense is actually reversible. We will come to this point in more detail in the next section. Right now, remember that the function that is one-to-one is reversible and it is reversible. We need to be more precise and I will give you a word for reversible function in the later lectures. But right now, it is an important observation that there are three cases. These three cases actually deal with a function.

The first one is not a function; first one is not a function, why? Because we had dealing with the coordinate plane. The vertical line test fails when then you pass a vertical line that vertical line will pass through two points or more than two points, then it is not a function.

So, then we said the second case. The second case was more than one x and one $f(x)$. In this case, what happened is we were able to find a function properly, but we were not able to revert the function. So, we said the function is not reversible. And, on what basis we have said? We have said on the basis of horizontal test, ok. And, we have related this with our first case that is, if the function is not reversible then the other part that is g is not also is not a function as well.

Third case where we have one x and one $f(x)$ we showed that it is a function and it is reversible also. This brings us to a question that, one-to-one functions how easy are they to identify; how easy are they to identify? So, let us properly define a one-to-one function. One in the domain, one in the co-domain and there is a clear cut association that is given by two.

(Refer Slide Time: 18:23)



So, our definition; give a function f is said to be one-to-one if, for every $x_1 \neq x_2$ which belong to the domain of a function, $f(x_1) \neq f(x_2)$, ok. The other interpretation is, if $f(x_1) = f(x_2)$, this should imply $x_1 = x_2$ this is the other interpretation of definition, but we will use this as a definition. But, sometimes it may be difficult to prove this thing in that case, you can prove the one written in the orange.

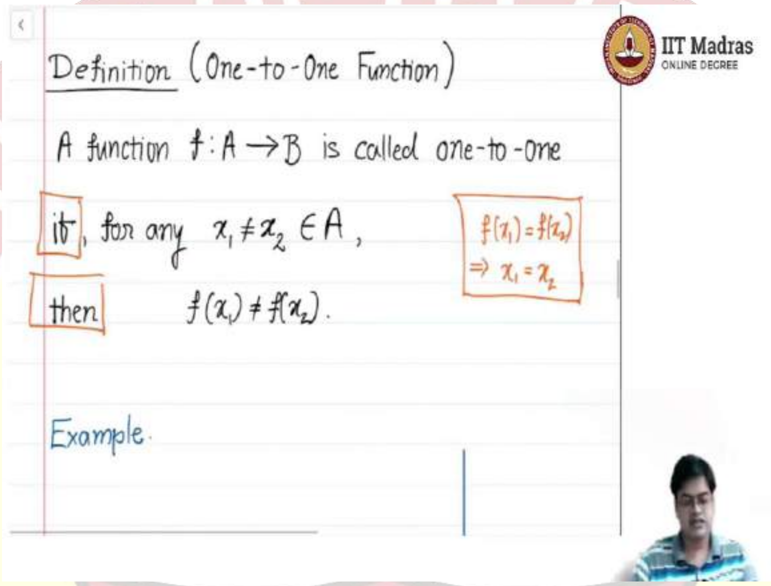
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

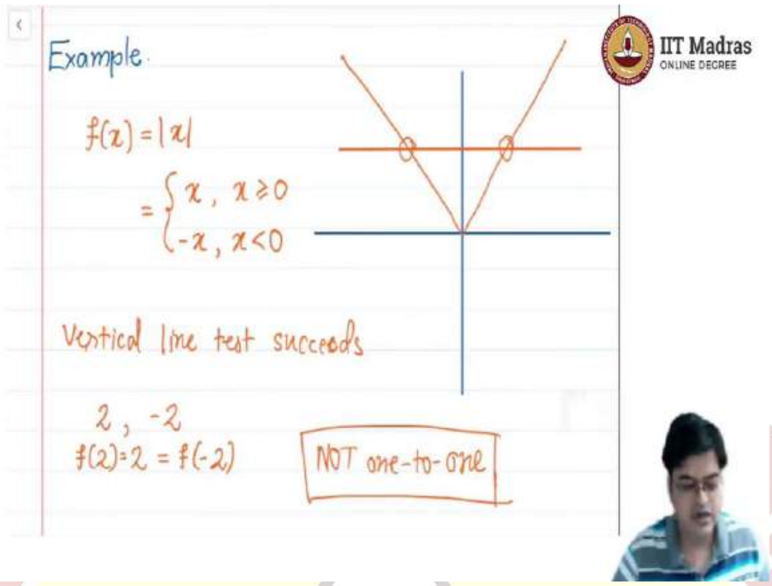
Lecture – 8.2
One-to-one Function: Examples & Theorems

(Refer Slide Time: 00:14)



So, let discuss some examples of a functions that are one to one and not one to one. So, for this let us first take $f(x) = |x|$; is this function one to one?

(Refer Slide Time: 00:20)

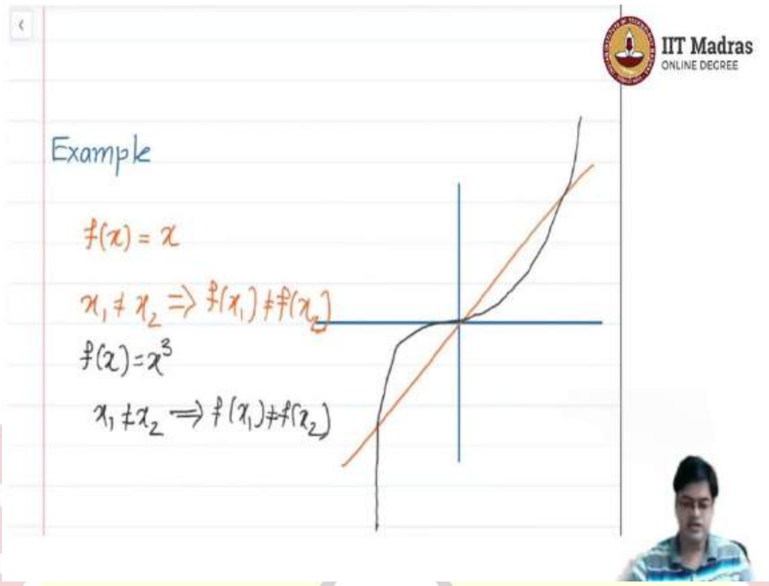


Try let us try to answer this question. So, let me write this function properly. So, if $f(x) = x$ for $x \geq 0$ and $-x$ for $x < 0$. So, it is actually a straight line on a passing through the origin like this 40 at a 45 degree angle and the $-x$ is this line ok. So, it is a V shape 90 degrees V; so, is this function one to one? First of all let us let us not take the argument, first of all is this a function $(x) = |x|$?

Pass a vertical line, take a vertical line and pass it through this; is there at if there is any point where two points more than one points pass through this function pass through that line then it is not a function. So, vertical line test is successful therefore, it is a function. Vertical line test says that it is a function succeeds and we know it is a function ok. Now, the question is the function one to one? Right. So, you pass a horizontal line. So, let me pass one horizontal line somewhere, let us take this horizontal line. Now, is the function one to one?

For $x_1 \neq x_2$, I got the same $f(x)$ ok. So, how will I prove it is not one to one? Let us take a value which is say 2 and -2; these are the two values, $f(2) = 2 = f(-2)$. Therefore, this is not going to be a one to one function. So, it is not one to one function. So, our conclusion is it is not one to one function. Then do we know functions that are one to one?

(Refer Slide Time: 03:05)



So, since we have taken $f(x) = |x|$, let us take a function $f(x) = x$; is this function one to one? It is a straight line passing through the origin, is this function one to one?

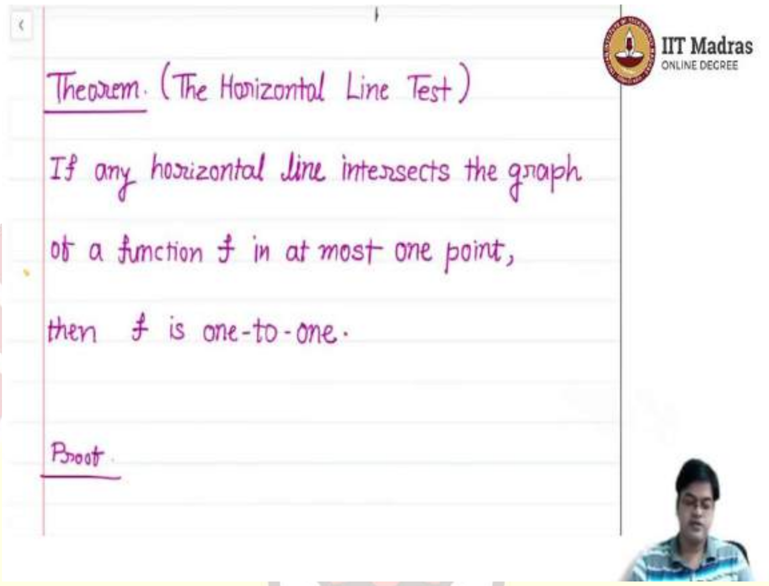
Let us take horizontal, first let us check whether this is a function, take a horizontal line, pass it through this pass it horizontally, the line parallel to x −axis. So, just drag x −axis up and down; do you see any point touching more than one point? No. So, it is a valid function; then, sorry yeah you have to pass the vertical line first ok.

Start with $f(x) = x$, take a vertical line which is y −axis, slide it to the left, slide it to the right. Do you see any where it has more than one points? No. So, it is a valid function. Then take a horizontal line, pass it from the top to bottom; see if you are getting any two points together for on that line; no. Therefore, this function is actually one to one because $x_1 \neq x_2$ implies $f(x_1) \neq f(x_2)$ which is more or less expected right.

Because $f(x) = x$ therefore, $x_1 \neq x_2$ will give $f(x_1) \neq f(x_2)$. So, what about it is an exercise then what about if you take a cubic functions? So, cubic function will pass like this sorry, it is not a correct diagram of a cubic function. So, cubic function let us change the color as well, cubic function will have something like this, symmetry will be retained and then this will go down.

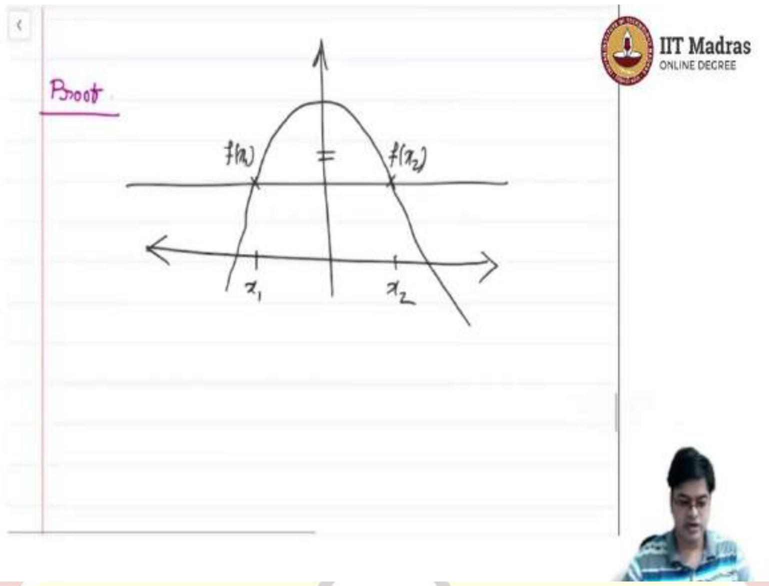
So, if this function, now check whether this function is one to one or not. Again the exercise is very similar, pass a let the x − axis go up and down, see if you are finding any two points together. So, let us say this function is $f(x) = x^3$ and now you can easily make out that for $x_1 \neq x_2, f(x_1) \neq f(x_2)$. So, again through horizontal line test, I have detected that the function is one to one.

(Refer Slide Time: 05:59)



So, let us write this particular test as a theorem. If any horizontal line intersects the graph of a function in at most one point, then the function is one to one ok. So, then what we will show here, if you want the proof of this what we will show here is if the function is not one to one then it will intersect some horizontal line will intersect the graph of a function in more than one point ok. So, that is very easy to prove. So, I will prove it graphically.

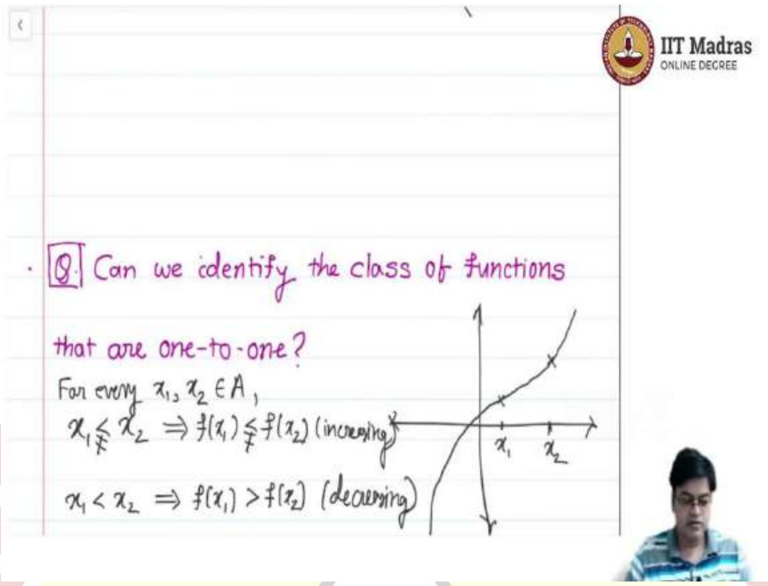
(Refer Slide Time: 06:41)



So, if the function is not one to one, let us say this is x −axis, this is y −axis. If the function is not one to one, I can take this point and call this as x_1 and I can take this point as call this as x_2. This is how I can make function not one to one and then pass a curve passing through these two points and pass the horizontal line over here which we have done several times now by now.

And therefore, $f(x_1)$ and $f(x_2)$ are same, they both are same. Therefore, the function is not one to one, that essentially proves the point that if a horizontal line intersects the graph of a function in at most one point then f is one to one good.

(Refer Slide Time: 07:37)



So, we are good to go now. Next thing that we will come is can we identify the class of functions that are one to one? So, what class of functions can you immediately think are one to one? For example, we have also seen some functions like if $x_1 \leq x_2$ then $f(x_1) \leq f(x_2)$ or let us not put this strict equality; let us put this way strictly increasing.

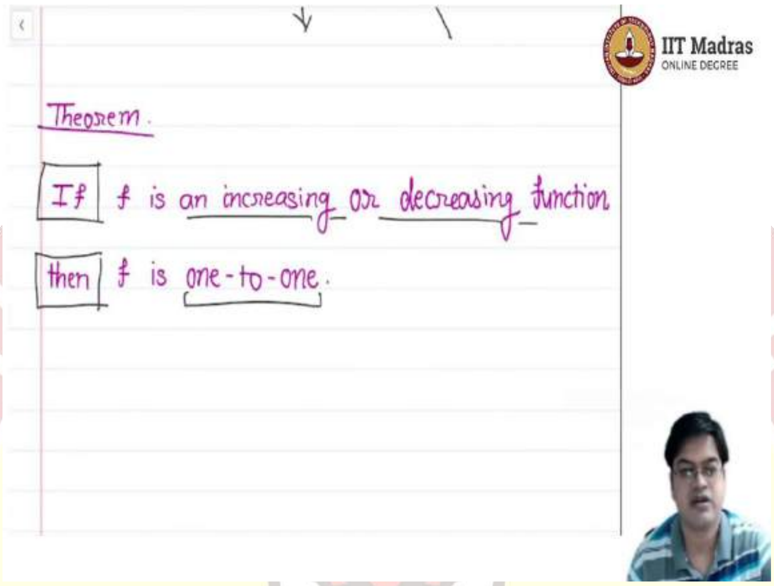
So, what does what do I mean by ok; let us can we question is can we identify the class of functions that are not one to one? So, I can; obviously, think of function of this form $x_1 < x_2$, $f(x_1) < f(x_2)$. Let me plot it and then the my imagination will work fine. So, this function is something like if x_1 is to the left of x_2 then $f(x_1)$ should always be less to the left of $f(x_2)$

Or, if you are plotting it on the y −axis then $f(x_1)$ below $f(x_2)$, this is the intuition and you can draw line joining these two points. Let it go ahead and this is true for every x_1, x_2 belonging to A this is true; then I am done.

But, this function have a name that is they are called increasing functions ok. In a similar manner, if I multiply this function with minus sign. Then I will get a function which is decreasing function and that can be written as $x_1 < x_2$ employs $f(x_1) > f(x_2)$ and this is called decreasing function.

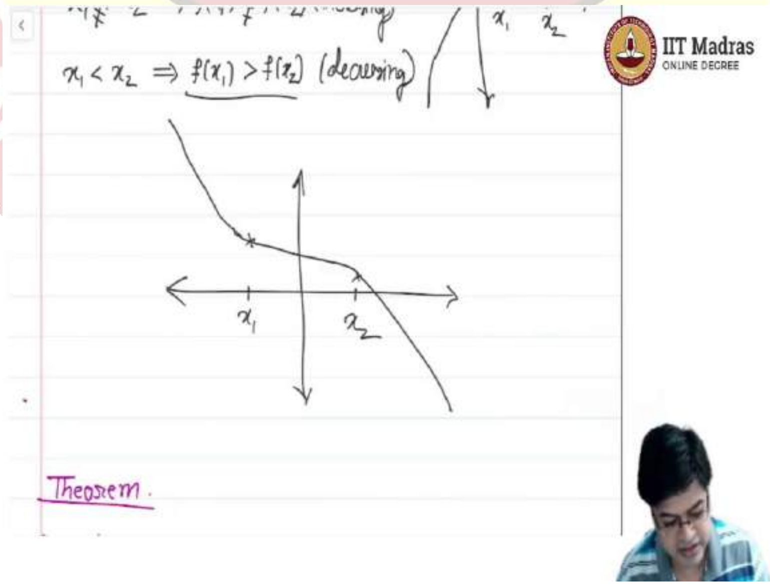
Now, you look at any increasing function and apply your horizontal line test. What is the horizontal line test? Just now we have seen that if you take the horizontal line, roll it across the axis across $y-$ axis and there should not be more than one point intersecting that line at any given point in time ok. So, this increasing function and decreasing function will satisfy this phenomena.

(Refer Slide Time: 10:47)



And therefore, we can easily write this as through horizontal line test that, if f is an increasing function or a decreasing function then f is one to one.

(Refer Slide Time: 10:59)



Let us see one decreasing function as well. What happens when the function is decreasing? As I go from left to right there is a x_1 is here, x_2 is here. As I go from left to right, I get x_1 here and now according to the condition $f(x_1) > f(x_2)$. So, it will be somewhere here and I can have a curve passing through this point in this manner ok.

This is true for every x_1 and x_2 belonging to the domain. And therefore, using our line test, horizontal line test we can easily see that the function whether it is increasing or decreasing, they are one to one. This gives us a big class of functions that are one to one. So, one to one functions are not rare to find, one to one functions are abundant in nature and they are reversible as well.

With this insight, we will go to our next topic which is exponential functions.

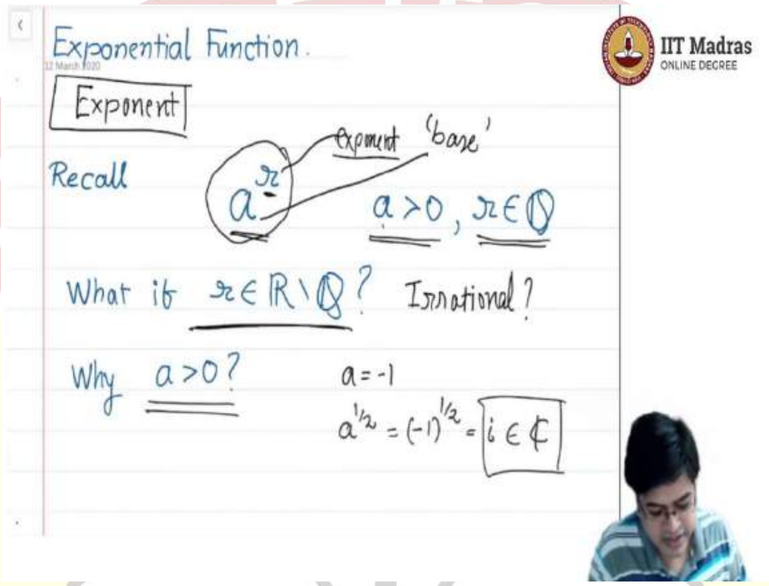
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

Lecture – 8.3
Exponential Functions: Definitions

(Refer Slide Time: 00:14)



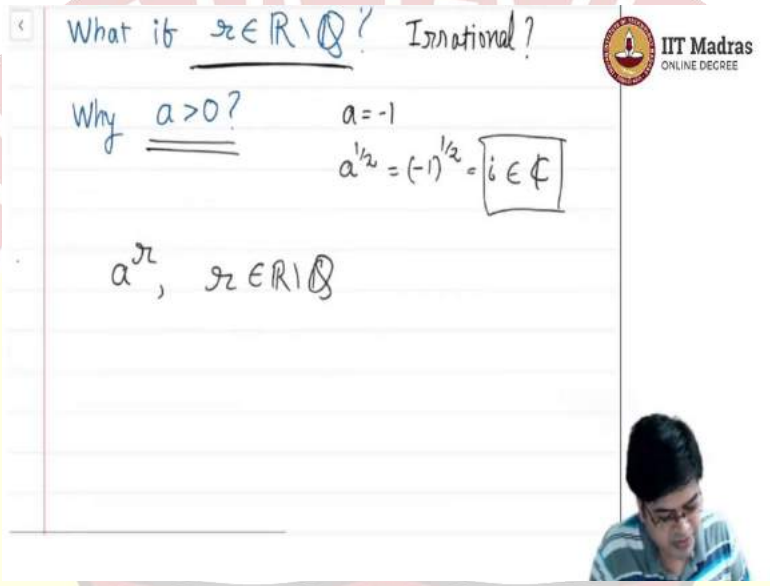
Welcome back. Next topic is Exponential Function. So, in this topic, what we will see is first we will identify with some known terminology that is exponent. We have already know seen this exponent. Where we allowed integer powers and then while defining the exponents, we allowed rationals also. So, when we defined exponents, they were of the form a^r and we always assume $a > 0$ and $r \in \mathbb{Q}$.

Now, I want to define an exponential function. So, as the name suggest, exponential has to do something with the exponent. So, what we are doing here when I am considering a function of this form, I am raising something, some number to the power of a where a will be popularly called as base and r is the exponent. So, this is base and this is exponent.

Now, if I want to define exponential functions on real line, then it is mandatory for me to define this a^r for $r \in \mathbb{R} \setminus \mathbb{Q}$. This is real line minus set of rational numbers that is I am talking about set of irrational numbers. So, I do not know as of now what is a definition of exponent form of set of irrational numbers ok.

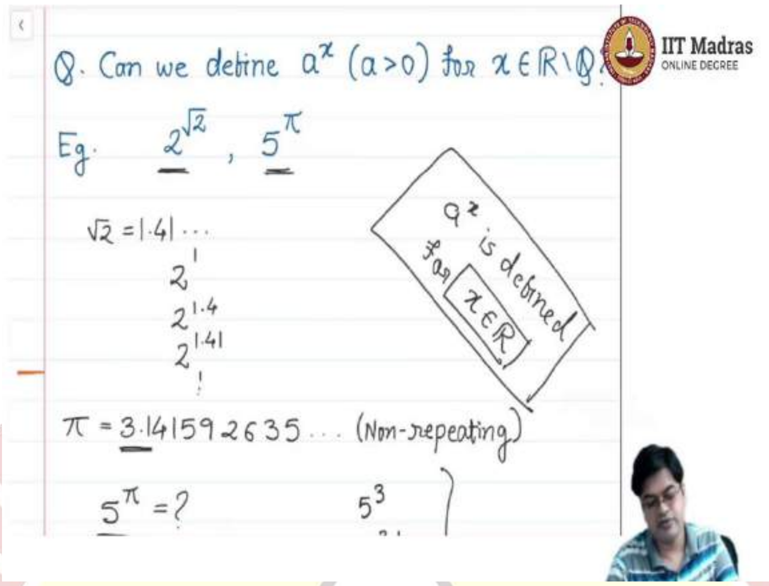
The next question that we have seen is why is $a > 0$ that for which you know the answer. Let us say $a = -1$ now$a^{1/2} = (-1)^{1/2} = i$ which belongs to complex set of complex numbers, but I do not want to deal with complex number so, I am avoiding a to be greater than 0. In general, you can define a to be a negative number and then deal with complex numbers. We do not want to indulge into that conversation. So, I do not; I do not want to define $a < 0$.

(Refer Slide Time: 03:12)



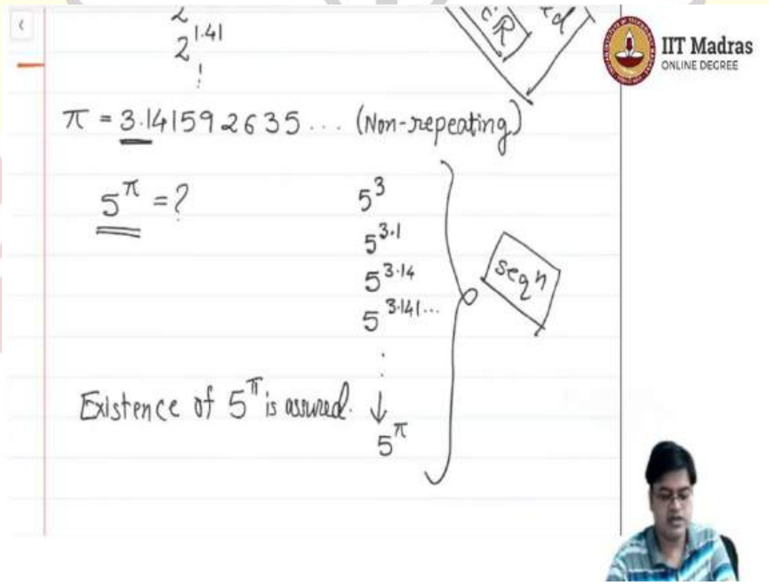
So, $a < 0$ is eliminated now the question is $a^r$and r belongs to irrational where $r \in \mathbb{R} \setminus \mathbb{Q}$ what will happen in this case? Or how will I define rational number?

(Refer Slide Time: 03:24)



To be precise, let us ask a question that is can I define $2^{\sqrt{2}}$ or 5^{π} ok. So, in this case, there is no direct way to answer this question, but I will definitely have a strategy which is a calculus-based strategy to answer this question.

(Refer Slide Time: 03:52)



Let us consider this π and the value the numerical approximation of π is actually we all know π is an irrational number and 3.14592635… and this thing is non repeating it will continue till infinity right. So, now, what I need to understand is from what I know, can I define the number 5^{π}? So, anyway I cannot define it accurately right now.

So, based on my understanding, I am asking you a question that is 5^3? Right if so, then next question is $5^{3.1}$ defined? So, what I am doing here is this if yes, can I define $5^{3.14}$? Now, you remember all these a approximations are actually a rational approximations 3 is a rational number, 3.1 is 31 by 10 which is again a rational number, 3.14 is 314 by 100 which is again a rational number and I can go on like this there is 3.141 and so on.

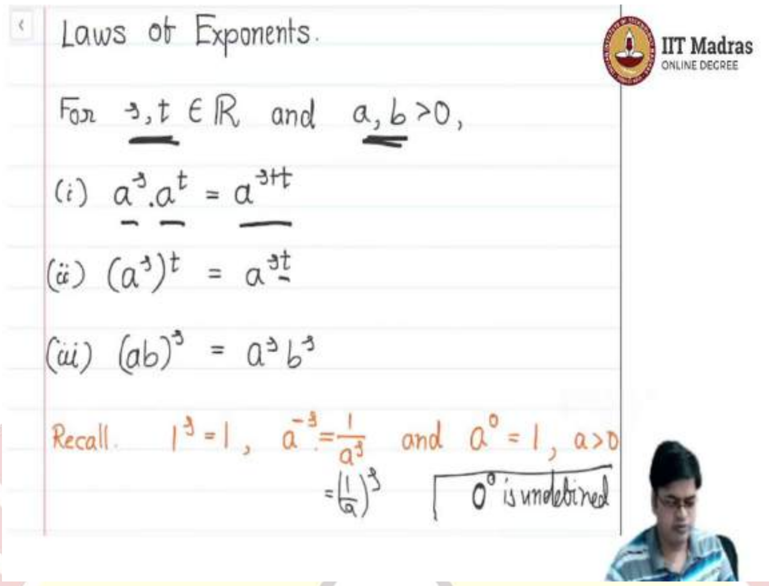
So, if I continue this way, I will reach somewhere; I will reach somewhere and that somewhere I will call as 5^{π}. So, in principle, I can actually define a raised to irrational number. This you will study when you will study a topic of sequences which is outside the scope of this syllabus. So, we will assume that you have to trust me on this that 5^{π} is well defined.

In a similar manner, you can do an exercise for $2^{\sqrt{2}}$. So, $\sqrt{2} = 1.41$ ... and something. So, again you will go with 2^1 is defined $2^{1.4}$ is defined, $2^{1.41}$ is defined and so on and you will reach somewhere that is $2^{\sqrt{2}}$.

So, this way we are very clear that a^x is defined for $x \in \mathbb{R}$. This sets up the platform for defining an exponential function, this is very important a raised to x is well defined for $x \in r$.

This answer is given by convergence of sequences which is outside the scope of the syllabus, but we know that it exist for sure. So, I am guaranteeing the existence of 5^{π}; existence of 5^{π} is assured. In case you are interested, you can take a basic course in analysis or; in analysis or calculus where you will study these things ok.

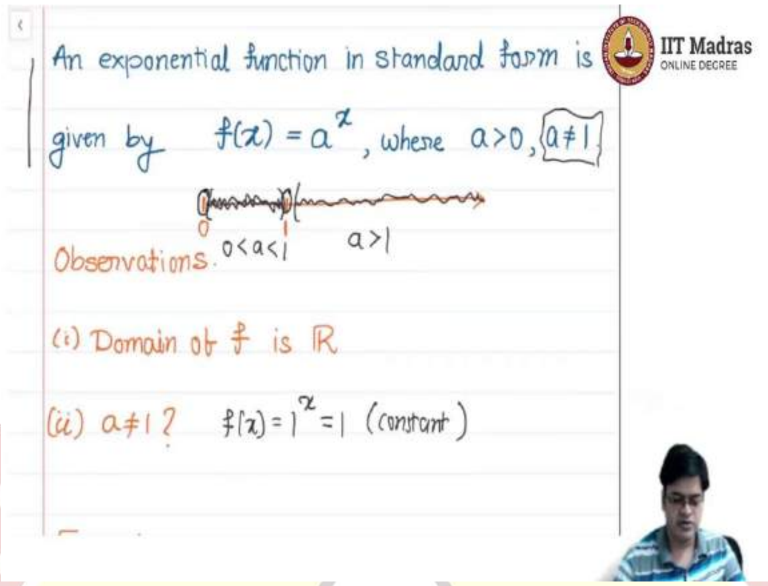
(Refer Slide Time: 07:27)



So, now let us go to let us recall all of these laws you already know simple laws of exponents. Earlier, we have we knew the laws of exponents for only rational numbers. Now, we are talking about the real numbers. So, $s, t \in \mathbb{R}$, $a, b > 0$, s and t will play a role of exponents, a and b will play a role of bases ok. So, then it is very easy to prove you might have proved. $a^s a^t = a^{s+t}$.

Remember here, product here is becoming addition here these are crucial points $(a^s)^t = a^{st}$. So, a raised to operation is becoming a product here. $(ab)^s = a^s b^s$ and then; obviously, you need to know that $1^s = 1$ for every $s \in \mathbb{R}$, $a^{-s} = \frac{1}{a^s} = \left(\frac{1}{a}\right)^s$, $a^0 = 1$. Remember where your $a > 0$ because 0^0 is undefined ok.

(Refer Slide Time: 09:18)



So, with this understanding, we have revised laws of exponents which will which we will use the left and right. So, you better remember all these laws and therefore, we are ready to set a framework of exponential function. So, here is our definition. An exponential function in the standard form is given by $f(x) = a^x$, where $a > 0, a \neq 1$. These are new condition that we have introduced.

We have seen why $a > 0$, but here they are saying $a \neq 1$. So, this needs further analysis, we will analyze it in due course. So, right now, if you look at the values of a, $a > 0$; that means, all these values are allowed and $a > 1$; that means, all these values are also allowed. Bearing the values 0 and 1 right.

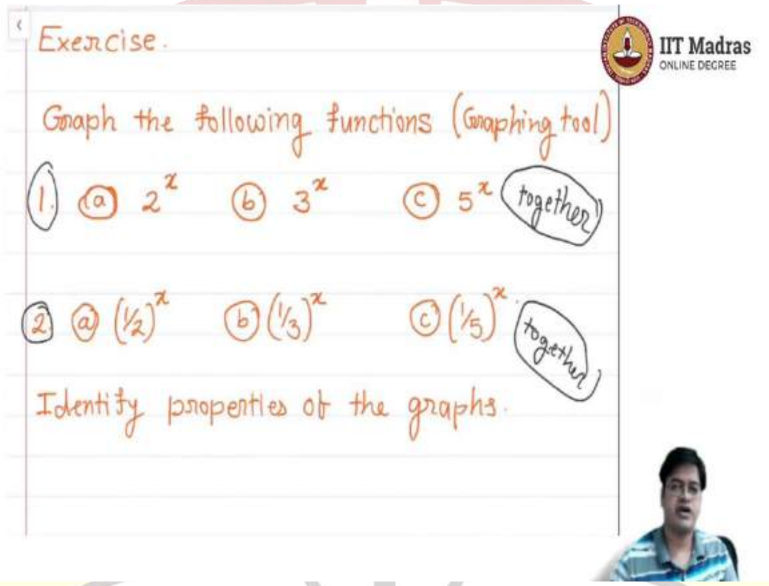
So, the first from the definition, the first observation that you can figure out is because you have bared the value 0 and 1, the function $f(x) = a^x$ will have a domain which is entire real line. For every $, x \in \mathbb{R}$, we should be able to compute a^x ok.

Then, let us analyze this is then observation: why $a \neq 1$? Let us put $a = 1$. So, $f(x) = 1^x$, but from the laws of exponent what you know? $1^s = 1$.

Therefore, $1^x = 1$ in fact; it is nothing, but a constant function. I am not interested in handling a constant function right which nothing, but a horizontal line $y = 1$ is the graph of a function; I am not interested in this. So, let us not call this as exponential function that is what we are saying in the definition.

So, hence forth, we will never talk about $a = 1$, $a = 0$ or $a < 0$. So, if you have a real line, you will have an expression of this form where you are talking about this interval, open interval and this interval, which is an infinite interval. So, you have two characterizations which is $0 < a < 1$, $a > 1$ these are the two characterizations that you got over this thing.

(Refer Slide Time: 12:18)



Now it will be interesting to use some graphical tool and see what are some functions of this kind look like. So, here is an exercise that I will give you. Use some graphing tool like Desmos and plot these functions together. For example, you plot the functions given in 1 using Desmos we just put $f(x)$ is equal to this, $f(x)$ is equal to this and this and plot all these three graphs together without any understanding about the behavior of the function you plot all three of them together.

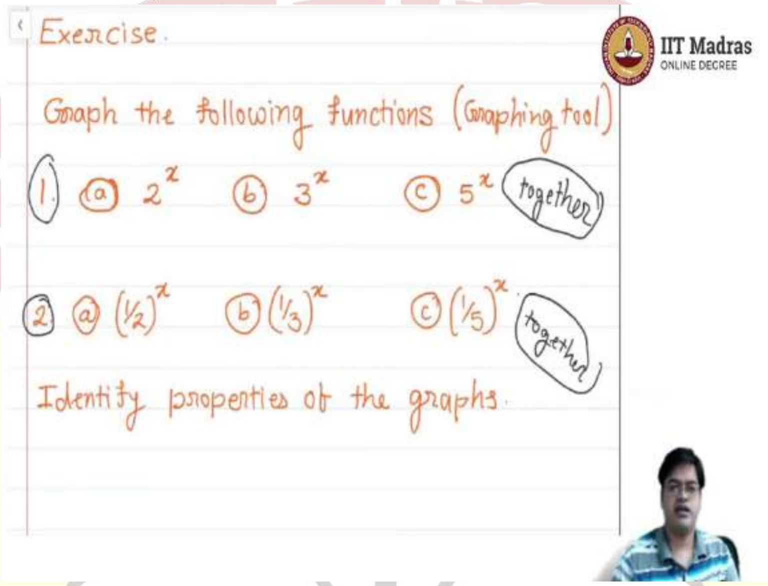
Then, use the 2nd graph and put all these three things together. Identify the properties of the graph that is through which points they pass through is there any difference in the graphs of 1 and 2. So, identify all these properties like we did in polynomials and after doing that again return back to this video and we will see some of the functions that are given here by a graph and we will analyze those functions. So, right now you pause the video, come back in the next video.

IIT Madras
ONLINE DEGREE

Mathematics for Data Science 1
Prof. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras

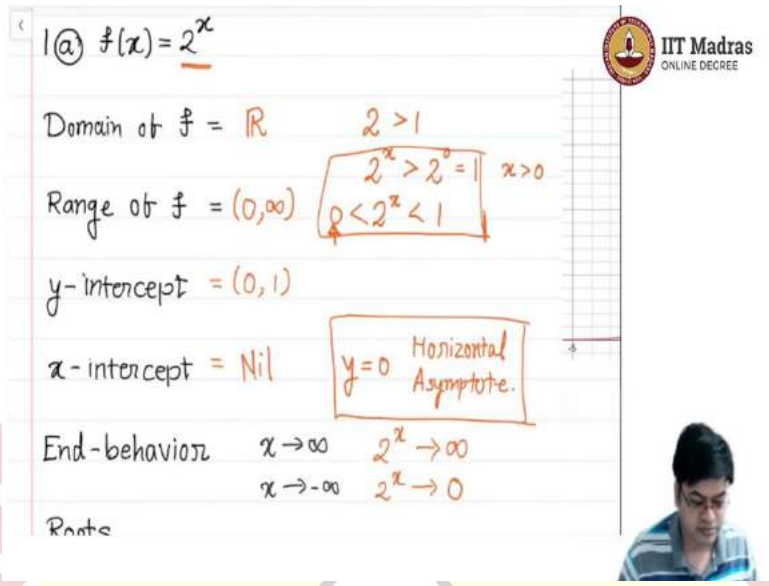
Lecture – 8.4
Exponential Functions: Graphing

(Refer Slide Time: 00:15)



Welcome back. So, I hope you must have done your exercises and you must have developed some understanding about the exponential functions. Let us try to collect recollect that understanding through 2 examples given here.

(Refer Slide Time: 00:31)



So, let us first take 1 a which is $f(x) = 2^x$. If you have used DESMOS, you must have got the figure of the function. But prior to receiving the figure of the function, let us see what should be the domain of a function.

We have already discussed in greater detail that the domain of a function can be a $\mathbb{R}$, entire real line. Now, if you look at this function which is 2^x, this $2 > 1$ and the $2^x > 2^0$ which is equal to 1, $2^x > 2^0$ whenever x is positive correct.

Now, because $x > 0$, then $2^x > 2^0$ ok. So, if $x < 0$, what will happen? 2^x, when $x < 0$ will always be less than 1. This is also possible. But when this 2 raised to; can this 2^x become negative? No. So, it is always greater than 0.

So, if you have this understanding, then you can easily write the function has a range which is $(0, \infty)$. So, there is a split from when you consider a point 1, there is something happening at point $(0, 1)$ right. What is $(0, 1)$? $(0, 1)$ actually is an y −intercept ok, something is happening at $(0, 1)$ because I have put 0 here for then it is I am getting 1.

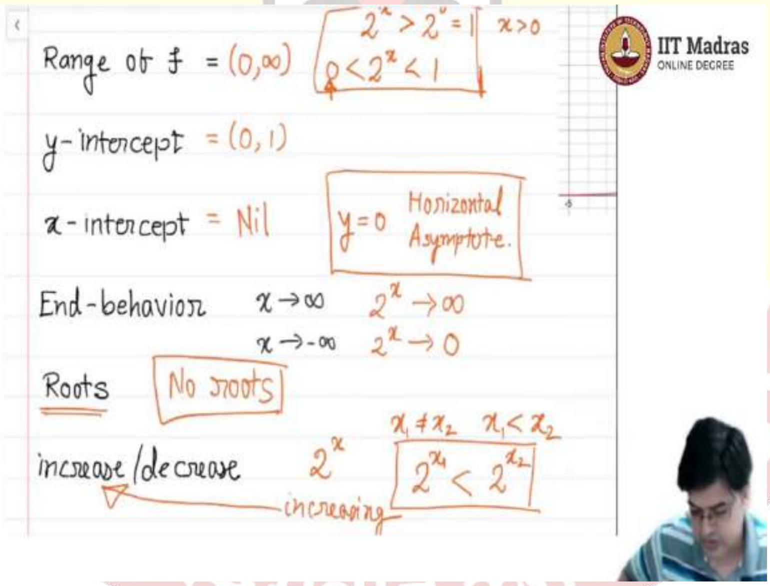
So, $(0, 1)$ is also y −intercept and there is something happening which is going below 0. Is going below 1, your graph is going below 1 and therefore, this particular thing is going down, but it never goes below 0. This is an interesting fact because if you consider 2^x, it never goes below 0.

It cannot go to a negative number. Therefore, will it touch the $X -$ axis? It will not touch $X -$ axis. In fact, $x -$ intercept is nil ok, but it is approaching 0. So, the something that is approaching 0, so $x -$intercept is actually it will never touch it; but it will actually go along that line. So, this $y = 0$, it will touch at infinity ok. So, such a thing, we call as horizontal asymptote ok.

So, such a thing you call as horizontal asymptote. So, with this understanding, these are the things that I can make out directly without looking at the graph. So, let us now look at the graph ok, before going to that, let us see what happens to the end behavior. End behavior of a function as $x \to \infty$.

So, as 2^x, you consider a function 2^x as x increases, this also increases. In fact it increases at a rapid rate than x. So, this also should tend to infinity and as $x \to -\infty$, we have already figured out $y = 0$ is the horizontal asymptote. So, 2^x will actually go to 0 ok.

(Refer Slide Time: 05:29)

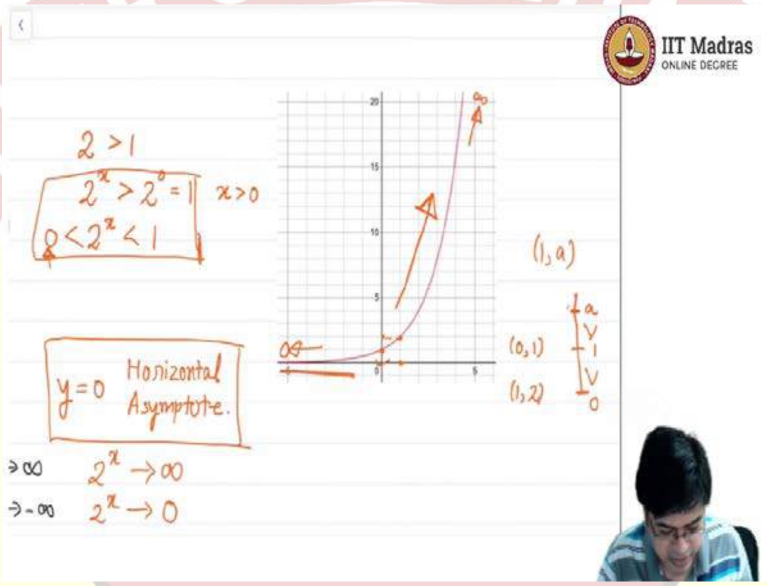


Then, the question that we used to quantify while considering the function, what are the roots of this function. So, do they have any roots? In fact, using graphical method, it is very clear that it never touches 0. So, there are no roots and the functions increase and decrease.

So, the domains of increase and decrease like polynomials, we studied domains of increase and decrease; but here, I think my claim is no need to identify the domains of increase and decrease. Why? Because you look at a function 2^x, let us take $x_1 \neq x_2$ or $x_1 < x_2$, without loss of generality, we can take this. Then, what can you say about 2^{x_1} and 2^{x_2}?

See $x_1 < x_2$, so naturally if it is raised to the power 2; 2^{x_1} and 2^{x_2}, this relation should hold. So, what I am saying is the function is actually an increasing function and increasing functions are 1 to 1. Therefore, I do not have any doubt that the increase and decrease, it is only increasing; throughout the real line, the function is only increasing.

(Refer Slide Time: 07:19)



So, let us look at the graph of a function $f(x) = 2^x$. Let us identify the points. So, here you can identify a point right. So, this point we have seen as $y -$ intercept and that point was (0, 1) right. Then, the one in this case, let us look at this point which is 1 and where will it go? It will actually tell you 2.

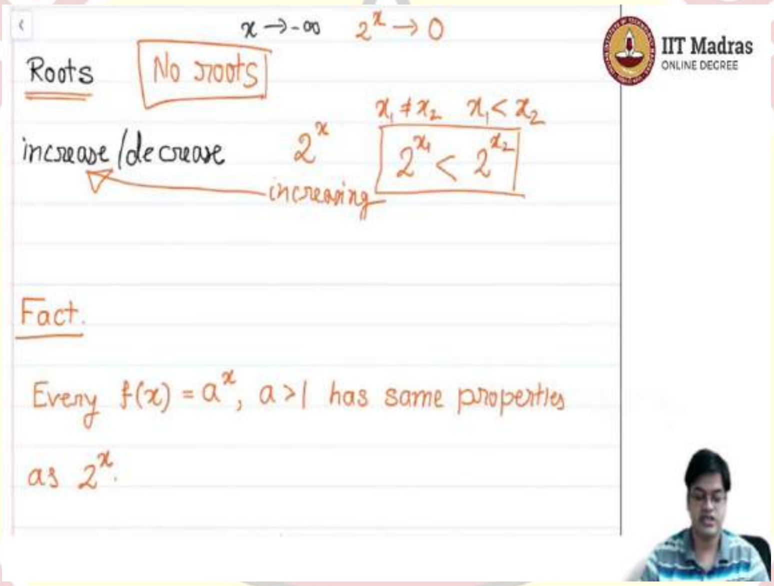
So, the point is (1, 2), the second point ok. So, these 2 points are very special points, they tell you something. So, in particular, had it not been 2^x, but a^x, then that point would have been (1, a) and if you mimic this graph over here y x is over here ok, this is a point 1, this is the point 0 and this is the point which is a.

So, that says $a > 1$; this relation is there, is greater than 0 yeah and therefore, the graph was a point which lies here, which is here right. As $x \to \infty$, this graph actually goes to infinity; as $x \to -\infty$, this graph goes to 0. These two points is these two point and this is an increasing function.

As you come from left to right, it increases. So, this is an increasing function, $y = 0$ is the horizontal asymptote, that is very clear ok. The range of a function is 0 to infinity, that is also very clear. The domain of a function is entire real line, $\mathbb{R}$.

So, we have got all the details necessary for finding this. Now, what it so special about 2^x, if I replace this 2 with 3, still I will have y −intercept to be 0, 1 because 3^0 is also 1 and I will again have domain of f to be equal to $\mathbb{R}$; range of f to be equal to 0 infinity; no x − intercept; $y = 0$ will be horizontal asymptote; $x \to \infty$, $a^x \to \infty$, $x \to -\infty$, $a^x \to 0$. There are no roots. The function is only increasing.

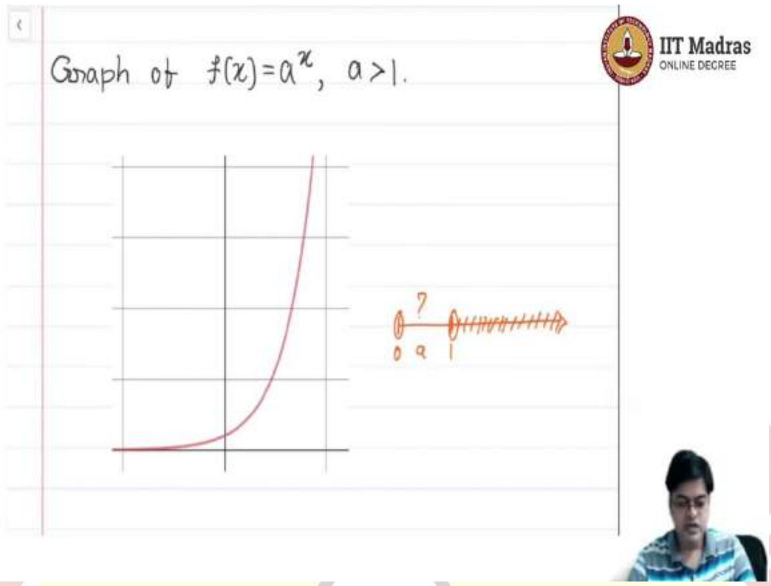
(Refer Slide Time: 10:19)



And therefore, I will state this as a fact that every $f(x) = a^x$, for $a > 1$ will have same properties as 2^x. So, I do not there is no need to draw different different values. The behavior is same only the values will change.

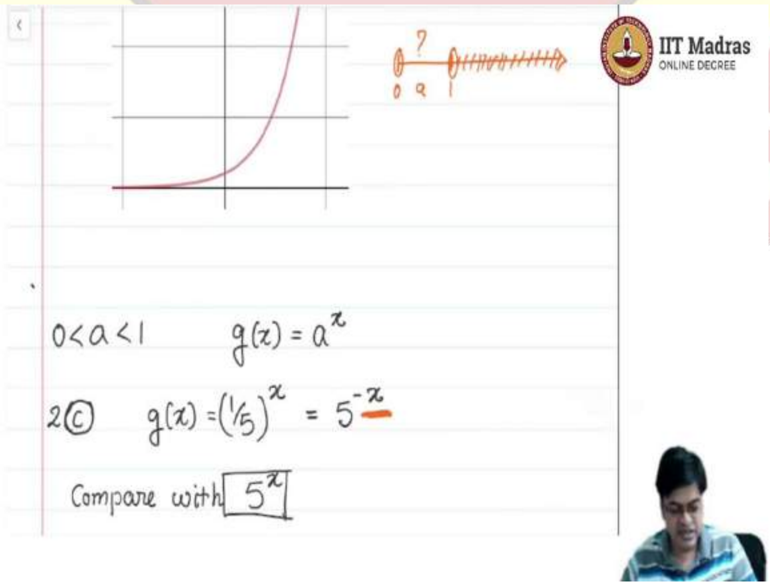
For example, in this case, where you have seen the graph of this (1, 2) is a point; (1, 2) is a point, suppose I consider 3^x, (1, 3) will be the point. So, only the values are changing; but the shape, the behavior, everything else that is listed here remains the same. Therefore, you do not have to draw a graph every time, only thing is you need to evaluate the values in general.

(Refer Slide Time: 11:15)



So, what is the graph of $f(x) = a^x$ in general? It is this way for $a > 1$. So, remember that line that we have drawn which is that the line for a, where we have eliminated these 2 points such as 0, this is 1, we have identified what is the case for $a > 1$. You have also identified the case, where $0 < a < 1$. So, let us go back and see what happens when $0 < a < 1$. So, if a lies here how is the behavior? So, you have already analyzed.

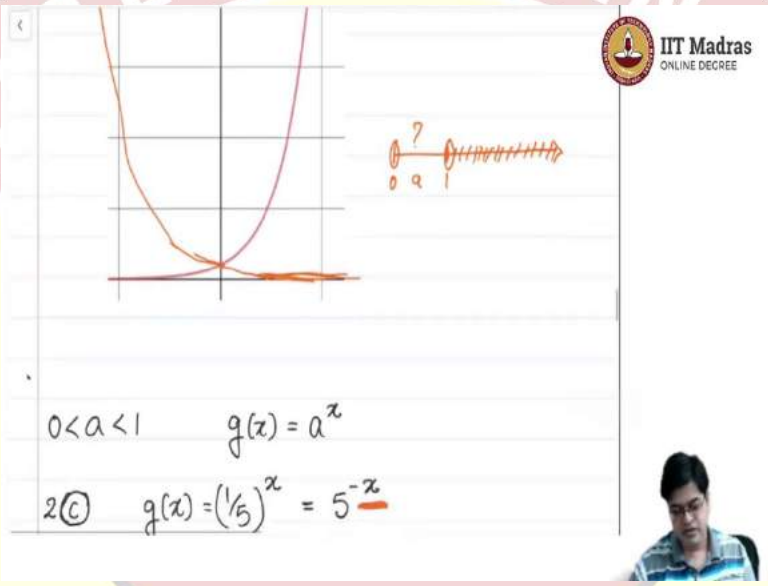
(Refer Slide Time: 11:59)



And let us take this function as $g(x)$ and take it to be $g(x) = a^x$ and this is $\left(\frac{1}{5}\right)^x$. Now, you do not really have to draw this graph, what you can do is ok. So, $g(x) = 5^{-x}$. So, here x is replaced by $-x$. So, what will be the change in the behavior?

So, when x is replaced by $-x$, you know its reflection across $Y -$ axis, you have solved many examples in the assignments. This $Y -$ axis, this is x; then when I put it as $-x$, it will be simply reflected along $Y -$ axis.

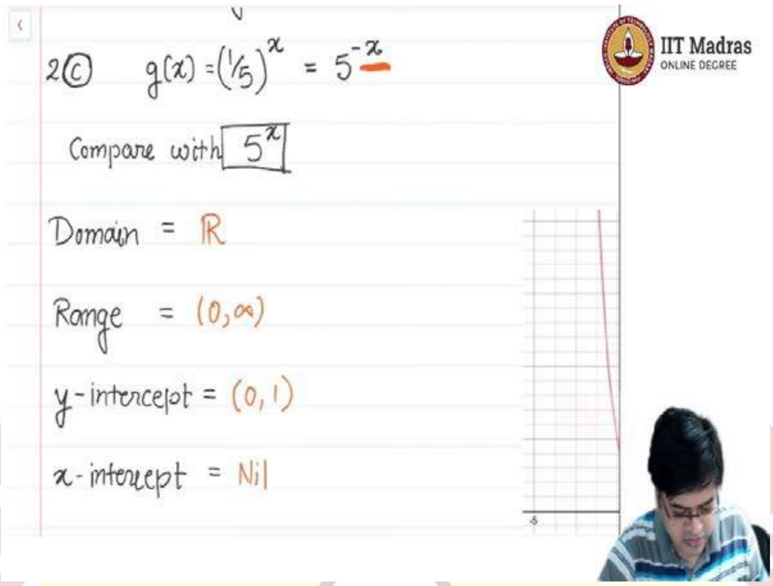
(Refer Slide Time: 12:53)



So, if you look at this graph and try to draw a graph of this function, then it should be something like coming from here going here, it should be something like this, it should actually look like a reflection along $Y -$ axis. So, let us try to show it as reflection ok. This will actually go very close, but never touch.

So, let me erase this ok. So, this is how it will look like. So, without actually thinking about anything else, you can simply draw a graph of $\left(\frac{1}{5}\right)^x$; but still let us try to do it in regular set up.

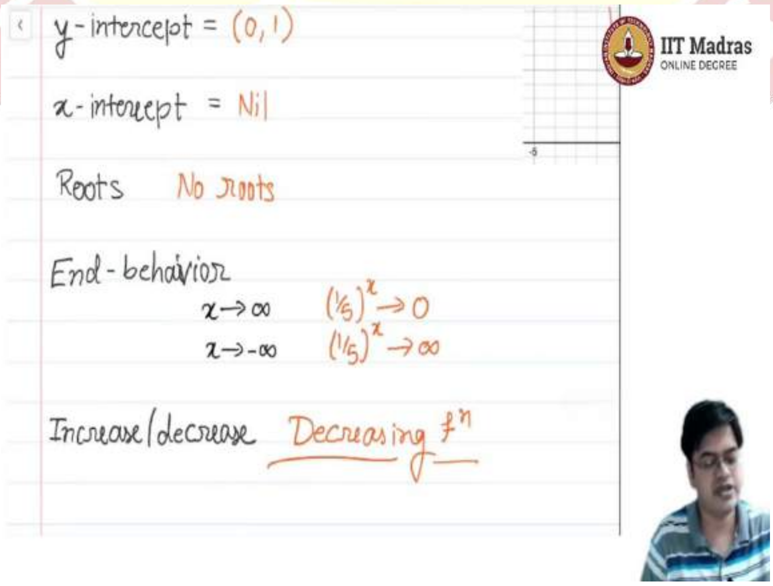
(Refer Slide Time: 13:43)



So, what will be the domain of this function? The domain of this function is very clear because we have used it several times, the domain of this function will be real line. Range, nothing changes; (0, ∞) because it is a reflection across $Y -$ axis. So, let us look at this function.

So, the domain will be ℝ; range will be (0, ∞). What will be the $y -$ intercept? Because it is a reflection, so $y -$ intercept would not change, so it will be 0, 1 only. $x -$ intercept will be nil, there would not be any $x -$ intercept.

(Refer Slide Time: 14:25)

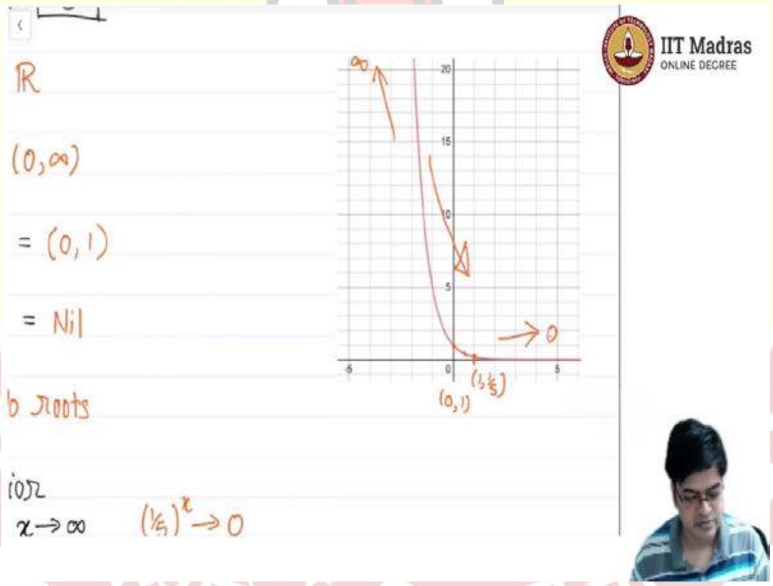


And therefore, no roots and what about the end behavior? End behavior is like $x \to \infty$, $x \to -\infty$. So, when $x \to \infty$, the end behavior will be because it is a reflection you see.

So, when $x \to \infty$ there, it was going to ∞. So, and $x \to -\infty$, function 5^x would have behaved, it will go to 0. So, that reflection will make this a^x or $\left(\frac{1}{5}\right)^x$ whatever is the function $\left(\frac{1}{5}\right)^x$, let me do it properly.

So, this will make $\left(\frac{1}{5}\right)^x$ to go to 0 and this function $\left(\frac{1}{5}\right)^x$ will go to infinity ok. Good. Then, because it is a reflection, the increasing thing will become decreasing. So, there is no intelligence here. So, this will be in fact a decreasing function wonderful. So, we have analyzed everything without taking much efforts. This is the beauty of once you understand the functions on graphical plane.

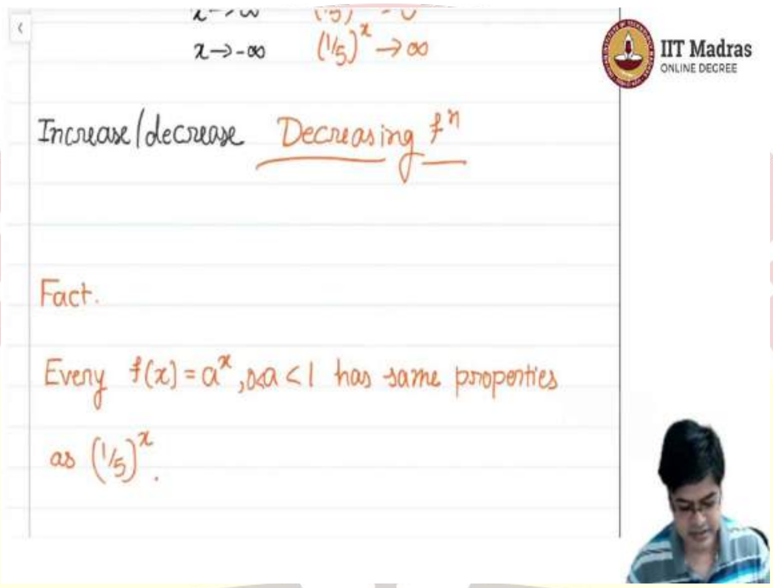
(Refer Slide Time: 16:03)



So, here is the graph of a function which is given to us $\left(\frac{1}{5}\right)^x$, you also might have plotted and naturally, the we will analyze whether it coincides with our thing. So, this is a point (0, 1), now it is $\frac{1}{5}$. So, your point will be somewhere here, sorry this is 5. So, the point 1 is here and this point is $\frac{1}{5}$.

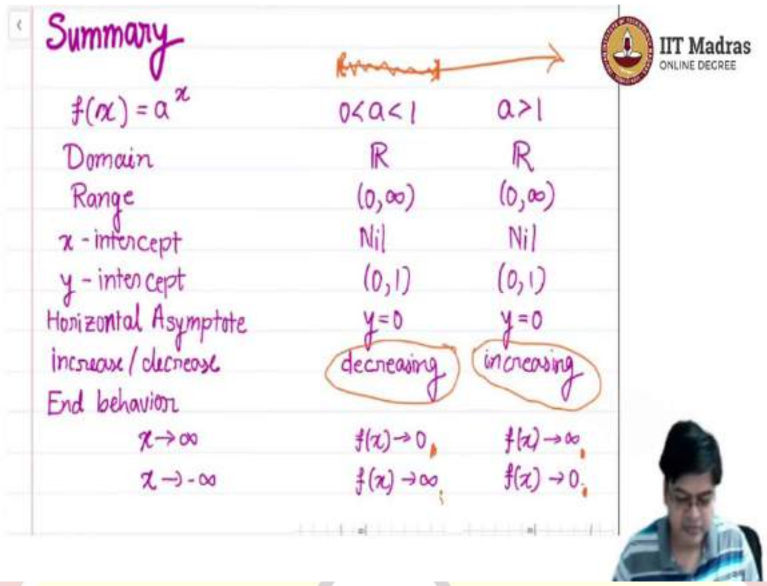
So, $\left(1, \frac{1}{5}\right)$, this is done. Then, as $x \to \infty$, this function goes to 0. As $x \to -\infty$ that is this way, this function actually goes to ∞ and this function is decreasing. From left to right if you come, you are actually coming down. So, it is a decreasing function. So, this completely gives us an understanding of what the graph of a function will look like.

(Refer Slide Time: 17:17)



Also, the same fact is true that every $f(x) = a^x$, where $0 < a < 1$ has same properties as $\left(\frac{1}{5}\right)^x$. Therefore, it is a representative class. So, you do not have to worry about the because it is a representative class, you have to worry about all other functions. All other functions will have a similar behavior.

(Refer Slide Time: 17:47)

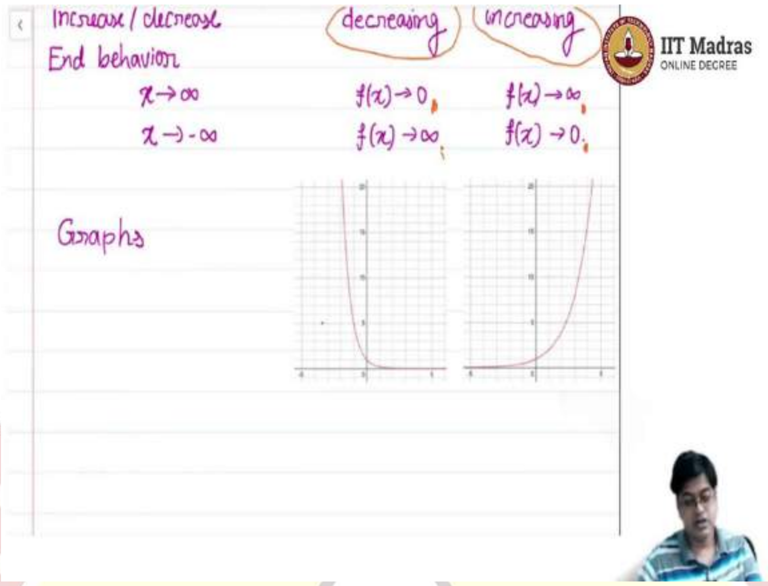


So, we have done a lot, let us summarize these things in a neat table which is this. So, this is the summary of the table. So, if I have been given a function $f(x) = a^x$, then to be more precise, let me draw a line here. This is a line; it does not look like a line, but assume that this is a line.

This is the point 1, then I am talking about $0 < a < 1$ that this zone. In this zone, the domain of a function is ℝ; range of a function is (0, ∞) . There are no $x -$ intercepts, no; $y -$intercept is 0, 1. Horizontal asymptote $y = 0$ is there. The function is decreasing. The end behavior as $x \to \infty$, $f(x) \to 0$; as $x \to -\infty$, $f(x) \to \infty$ correct.

Then, you look at the function which is $a > 1$, domain is real line, range is (0, ∞), nil; (0, 1), $y -$ intercept is (0, 1). Horizontal asymptote is $y = 0$. The only distinguishing feature is the function is increasing here and a function is decreasing here and because it is increasing and decreasing, the end behavior changes that is because it is decreasing, it will decrease to 0 because it is bounded below by 0 and because this is increasing, it will increase to infinity, but here it will go to 0 ok.

(Refer Slide Time: 19:21)



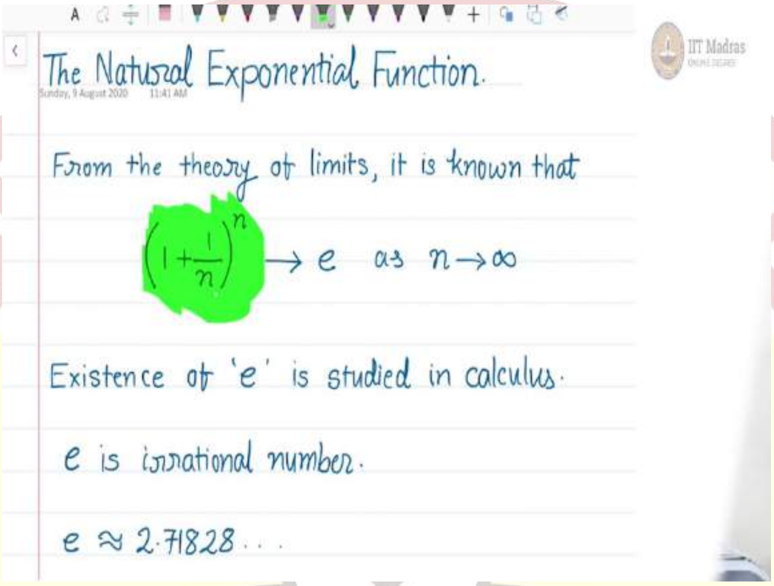
Then finally, you see the prototypes, just look at the graphs of these two functions ok. This ends our topic on exponential functions. Now, we will introduce something which is called natural exponential function in the next video.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 47
Natural Exponential Function

(Refer Slide Time: 0:14)

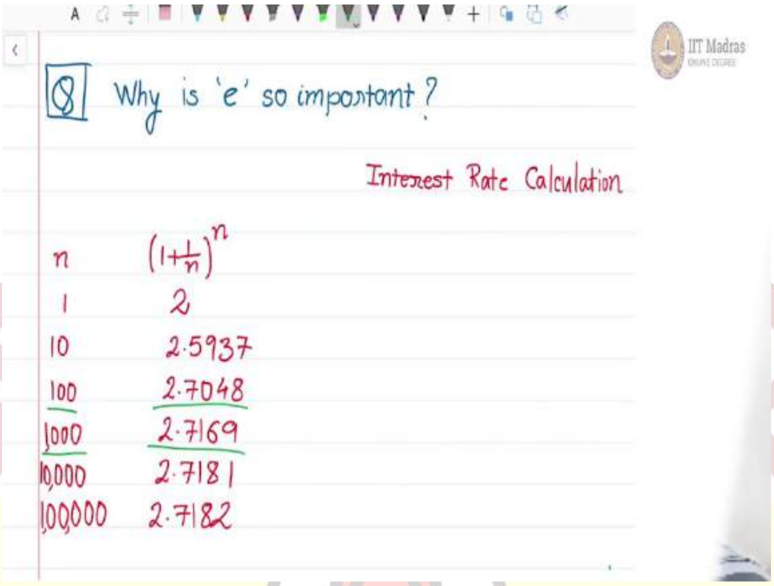


Hello friends, in this video we are going to talk about yet another important function among all exponential functions is natural exponential function. So, basically the theory of natural exponential function is derived from calculus. So, in order to understand how relation natural functional, natural exponential function arises, we need to study the theory of limits. In particular, this natural exponential function is dependent on something raised to the exponent that something is in irrational number that is called e.

And I will make sure by the end of this video you will understand why this number e is very important. In particular, when we talk about number e or ratio or a limit of some quantity is important which is shown here. So, from the theory of limits it is known that whenever you are talking about $(1+\frac{1}{n})^n$ this particular limit it actually converges to e.

So, now unless you understand the concepts of limits, you may not be able to have complete understanding of this concept, but still I will give you some intuition behind this number e. So, though, as I mentioned earlier, the existence of e is actually studied in the field of calculus. For that you may have to do the course which is maths 2, Math for Data Science 2 and you have to agree with me on certain facts without knowing them or you have to trust me that e is an irrational number and e is approximately equal to 2.71828 and so on. It is an irrational number. So, it will go to never ending decimal representation, it will continue on the right.

(Refer Slide Time: 2:26)



So, these are the facts about e. Now the question that we asked at the beginning of the video is why is 'e' so important? So, to answer that question, let us first look at the behaviour of this particular number as a limit. So, when I say that n goes without bounds, the number the this particular function $f(n) = (1 + \frac{1}{n})^n$ converges to e. What do I mean? Let me put it in a proper formal way.

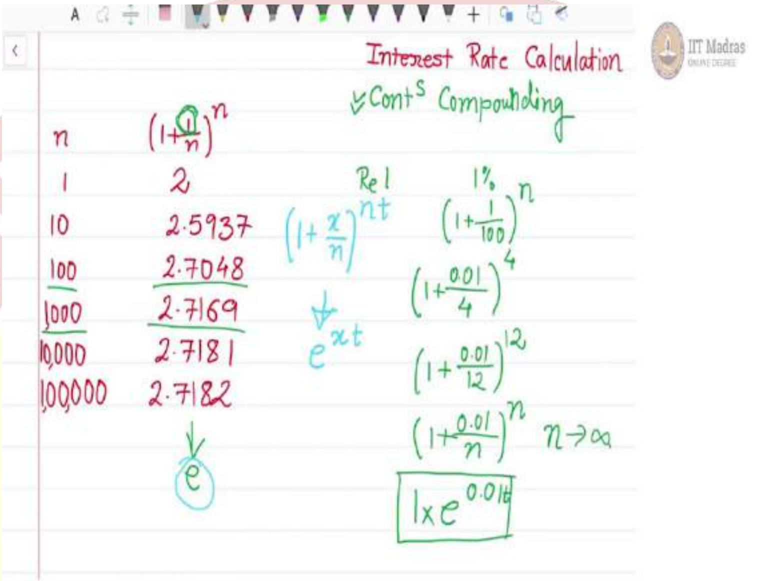
So, I have generated a table over here for our convenience and let us understand. So, when I substitute this n, the value of $n = 1$, this number $(1 + \frac{1}{n})^n$ is simply 2. When I substitute $n = 10$, the number becomes 2.5937. So, does that mean this function will go without bounds? The answer is no that is why we get the convergence. And such type of questions are studied in Calculus.

So, when you substitute the further values of n that is, $n=100$ you have substituted, you got 2.7048. When you substitute $n = 1000$, you will get 2.7169. Now, you can see that you approaching closer to the ideal value of e. And when you put $n = 10000$, you get 7181 still because we are writing up to 4 decimal places, we are not really very close to it, but we will be, we are very close to it, but we are not at that point.

But when I put n is equal to 1 lakh, then I get a value of e which is 2.7182 and that is actually exact representation of this number e up to 4 decimal places, correct up to 4 decimal places.

Now, if you go on further and put higher and higher values e like you can put it to be a 1 million, 10 lakhs and then you will see you will get further improvement, but because we are focusing only on 4 digits after decimal, we this requirement is enough for us, this much calculation is enough for us. Another thing, another aspect in which this e becomes very crucial in accounts this interested calculations.

(Refer Slide Time: 4:57)



You must have heard the term by now of continuous compounding and continuous compounding actually means the taking ratios with respect to e or taking exponents with respect to e. So, let me demonstrate to you in this manner. Let us say you have invested rupee 1 in a bank and bank is offering 1 % interest rate and you have invested it for 1 year. So, in that case what is the answer? 1 plus 1 upon 100 raised to 1, this is the answer, 1.01 1 % you will get if you have invested rupee 1, you will get 1 paisa of interest.

So, now if you go on like this you will have something like $(1+\frac{1}{100})^1$ raised to so whatever number of years you have invested in raised to n. Now, when you actually look at the procedure of the bank, banks do not give you the interest which are given annually, but they credit the interest quarterly. So, in that case what you need to understand is the interest rate is actually given in a quarter, so it is computed on quarter and whatever interest you have accumulated, that interest will be taken into account for the next quarter.

So in that case, basically what bank is doing is bank is taking this interest rates which is 0.01 and it is actually dividing it into 4 quarters, 4 parts because they are giving you a quarterly interest and then you are actually getting this multiplied in this fashion. So here, if you look at the interest rate, instead of 1, I have 0.01 as the number and for 0.01 I got this number. So, if the bank decides, so I will revise the bank decides that I am revising, bank is revising the interest rate every month, then what will happen?

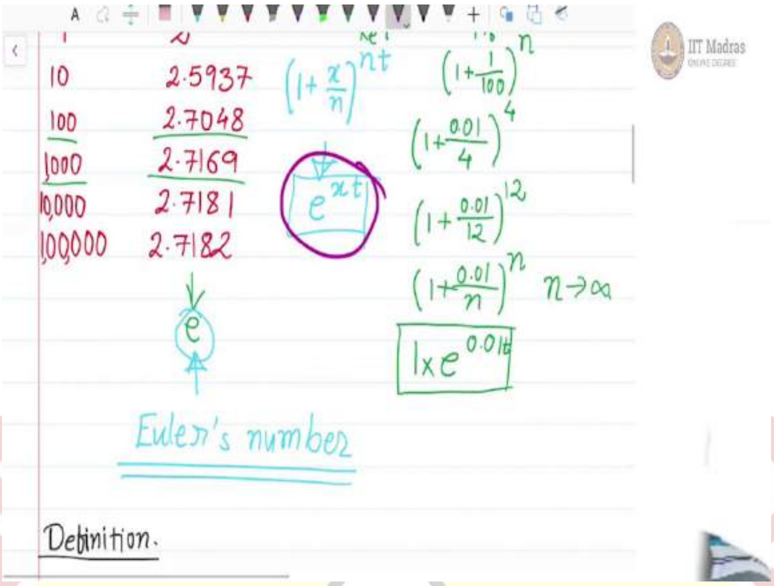
Then the same logic for single year, for single year remember $(1+\frac{0.01}{12})^{12}$. So, if the bank starts revising the interest rate infinitely often, then we are actually talking about something like $(1+\frac{0.01}{n})^{n}$ and this $n \to \infty$. In this case, according to our judgement, according to this, this number was 1 here, now it is 0.01 and this number converge to e. So, based on this understanding, if you apply the same logic and try to calculate this thing, then it will converge to e raised to 0.1.

This is an interesting revelation. That means, if you invest rupee 1, you just take that rupee $1 * e^{0.01}$ that will be the interest accumulated along with the original capital in are bank if the bank follows continuous compounding. This is how whenever you study finance, you calculate the interest rate. So, this is for the period of 1 year. Now if you add the period in terms of time, then it will be $e^{0.01t}$. So, this is how e becomes important. Let us now replace this 1 % by a generic number which is x.

So, what I am talking about now is $(1+\frac{x}{n})^{n}$ and now from the discussion that we have done this will converge to e^{x} and when I add the time that is it is more than 1 year, then I have something like n_t and that is where I will get x_t. So, these are simple understanding why the number e is very important. e typically comes when you are considering a continuous compounding.

I hope I have made the relevance of the number e, irrational number e very clear and it is an irrational number and its exact value is given by this particular expression. It is not exact but it is approximate which is suitable for our purposes.

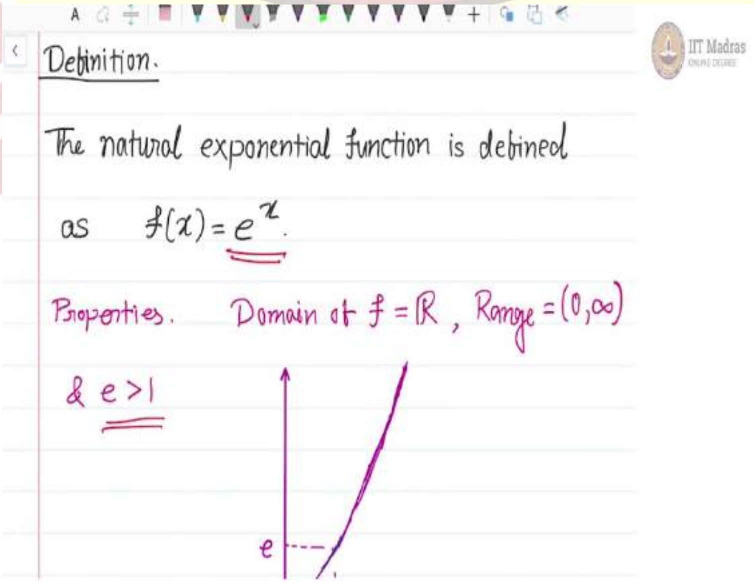
(Refer Slide Time: 9:53)



Now let us go further and understand what is a function that we have defined here and it is why it is called natural exponential function? So, this number *e* as I mentioned now naturally comes when you are considering continuous compounding. It also comes very naturally in the field of Differential Equations which is also relaying on our calculus. So, this number *e* has a special name when you consider differential equations as a area which is called Euler's Number.

So, you can Google and you can search the meaning of Euler's number and why it is relevant. So, that is how this *e* is called a natural exponential. So, now let us formally define the function that we have just now seen which is *e* raised to x_t as a natural exponential function.

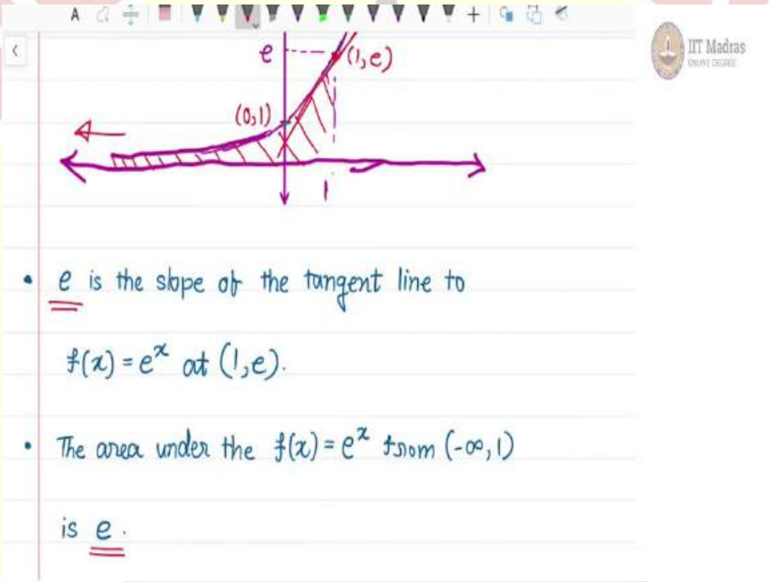
(Refer Slide Time: 10:50)



So, a natural exponential function is defined as $f(x) = e^x$. Then, you may ask a question, what are some interesting properties of this natural exponential function? Now, the properties will be very similar to the exponential function that we have studied, but it is special in some sense. We will see its specialness in a when we will study some special properties of this natural exponential function e^x. Let us list all the properties.

Domain of f, domain of this function will be set of real numbers and range of the function is positive real line that is 0 to ∞. As you have seen e, the value of e is 2.7182 so e is natural greater than 1.

(Refer Slide Time: 11:51)

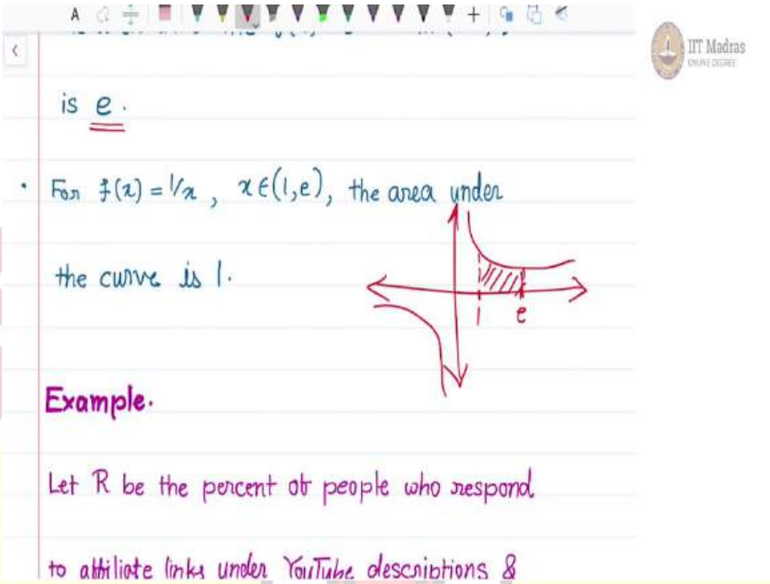


So, if you recollect whatever we studied for exponential functions, you will get the graph of exponential function in this manner where there are two typical points 0,1 is one typical point that it passes through and it will always pass through 1,e. As x tends to infinity, the function goes without bounds, as $x \to -\infty$, the function asymptotically goes to the x axis so $y = 0$ is the horizontal asymptote for this function, we have already seen that. For general exponential function same properties hold true.

Now what makes e special? And what is something special that is not true with general exponential function. So, in this case, if you look at the point (1,e) and if you draw a tangent to a line tangent to the curve, that is a line passing through this particular point, the slope of this line will be e, that is very special. So, e is the slope of the line that is tangent to the curve y is equal to e raised to x at (1,e). So, that is one thing.

Then if you look at the area that is covered under this curve from $-\infty$ to 1, that area is actually *e,* the irrational number *e.* This you will learn when you will study calculus in maths 2. So, that is very important.

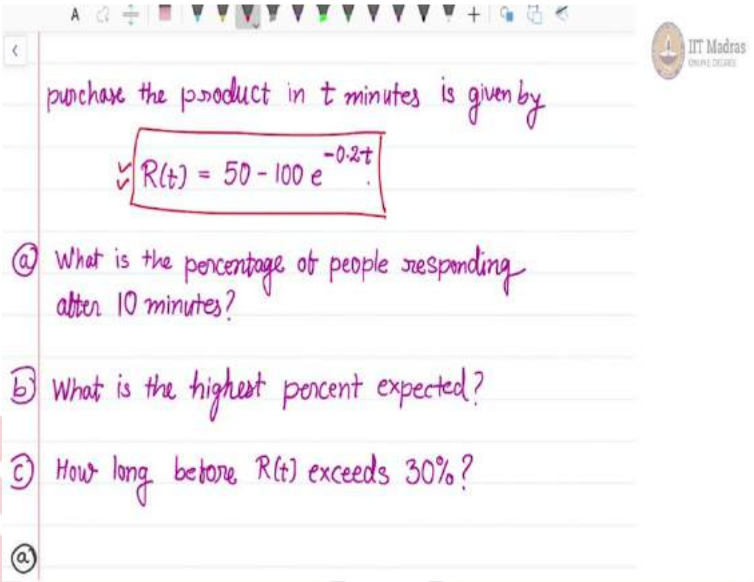
(Refer Slide Time: 13:38)



And the third thing that is very important which will not happen in general with other exponential function is if you draw a curve *f(x)* is *e*, if you draw a function *f(x)* = $\frac{1}{x}$. So, you may be familiar with that function, it will be something like this and something like this. And in this particular case, if you look at the area under the curve in the in the range 1 to *e*, this particular area, this area is a unit area for *f(x)* =*x* and remember this *e* is an irrational number so still it will be a unit area.

Why it is so? this is a matter of calculus to explore, but these are the things, these are some of the things that makes the function *f(x)* =e^x special function. Let us now understand this function better by considering an example which will deal with our real life problems.

(Refer Slide Time: 14:43)



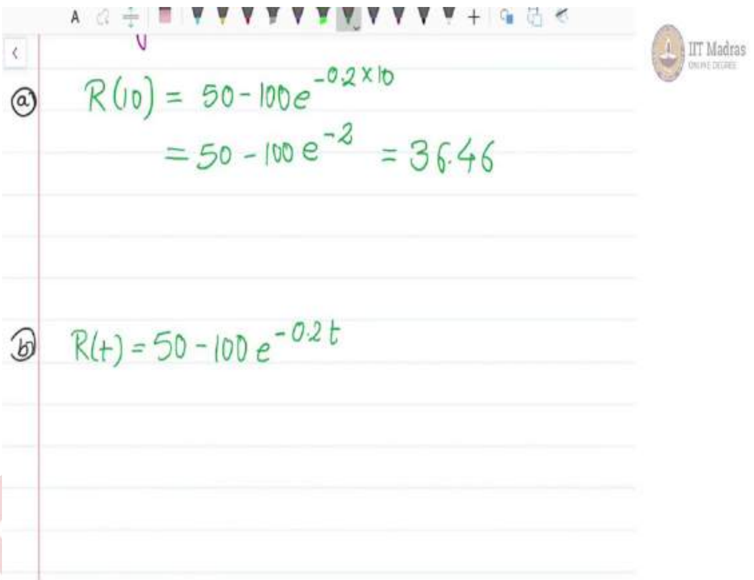
So, here is an example which says that let R be the percentage of people who respond to affiliate links under YouTube descriptions and the purchase and they purchase the product in t minutes and that particular purchasing thing is a function of time so it is given as the $R_t = 50 - 100e^{-0.2t}$. So, let me give you a brief understanding of the problem.

So, now when you watch some video on YouTube if you, the speaker in the YouTube says that there are some affiliate links below in the description. Now, if you click on that link and go to the affiliate site, then what you will do is, either you will purchase or you will not purchase. If you will purchase, the speaker or the channel owner will get some amount of commission.

Now, here the person who is actually giving the affiliate links is interested in finding the number of people who are responding in t minutes. So, he has devised a function which is available in YouTube statistics so based on the data available, he has derived a function, we are taking the function as it is. So, that function is $R_t = 50 - 100e^{-0.2t}$.

Now, he is interested in answering these questions. What percentage of people responding after 10 minutes? So, how many percentage of people responded after 10 minutes? Then, based on this function, what is the highest percentage expected? And the third question is how long before $R_t > 30$ %? The response rate being 30 % is also a good enough rate.

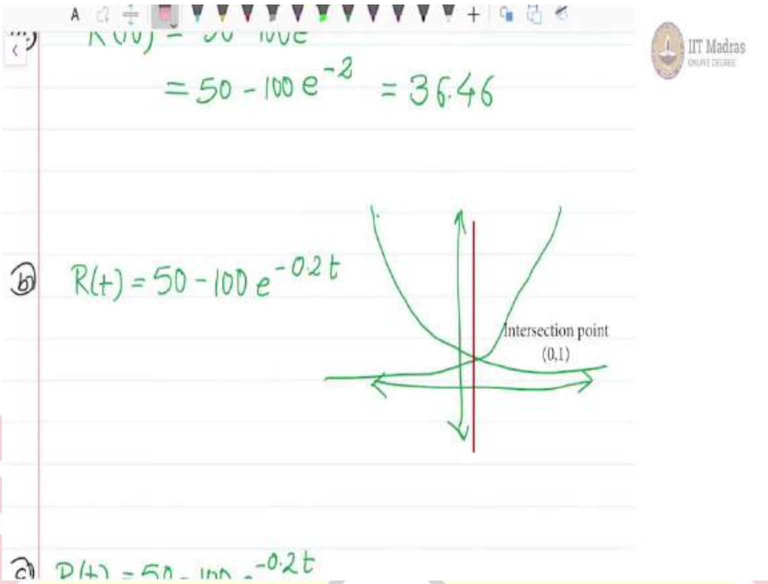
(Refer Slide Time: 16:54)



So, because you are just putting some affiliate links. So, let us try to understand what percentage or people will respond after 10 minutes? That means, I want to essentially evaluate the function as R_{10}. So, if I substitute this, it will be $50 - 100e^{-0.2\times10}$. that is simply if you rewrite this as 50 -100 times, this is 2/ 10 which can be simplified to e^{-2}. And then you can actually calculate the function e to by value of e^{-2} and you can put the value, that value is 36.46. So, this you can do it using calculator.

Now, let us look at the second question. What is the highest percentage expected? Now you have to think about this function which is $50 - 100e^{-0.2t}$. Now, you look at the function which is $e^{-0.2t}$ or e^{x}.

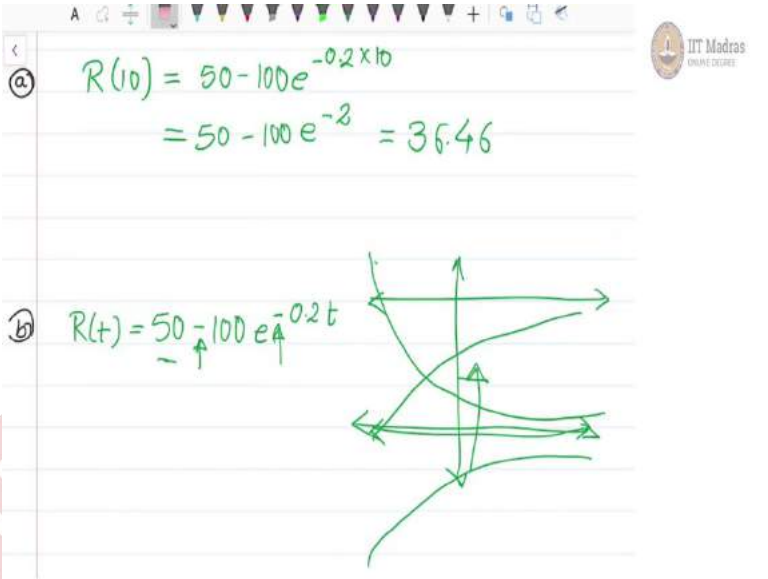
(Refer Slide Time: 18:19)



You already know the graph of the function which is e raised to x, $f(x) = e^x$. Now, how will the function look like when you are talking about $f(x) = e^{-x}$ has a graph of this form, roughly this form. Now, when you are talking about − of x, when you are talking about − of x, you are actually talking a reflection of this graph along this so that will give you some graph of this kind, it will never cross x axis but it will go this way.

So, now you have a understanding of how the graph of e^{-t} will look like. But here are some scaling versions, scaled versions like 100, this is 100 and this is 50. So, now this graph is actually multiplied with − 100, this graph is actually multiplied with -100, but multiplying with − 100 will again, what it will? It will actually keep the graph in a similar manner but it will actually because it is multiplied with − 100, it will shift in some sense like this.

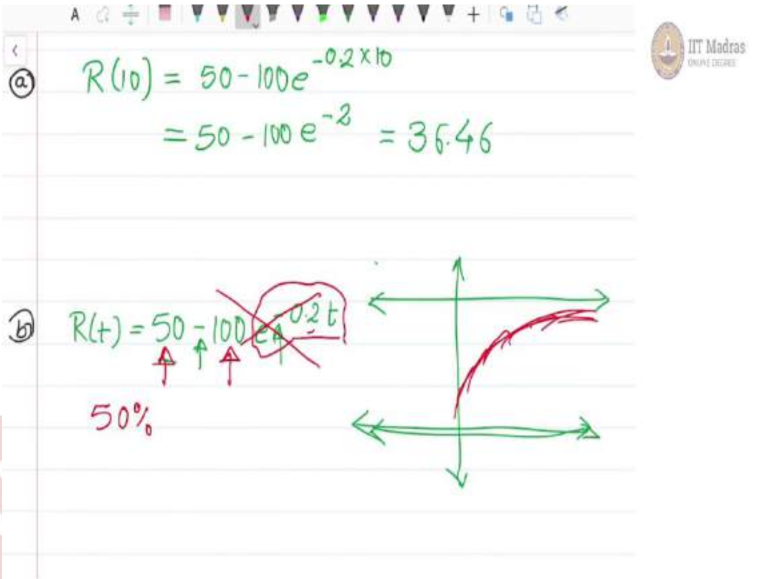
(Refer Slide Time: 19:33)

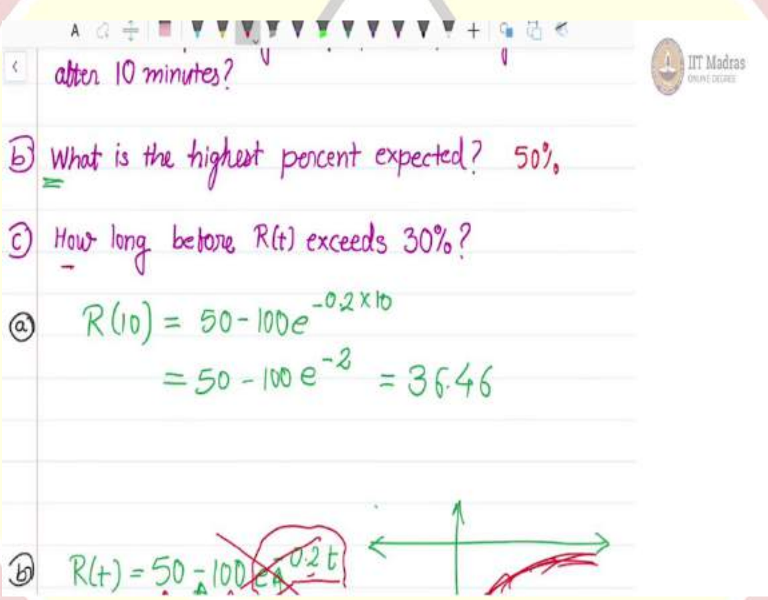


One minute let me chose and erase it. So, it will shift like, it will flip here and it will shift like this. And then, now when you are adding 50 to it, this graph will actually go up by 50 units. So, this way the changes will happen to the graph and finally graph will look something like this. You can actually check for yourself. So, basically first multiplying this − sign will have an effect of reflecting the graph along y axis, then multiplied with − 100 will reflect the graph along x axis and then adding 50 will shift the graph by 50 unit.

So, you have a fairly good understanding of the graph. Now, you just apply your knowledge that what is the highest percentage expected? So, in this case, if you understand this, the horizontal asymptote over here is actually shifted to 50 units because you are transferring to 50 and the graph actually let me clear up the image.

(Refer Slide Time: 20:49)

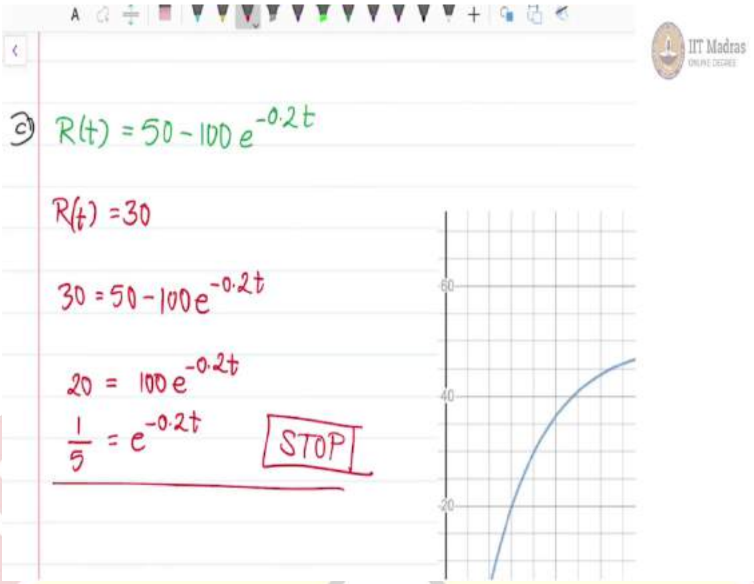




You can use your graphing tool also and verify that this is the graph. So the graph will look somewhat like this and it will asymptotically reach to 50 units. So, the highest percentage that is expected will be 50 %. It will not exceed 50 % based on the graphical analysis. Let us analyse this graph, instead of graphically analysing, let us look at this function e raised to − 0.2t, this function in itself will never exceed 1 and as x tends to infinity, this function will actually tend to 0.

That means, whether I am multiplying by 100 or I am multiplying by 10000, as $t \to \infty$, this function has to go to 0. So, this entire thing will go to 0 and therefore, 50 is the maximum that I can achieve. Therefore, my question b is answered as 50 %. Now let us look at this particular thing, how long it takes before R_t exceeds 30 %?

(Refer Slide Time: 22:06)

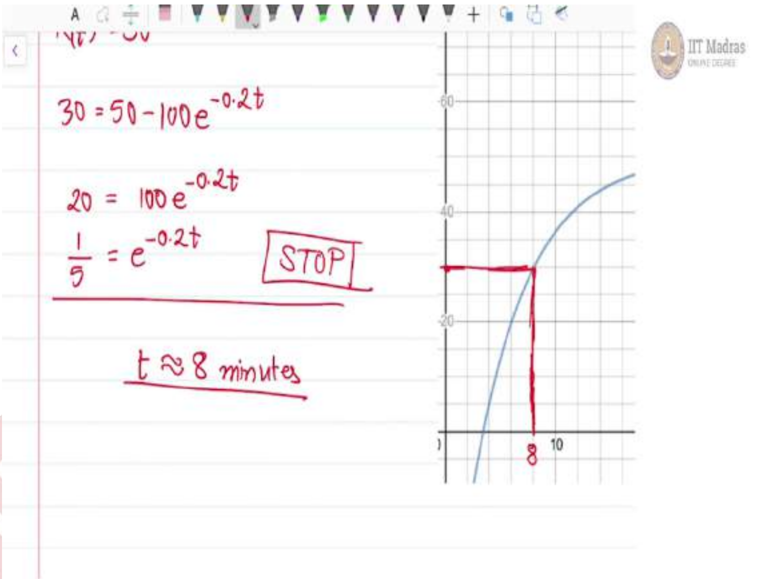


So, in this case you just look at the graph of the function, so as I have already described this will, this is how the graph of the function will look like. Now, there are two ways in which you can solve. Let us try to see whether we can go ahead formally and solve this. So, essentially I have been given that how long till R_t >30 %? That means R_t= 30, find are value of t such that $R_t = 30$?

So, 30 = 50 − 100$e^{-0.2t}$. So, 30 − 50 will give me something like − 20 = − 100$e^{-0.2t}$. Now, this is, these − signs will cancel themselves off so, this is this and then you simply rewrite this expression $\frac{20}{100}$is nothing but $\frac{1}{5} = e^{-0.2t}$. Now, I have to stop here because right now I do not have any ways to see what t will be when $\frac{1}{5} = e^{-0.2t}$. No analytical way is possible.

Then, what I will do is, so analytically I am stopping here and if I somehow I am able to figure out how to find t is equal to something, then I can answer this question. But let us now try to compute this graphically.

(Refer Slide Time: 23:47)



So, in this case, $R_t = 30$ is this point. So, now if you go along this line and then you map this onto x axis, remember x axis is nothing but the value of t. So, in this case, now you look at the mesh, this roughly turns out to be 8 that means, t will be approximately equal to 8 minutes. So, this is how we can without even solving the expression for this, graphically solve the expression for exponential functions. So, this is one live demo of that. That is all for now. So, we will meet in the next video where we will actually try to understand how to solve this particular problem. Thank you.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

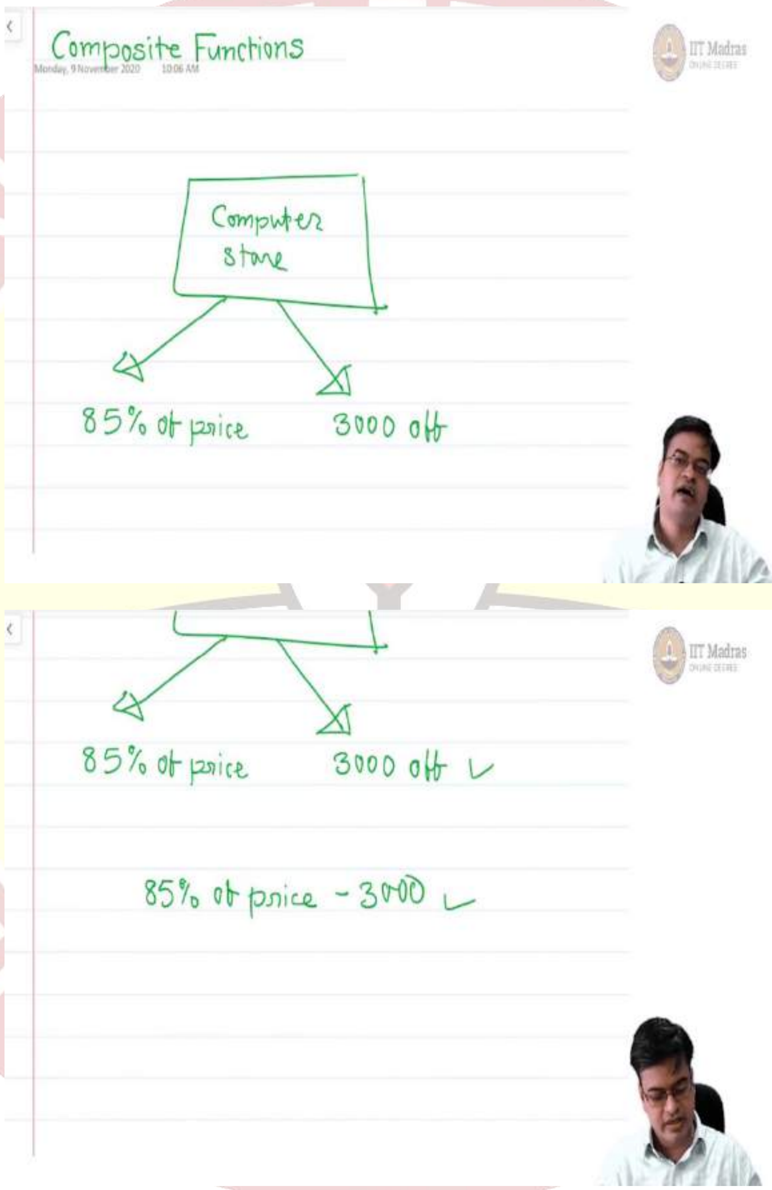
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 48
Composite Functions

(Refer Slide Time: 0:16)



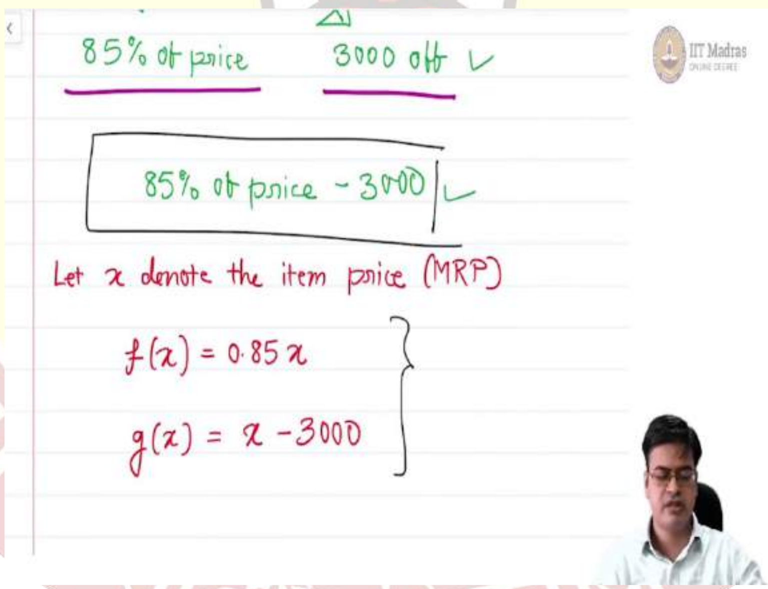
Hello students, today we are going to learn the concept of composite functions, what do you mean by a composite function? So, let me motivate this with an example. For example, it is known that you are a very good bargainer and your friend wants to buy a computer. So, your friend takes you to a computer store, so this is a computer store, in this computer store there are two offers available.

So, something is on sale, all items are on sale and there are two offers available, one offer is you will get 85% of the price, whatever you buy you will get the product at 85% of the price. And the other offer is you will get flat 3000 off on the MRP, the maximum retail price you will get 3000 off. So, these are the two offers that are available.

Obviously because you are a good bargainer, you bargain with a salesperson and you strike a deal that is the computer that you want to buy will be given to you at 85% of the price and of the amount, once the 85% of the price is decided further 3000 will be given you as a discount. So, there is a discount of rupees 3000 as well as you are getting 85% of the price.

Now, this kind of thing when we write mathematically can be considered as composite functions, you are in fact using these kind of tricks in a day-to-day life. So, let us see what happens when we put this mathematically and how composite functions arise. So, let us say the first draft that is 85% of the price. So, can I represent this as a function?

(Refer Slide Time: 2:20)

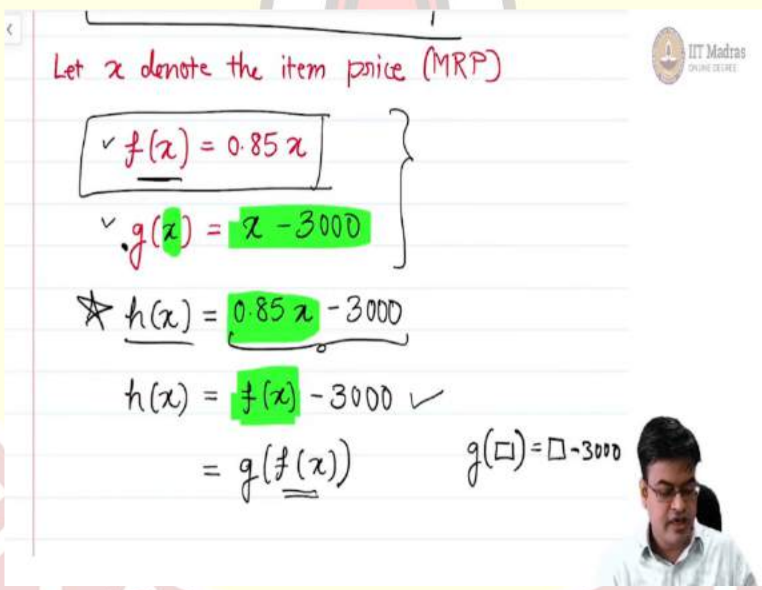


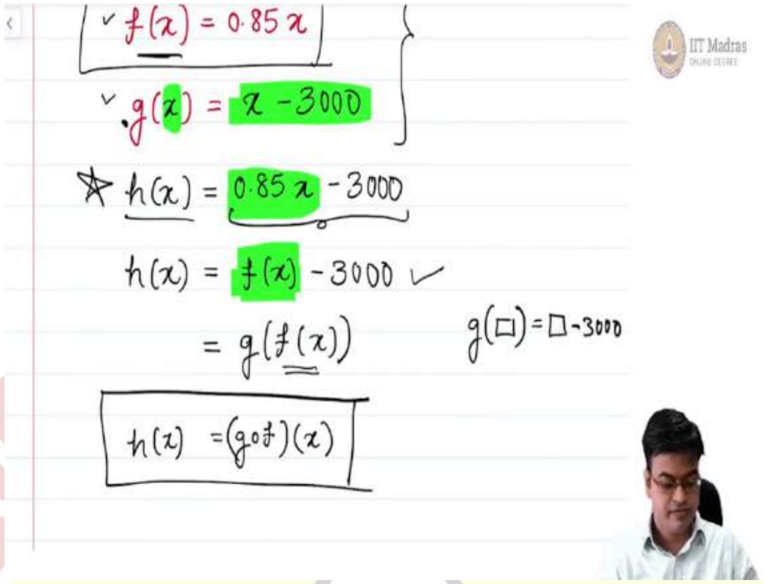
So, for cleanliness let us write let $x$denote the item price. So, let $x$denote the item price which is the MRP, you can write maximum retail price and on that you are getting 15% discount that is 85% of the price you are getting. So, I can write this particular offer as *f(x)* which is nothing but 0.85 times *x*.

Now, the other offer that is on in this particular computer store is this. So, I can write this as g of $x$to be equal to if $x$is the MRP I will subtract 3000 rupees from x, so these are the two offers that are available. Now, what we did is we want best of both the offers. Now, when a store is offering these two offers it is safe to assume that you may not have any item that is less than 3000 rupees, you may not have any item on sale which is less than 3000 rupees, so your $x$will always be greater than 3000.

Another thing that you can assume that because the store is offering you this kind of thing, that is 85% of the price -3000, the store has already taken care of that they do not have to pay back any money, that means after giving the 85% of the price, the price should be greater than 3000, so all these conditions are assumed implicitly, which we will deal with them in later when we will formulate a problem.

(Refer Slide Time: 4:23)





So, now the offer that you got if I want to write this offer mathematically I can write this as some function $h(x)$ which is equal to it is 85% of the price -3000. So, now when we are dealing with functions in mathematics it is good to see if I have some correspondence of the function h with these functions f and g, this is the question that we are trying to answer when we are studying composite functions.

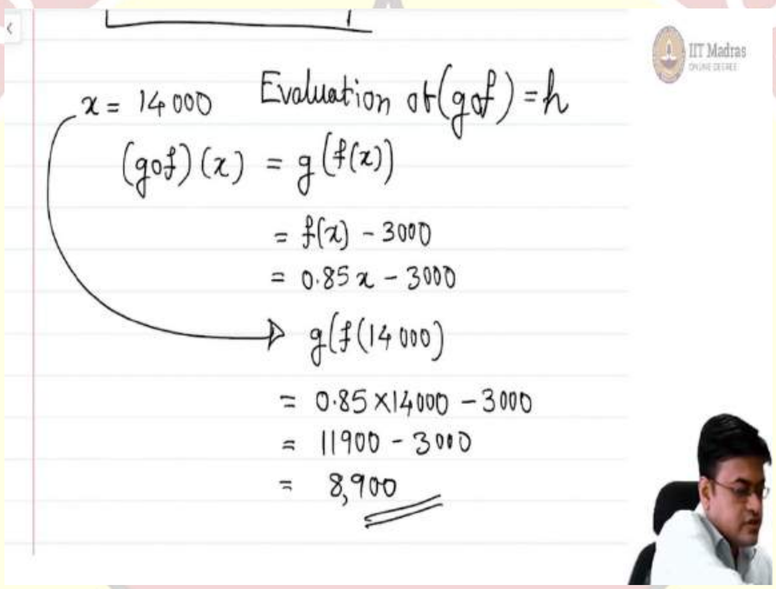
So, let us first see what is being done over here, that is if I use this f then it is 0.85 x, so if I want to do something like this then I can write this as $h(x)$ is equal to $f(x) - 3000$ is that a safe assumption to do? Yes, of course because $f(x)$ is 0.85 x so what I am essentially doing is, I am for this particular term I am substituting $f(x)$, so it is a perfectly valid guess, fine.

Now, if you treat this f, if you treat this $f(x)$ as one argument like $x$then what you are actually doing, you are actually saying it is $x$minus 3000 that means instead of this x had it been $f(x)$ you would have written $f(x) - 3000$. So, I will use that knowledge and I will try to do, I will try to rewrite this as, this is g times f(x). Is this acceptable? Let us redo the math.

For example, what is g times $f(x)$? So, if you look at g of, f of, $g(x)$, so whatever is x you will write that $x-$ 3000 or whatever, let me put it this way if g had some box inside it then I will write that box -3000. So, in particular, in that box right now $f(x)$ is written, so I will substitute it as $f(x)$ minus 3000, done?

And what is *f(x)?* Now, *f(x)* as you know is nothing but $0.85\times x$. Therefore, I can rewrite this function as *g(f(x))*. In mathematics you will rewrite this as *g(f(x))*, so my *h(x)* can also be written in terms of g and f in this fashion. So, this is the motivation for composition of two functions. So, in particular what we have seen is a practical example, we motivated it through a practical example of a computer store which is offering two kinds of sales, one is 85% of the price, another one is flat 3000 off on the MRP. So, after doing this you can easily guess that how will, how will I evaluate this function, how will I evaluate this function, that is what we have to see.

(Refer Slide Time: 8:07)

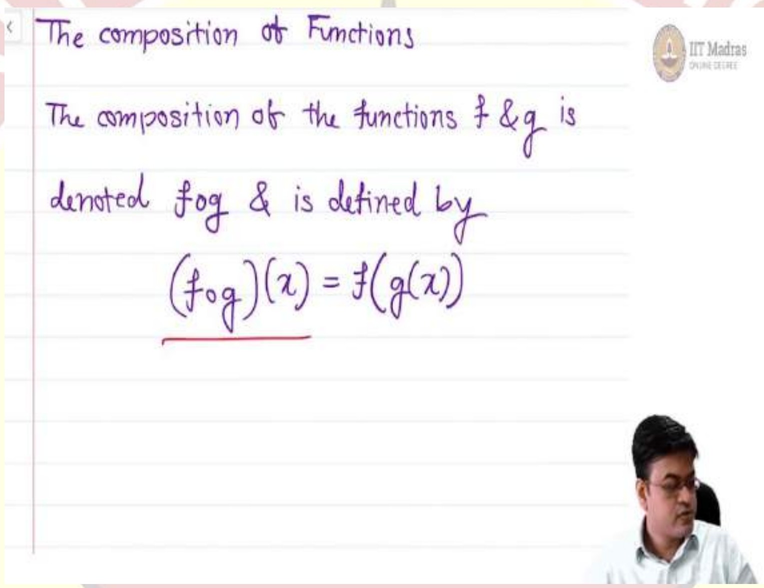


So, in particular let us say your $x$in this particular function is say you can take it to be 14000 let us say, 14000 is your $x$and you are asked to calculate *g o f(x).* So, how will you calculate? It is very simple, you will first insert g(f(x)). So, what is f(x)? f of x is nothing but point okay, let us follow the same notion the way we followed, so in particular in this case this is what will happen, this is going to be equal to $f(x) - 3000$.

What is $f(x)$? *f(x)* is going to be $0.85 \times (x - 3000)$, so I will substitute the value 14000 over here which will give me, so since my x is 14000 I will plug this value in, so I am calculating g of f of 14000. What will be g of f of 14000? Again you have to do a similar calculation which will give me 0.85 multiplied with 14000 -3000 so this I think comes out to be 11900 just check if I am calculating it correctly -3000 which will give me 8900, 3000, 900 as it is 11 -3, 8, yes, so the final answer is 8900, this is what, this is actually, what I have just now shown is evaluation of a composite function which is *g o f,* what is, which is actually *h*, there is nothing special in this, it is just a nomenclature that we are using.

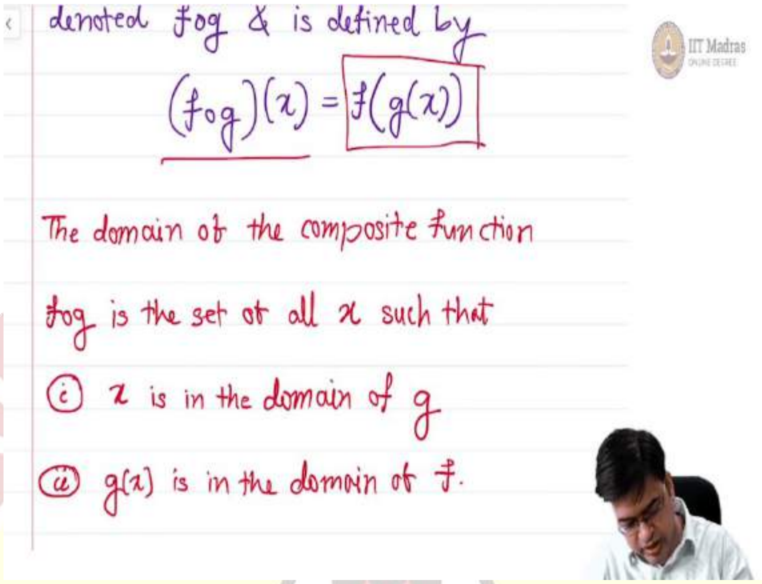
But this kind of composition helps you in understanding lot of things. So, let me formally define what is the composition of a function and how we are going to handle them mathematically. Because composition of a function as you must have seen is again a function. So, natural questions about domain, range will arise and we will try to answer them as and when they come.

(Refer Slide Time: 11:01)



So, let me formally define the composition of functions. What is the composition of function? So, in particular we can write as the composition of functions f and g composition of the functions, there are two, at least two functions you need, functions f and g or we can write the composition of the function *f* with g that is also a valid terminology is denoted by, I have already defined this notation *f o g* and is defined by *f o g,* this is one function of *x*, so you can write this as *f(g(x))*.

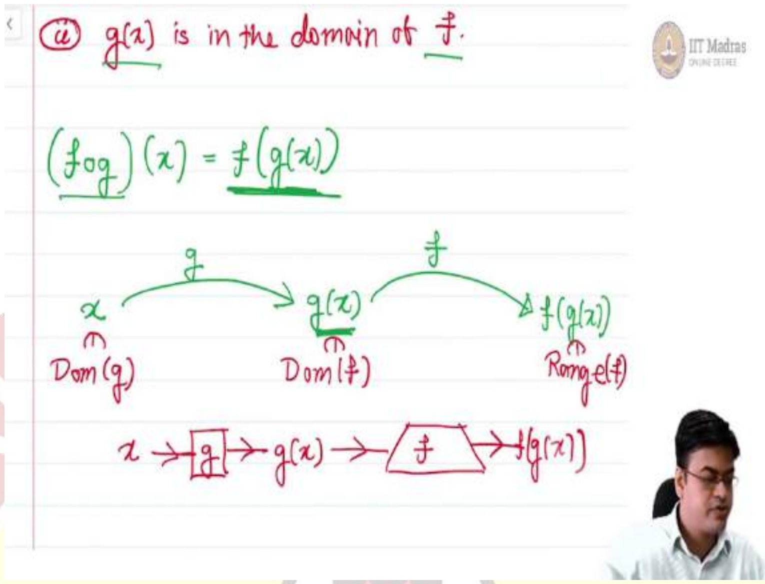
(Refer Slide Time: 12:34)



So, naturally the next question is what should be the domain of this function, so that we will answer as the domain of the composite function $f \circ g$, let me write it here, $f \circ g$ is actually the set of, is the set of all $x$such that the two conditions we require and they are pretty evident, as we go further we will realize how these two conditions are evident.

So, the first condition is x is in the domain of g and second condition is it will be about $x$so if g $x$is something that you are figuring out. Now, that $g(x)$ should be in the domain of f, $g(x)$ is in the domain of f. So, now why these two conditions are required that is what we need to figure out. For that you need to focus on this particular component $f(g(x))$. Let us use this particular component and try to answer the question.

(Refer Slide Time: 14:21)



So, I have, if when I talk about *f o g(x),* what I am talking about is *f(g(x)).* Now, let us look at the first condition. If I want something to be in the domain of *f o g* that means it should be well defined, so when I input the value it should give me the output, if there is some ambiguity then it is not a properly defined function. So, let us say why this condition x is in the domain of g.

What if x is not in the domain of g? g of x is not defined, because g is defined only over domain of g, so g of $x$is not defined and therefore you need this condition that $x$should be in the domain of g. Now, when I am using this composite function, I am applying f to the value that is obtained by applying g, so it is g of $x$that is playing the part.

So, now if this g of x that is the value of $x$which is in the domain of g if that particular value g of $x$is not in the domain of f then again this f(g(x)) is not defined. Therefore, I need g of $x$also to be in the domain of f. So, in particular you can visualize it this way. So, if I have $x$then there is a map which maps everything that map is g and that maps it to a value called g of x.

Now, this *g(x)* should be in the domain of f because I will take this value to a function which is f(g(x)). So, this is another value and what is the application? *f* is the application, we are applying the function f to the value *g(x),* if this $g(\ x)$ is not, does not belong to domain of f then our function is not defined.

So, you can actually remember this diagram by using this particular, this belongs to, what it belongs to? It belongs to domain of *g*, this particular thing actually belongs to domain of *f*, this is my abbreviation for domain and this is nothing but the range of f, so this will be in the range of f but it can be smaller than the range of *f* because $g(x)$ may not cover the entire domain of f.

So, it can be smaller but this will belong to range of *f* or if you want to visualize it in a better manner there is something which is box, you feed an input to this box *x*, *g* is this box and it will throw out $g(x)$, so when you feed *x*, this will spit out $g(x)$. Now, for $g(x)$ to be fed into another let us say this is a trapezium and that trapezium is a function f, this is a machine, function machine that is $g(x)$.

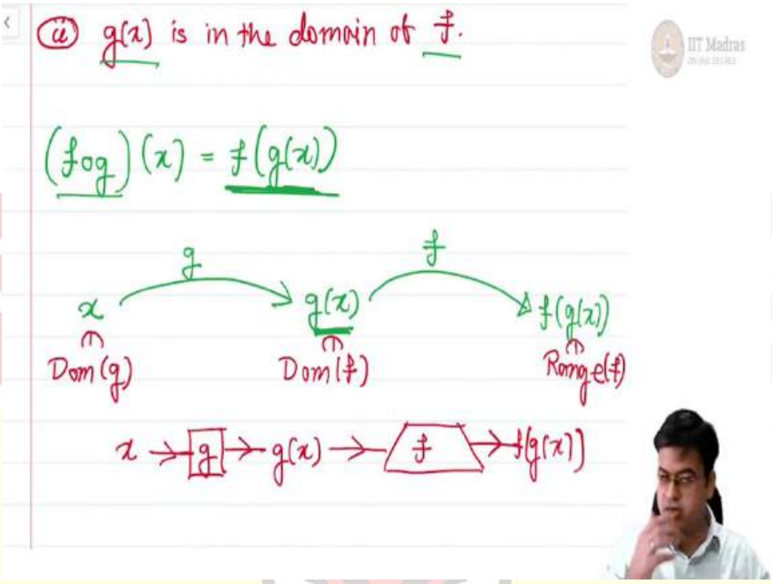
If this $g(x)$ is not in the domain of *f*, then this machine is unable to produce the output, so I need this $g(x)$ to be in the domain of f in order to get the output which is $f(g(x))$. So, this is how you can always remember how to compute the composition of two functions and what are the necessary steps that are required to compute this composition.

IIT Madras

ONLINE DEGREE

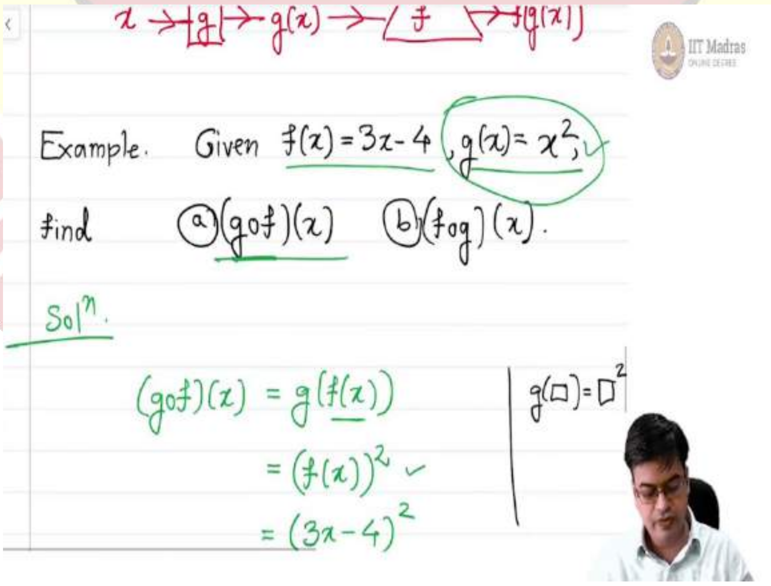
Mathematical for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 49
Composition Functions: Examples

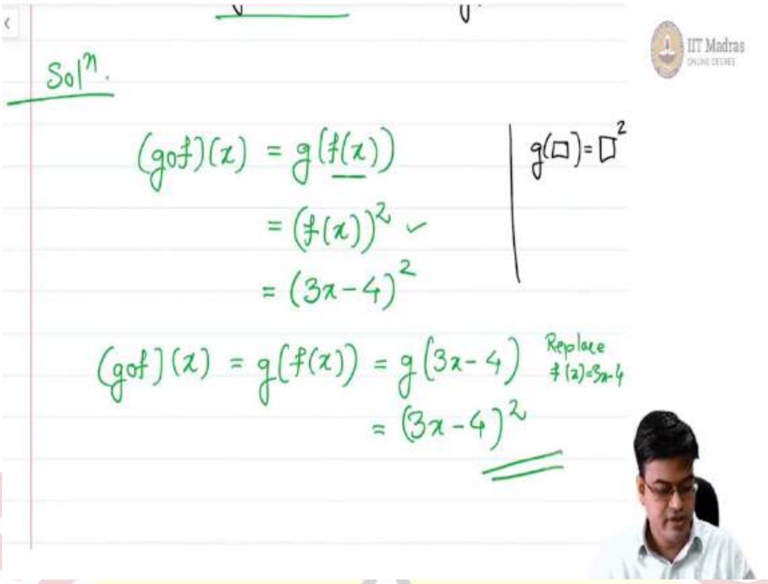
(Refer Slide Time: 00:15)



So, we have understood the theory, roughly the theory behind the function, composition, composite functions or composition of two functions. So, it's time to get some practice.

(Refer Slide Time: 00:29)





So, let me start with an example. And in that example, let us take you have been given two functions f(x)=3x-4, and g(x), which is equal to let us say x^2 these are the two functions that are given, then you are asked to find two things one is *g o f(x),* and the other one is obviously *f o g(x),* how to find this? Let us start, let us start with a solution.

So, what can be the solution let us take this function. So, let me write it properly, it is *g o f(x).* So, as per our theory, we have to write this as *g(f(x)).* So, *g(f(x)),* you can treat this as, what is *f(x)* now? $f(x) = 3x - 4$, and g(x) is x square. So, naturally *g(f(x)),* so, you go to this function, you treat this g as g. So, let me write it here, you treat this *g* as a *g* of a box, and *g(x)* is nothing but this box squared. So, in particular, if I want to write something about this function, this box right now has an argument which is *f(x).*

So, I will simply write this as f(x) squared, that is all. Now, the entire process is simplified. So, now, you do not have to worry about what g is, now it simply $f(x)^2$ what is the *f(x)* fit that when you in and you will get $(3x - 4)^2$. Another way to handle this is you can simply write *g o f(x)* as *g(f(x))* fit in the value of *f(x)* that is $g(3x - 4)$ and what is $g(3x - 4)$ as per our question, it is x^2. So, $3g(3x - 4)$ will be $(3x - 4)^2$.So, anyway whichever way is convenient to you, you proceed and you will get this answer correct.

So, what I have done here is I have replaced *f(x)* in this particular case, I have replaced $f(x) = 3x - 4$ in this particular case I have written *f(x)* and replaced what is *g(x).* So, both ways you can go now.

(Refer Slide Time: 04:01)

$= (3x-4)^2$

$$(f \circ g)(x) = f(g(x)) = 3\,g(x) - 4 = 3x^2 - 4.$$

$f(\Delta) = 3\Delta - 4$

Sol^n.

$$(g \circ f)(x) = g(f(x)) = (f(x))^2 = (3x-4)^2$$

$g(\square) = \square^2$

$$(g \circ f)(x) = g(f(x)) = g(3x-4) = (3x-4)^2$$

Replace $f(x) = 3x-4$

$x \to \boxed{g} \to g(x) \to \boxed{f} \to f(g(x))$

Example. Given $f(x) = 3x - 4$, $g(x) = x^2$,

find (a) $(g \circ f)(x)$ (b) $(f \circ g)(x)$.

Sol^n.

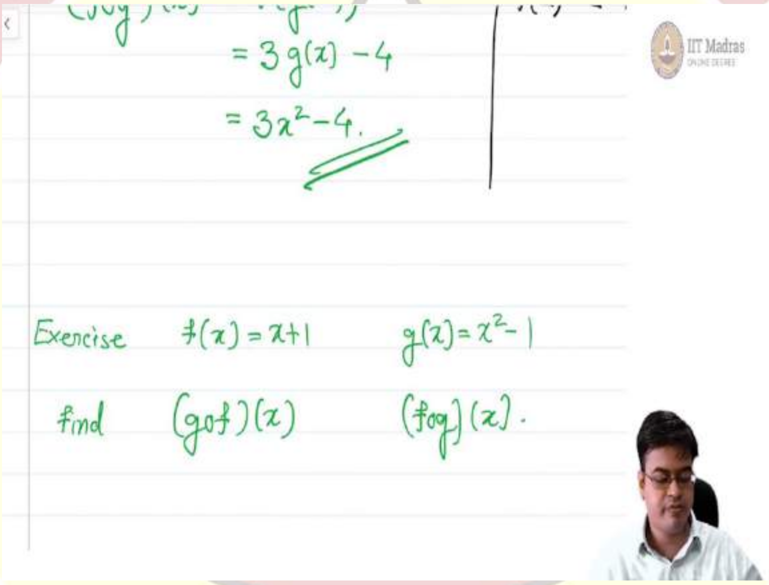
$$(g \circ f)(x) = g(f(x)) = (f(x))^2 = (3x-4)^2$$

$g(\square) = \square^2$

Let us go to the second problem that is a *f o g(x)* and *f o g(x)* is again can be written as *f of g(x).* Clear, there is no question, then there are two ways let us go it the first way, what is *f(g(x))?* So, what is *f(x)* here? $f(x) = 3x - 4$ here. So, I will write this as to be equal to $3g(x) - 4$.

So again, let me be very clear about this there should not be any confusion in this. So, what is $f(\Delta)$? Δ is an argument. So, this Δtriangle will be $3\Delta - 4$. So, now this triangle is replaced with *g(x)*, that is all. Therefore, your answer is *3x-3g(x)- 4*. But what is *g(x)?* Again, go back to the question *g(x)* is x^2 So, substituted here that means it will be $3x^2 - 4$ and this is the final answer for you in terms of *f o g(x).* So, we are seen how to write the compositions in both ways *g o f* and *f o g.*

(Refer Slide Time: 05:26)



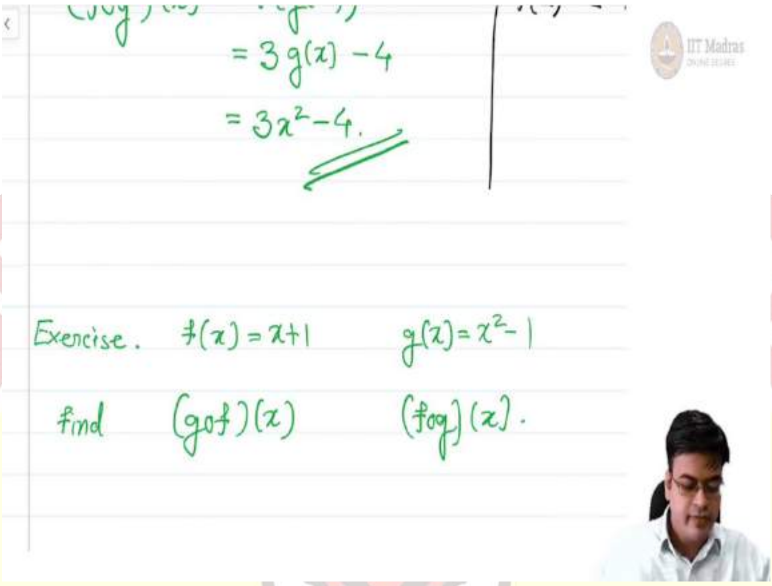
So, here is a quick exercise for you pause the video, do the exercise and get back the get the answer. So, $f(x) = x + 1$ and $g(x)$=$x^2 - 1$. Then simply find *g o f(x)* and *f o g(x)*. This is an exercise you stop and get the answer. It will be a good practice to revise the concepts.

IIT Madras

ONLINE DEGREE

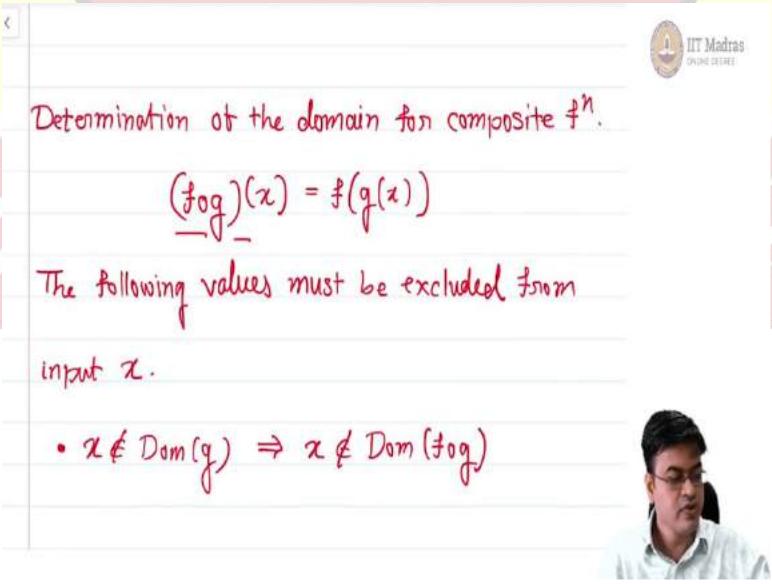
Mathematics for Data Science 1
Professor Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture 50
Composite Functions: Domain

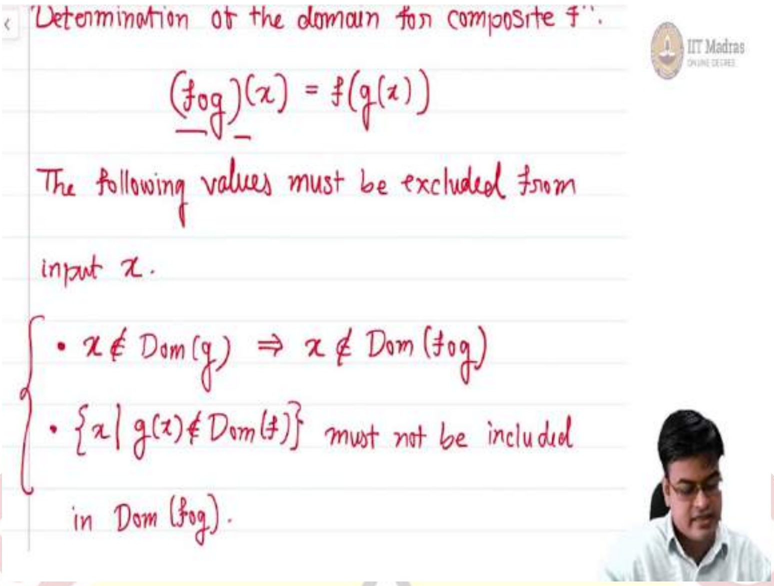
(Refer Slide Time: 00:15)

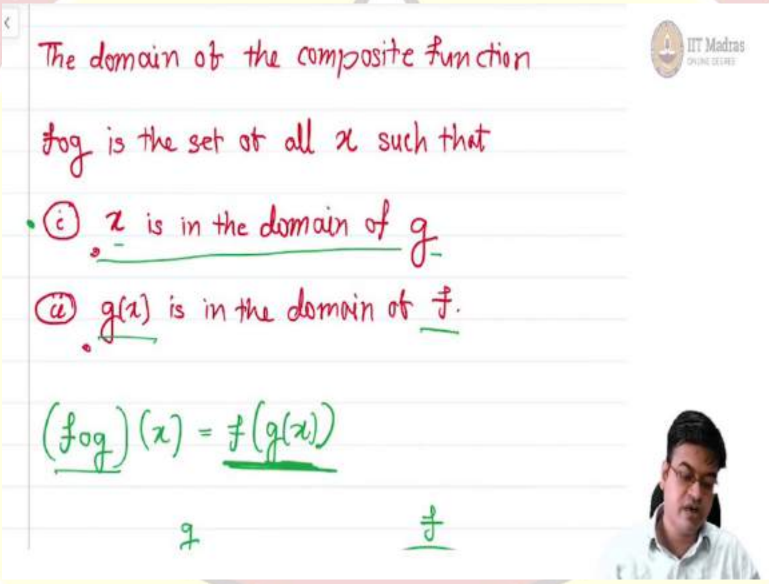


Let us now go further and talk about how to determine the domain of composite functions. So, this will be any important question that is determination of domain of a composite function.

(Refer Slide Time: 00:33)







Determination of the domain let us say for domain for composite function, how will you determine this? So, I have let us say $f \circ g(x) = f(g(x))$, we are talking about all functions that are real value. So, in order to determine the domain there must be some rules that you should follow I will list the rules and that essentially says the following rules, the following rules must be followed and therefore, the following values must be excluded from input values of x.

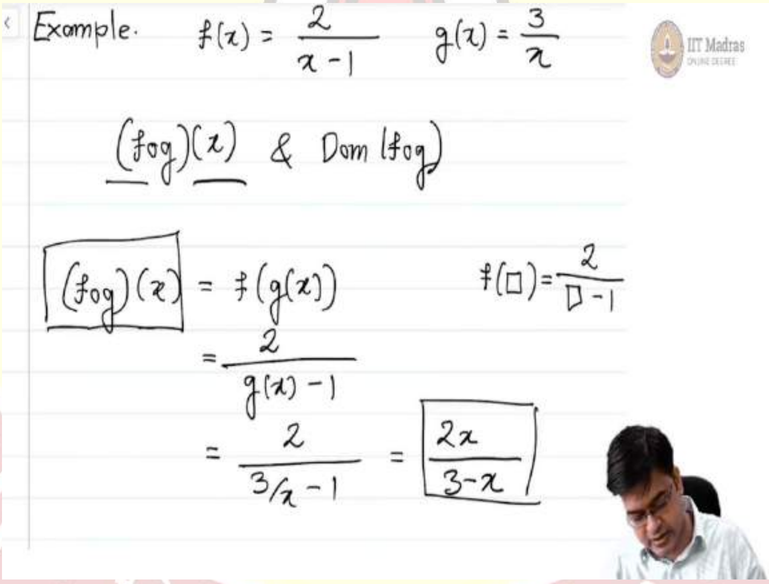
So, this is again in concordance with what we have seen earlier that if you remember we have seen some conditions right, where x should be in the domain of g and $g(x)$ should be in the domain of f. So, again what we are discussing now is in concordance with that, but here we were seeing what are the possible values.

Now, what we are seeing is what are the possible exclusions, that means, what value should be excluded from the input values. So, there are basically two rules the first rule which corresponds to the first rule of this that x should be in the domain of g that means, x if x is not in the domain of g then I cannot include it then x cannot be in the domain of the function $f\ o\ g$.

So, I am talking about $f\ o\ g$, when you talk about $g\ o\ f$, you will talk about the $x$belonging to domain of f implies. So, x does not belong to the domain of f implies x does not belong to domain of $g\ o\ f$. So, just remember the function the order in which they are taken it matters and in the similar manner, when I talked about $g(x)$ belonging to the domain of f. So, the set of all x's such that g $x$does not belong to the domain of f.

So, this is the set that you need to be careful about this set must not be included in domain of our function $f\ o\ g$ that is a composite function otherwise, we will have some ambiguity. So, in order to eliminate the ambiguity, we need to follow these two rules strictly very strictly. So, let me demonstrate how these rules can fail and then we will I will demonstrate it through an example and let me take that example as that is write it here.

(Refer Slide Time: 04:38)

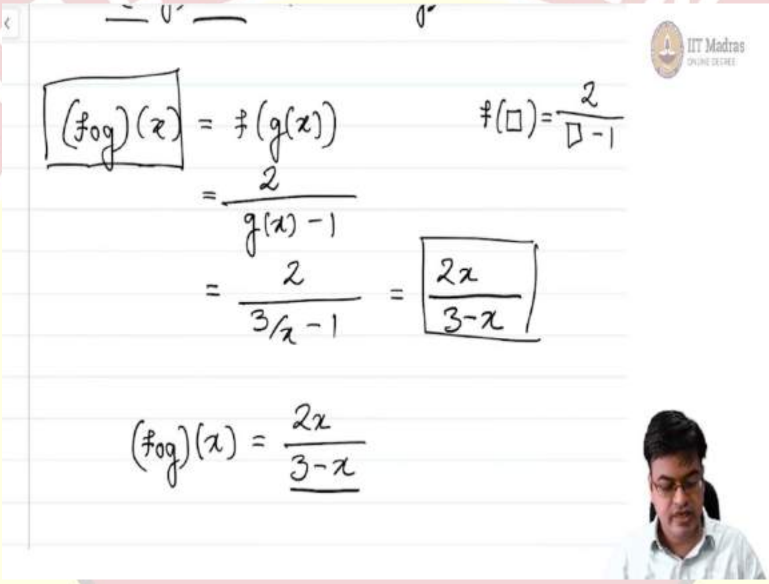


So example, so I have been given a function $f(x) = \frac{2}{x-1}$ and another function that is given to me is $g(x) = \frac{3}{x}$ and you want to find $f\ o\ g(x)$ and you also need to find a domain of this function $f\ o\ g$. What is the domain? Domain if you recollect from your week 1 it is nothing but the set of allowed values for which the function is well defined whatever input values you are fitting into the function, this function should be well defined this is the domain This is the notion of domain.

So, let us first see what is *f o g(x)?* And let us see if it gives you some hints about what can happen, correct? So, what is *f o g(x)?* Simply apply our definition it is *f(g(x)),* fine no confusion in this, then again you use that $f(\boxdot) = \frac{2}{\boxdot - 1}$. So, that gives me $\frac{2}{g(x)-1}$.

Now, what is *g(x),* it is $\frac{3}{x}$. So, substitute what is *g(x)?* So, it will be $\frac{3}{x-1}$, simplify this assume $x$is not equal to 0 and simplify this you will get $\frac{2x}{3-x}$. So, this is my *f o g(x).* Now, the question the second question that is asked is, so I have given answer what is a *f o g(x).*

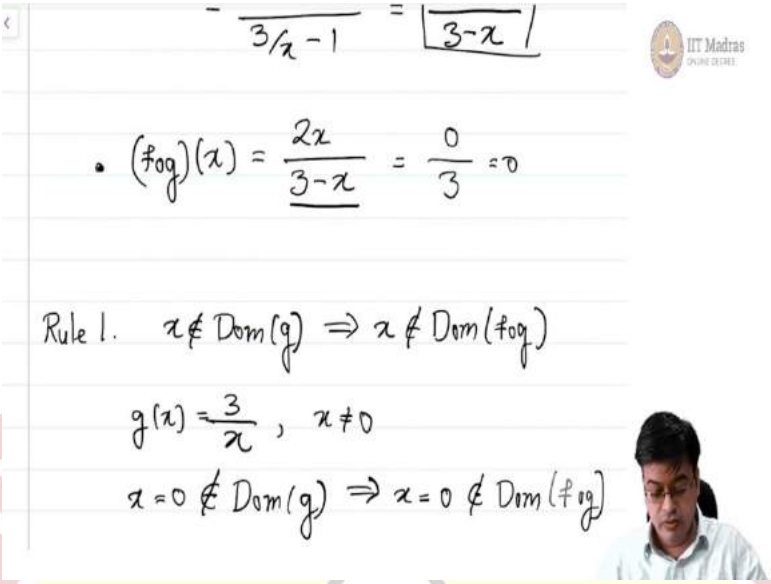
(Refer Slide Time: 06:52)

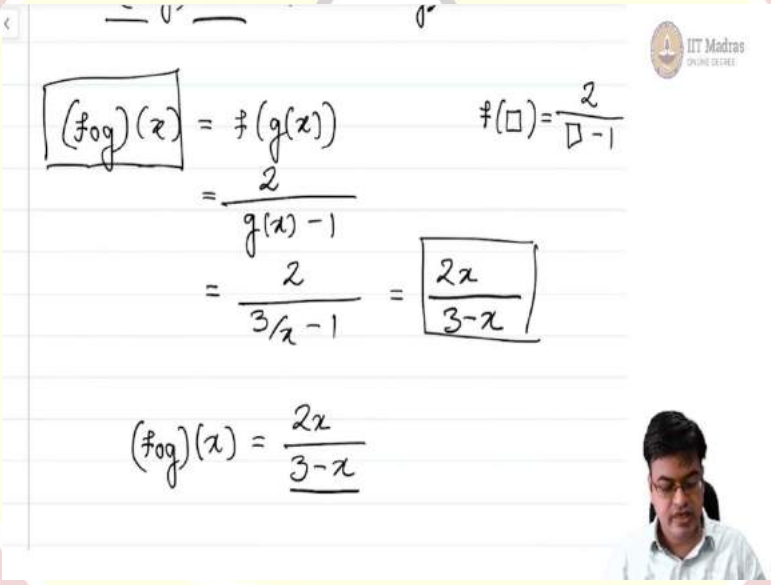


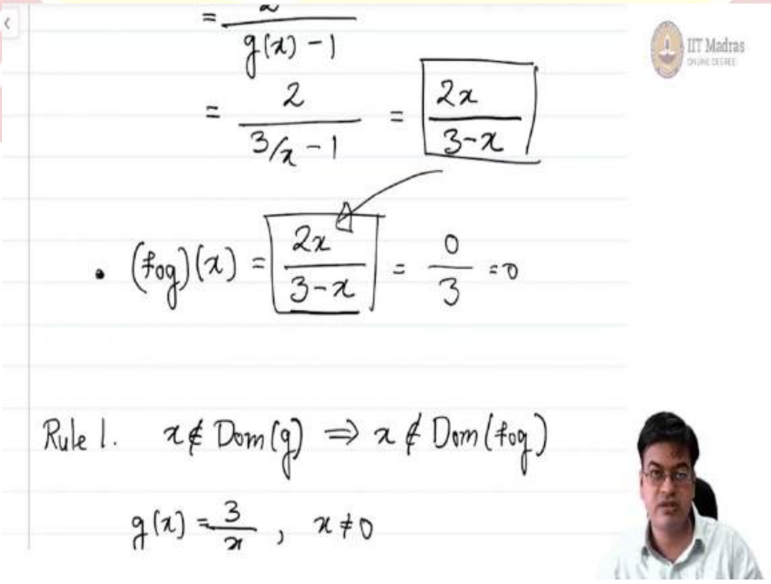
So, my $fog(x) = \frac{2x}{3-x}$. Now, if you look at this function, if you look at this function, you can simply see that at *x=3* this function is not defined, because the denominator is becoming 0. So, 6 / 0 is undefined. So, this function is not defined at *x=3*. So, the domain of this function must exclude 3 that is very well known.

But let us now see because of composition if I am eliminating any points, so here you look at this function which is *f(x).* And you look at this function which is *g(x)* and I am calculating *f o g(x).* So, if *x* does not belong to domain of *g*, then that function that particular value of *x* should not belong to domain of *f o g* that is the first rule that we have to implement.

(Refer Slide Time: 08:06)





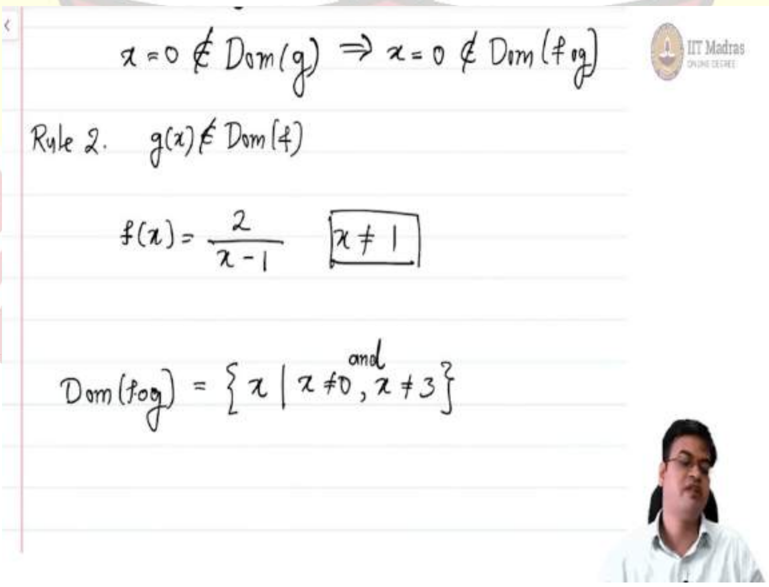


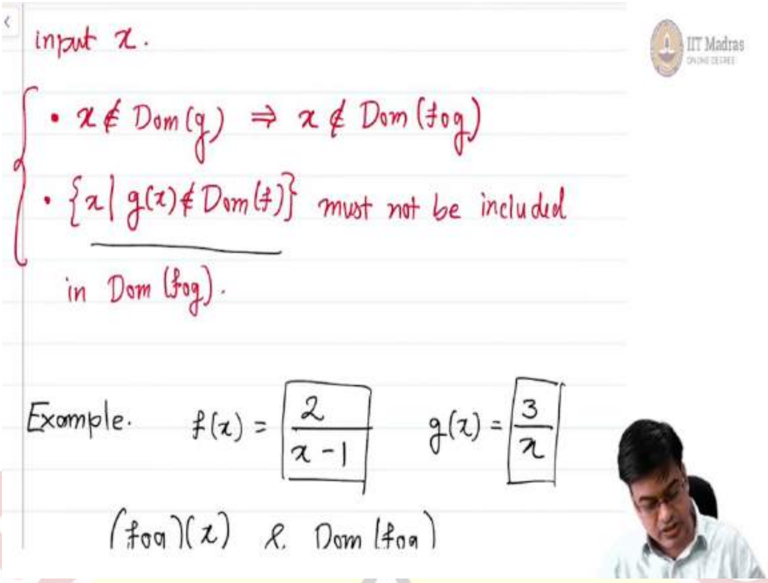
So, rule 1, what is the rule 1? If x does not belong to the domain of g that must imply x does not belong to the domain of f o g. So, what is that point? Let us look at what is *g(x)?* $g(x) = \frac{3}{x}$. So, this function is well defined only when $x \neq 0$. So, $x \neq 0$ not equal to 0. So, $x = 0$ cannot belong to domain of g. So, $x = 0$ do not belong to domain of g. So, naturally I will enforce that $x$equal to 0 should not belong to domain of $f \, o \, g$.

So, now, you may come up with some argument that when you look at this function, when you look at this function, if I substitute $x = 0$ if I substitute x=0, I am getting 0/3. Then this function is well defined because the answer is 0. That is what your argument will be. But no, why? I will tell you because when we when, when we were while we were coming to this particular form, what we were doing actually is we were multiplying a numerator and denominator by x or we are taking assuming $x$not equal to 0.

We are taking this x on the numerator on the numerator side and multiplying by $x$and that is where we have reached this point. If we had not assumed $x \neq 0$, then we would not have reached this point. Therefore, $x \neq 0$ is a valid condition still even when you cannot see anything visible over here, because I am composing the 2 functions where $x \neq 0$ is outside the domain.

(Refer Slide Time: 10:28)

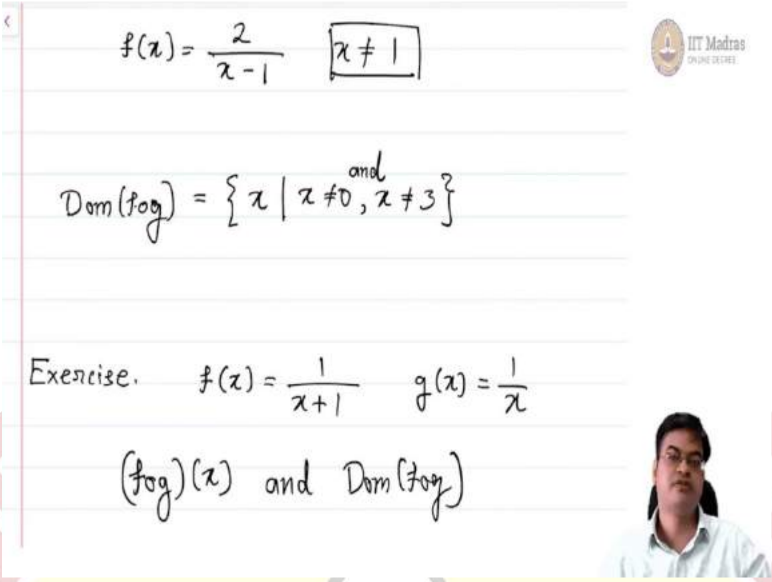




So, let us come to the next rule 2 that rule 2 was if *g(x)* does not belong to the domain of *f* then I am having a problem. So, that rule we have figured out like x says that *g(x)* does not belong to the domain of *f* must be excluded. So, let us look at our function *f* what is our function f it is 2 upon *x*minus 1 in this case *x=1* I have a function where the denominator is 0.

So, x is equal. So, let me write for the sake of completeness $f(x) = \frac{2}{x-1}$ this is well defined when $x \neq 1$. So, so this also this point $x \neq 1$ should also be eliminated from the domain of *f o g*. So, what should be the domain of a *f o g*? All other points the function *f* and *g* are well defined. So, domain of *f o g* must be set of all *x*'s belonging to real line such that $x \neq 0$ and $x \neq 3$, this comma means and or if you want me to be precise, I will write.

(Refer Slide Time: 12:02)

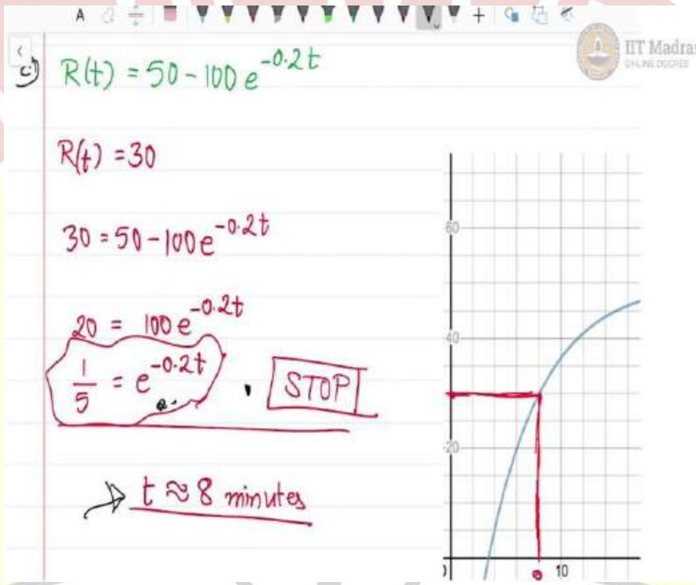


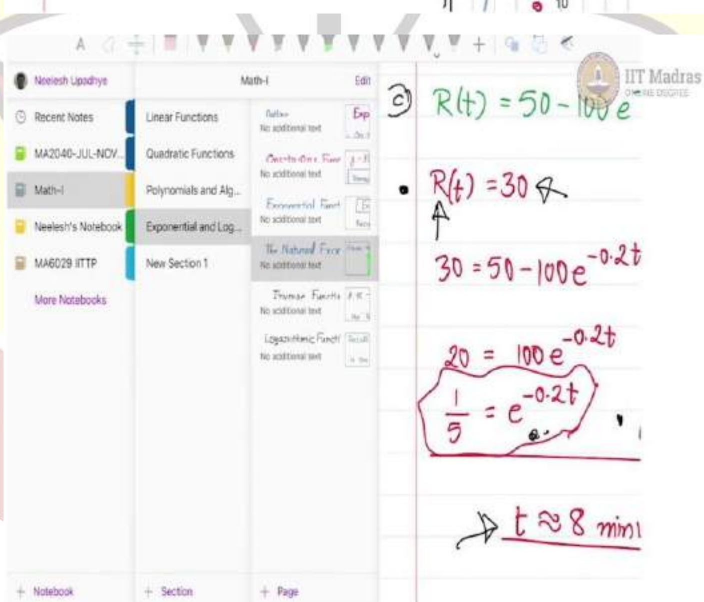
Another quick exercise that you can do in order to verify whether you have understood the concept of composition of function and the domain is you have been given 2 functions $f(x) = \frac{1}{x+1}$ and $g(x) = \frac{1}{x}$ and you are asked to find $f \; o \; g(x)$ and the domain of $f \; o \; g$. So, you can quickly solve this problem and check whether you have understood what we are supposed to understand. Thank you.

IIT Madras
ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture No. 51
Inverse Functions

(Refer Slide Time: 0:14)



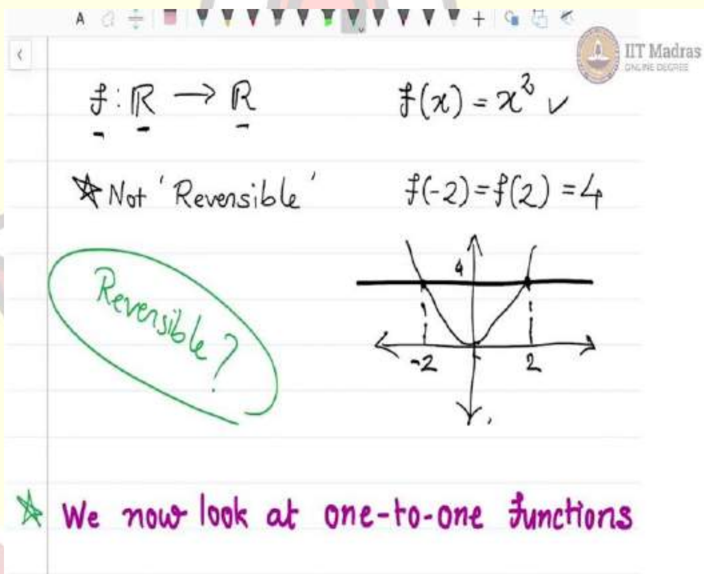


Hello Students, in the last video we have stumbled upon one concept, where we could not proceed. Then we came to…Let us go to the last videos last slide. So, here if you look at this particular concept we actually stopped while computing. And why we stop while computing is because we did not have enough information on, how to write t is equal to something given this equation.

So when, what we did we found escape out by plotting the lines across x and y a$x$is, horizontal and vertical lines and figured out that the answer is 8 minutes. And that is how we concluded this is 8 minutes. Now when we start such a thing analytically that is R_t is given to be 30. What is the value of t? We want to answer such questions then we need to look at the function R and we need to understand whether this function is reversible or not.

Which is the case, in this case, in this particular function because we were able to map it uniquely. So, what are the important trades of this function R_t? R_t was a one to one function and it was increasing function. Hence, it was one to one. Therefore, we were able to find a reversal of the value 30 to the value of t which is 8.

(Refer Slide Time: 1:58)



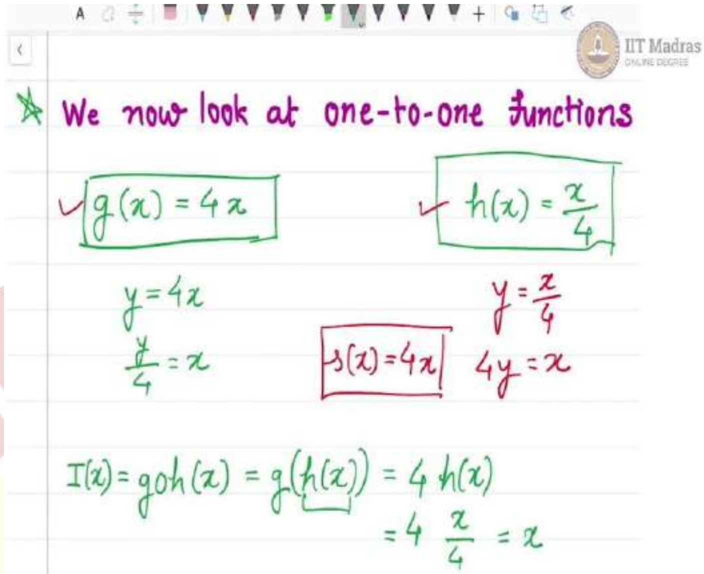
So, in order to find such reversible functions, we need to understand the theory which we will discuss now is the theory of inverse functions. So, when I talk about inverse functions, I am talking about functions from domain which is real line co domain which is real line. So, a function is define from real line to real line, then the immediate question that comes to our mind, are all functions reversible? And the immediate answer is a very well-known function that we have seen is, $f(x) = x^2$.

Now this function is not reversible because it feels to pass the horizontal line test, if you remember. So, $y = x^2$, if I try to plot, it will be something like this. Very close to something like this. And when I pass a horizontal line through this it passes through 2 points. And let these points be 2 and -2 . And that essentially means this, when I feed in the value 2, it will give you four. And when I feed in the value - 2, it will give you the answer to be equal to 4.

Now if this function is reversible, when I feed the value to 4 it can spit out the two values 2 and -2. So, it is not uniquely spitting out the value. Therefore, this function is termed as not reversible function. Such functions we cannot study the inverse properties or the properties of inverse functions. However, if you restrict the domain of this function instead of real line to only positive half of the real line, then you will get one to one correspondence between the values on x a$x$is and y a$x$is and then you can talk about inverse of these functions, when it is defined from 0 to ∞.

Now let us look, then the question is, this function is not reversible then which functions are reversible? That is a question that we can ask now, in order to answer this question, we need to study some class of functions. So, in last few videos we have already seen that one to one functions are nice functions. Any function that is either increasing or decreasing is one to one and therefore we can look at one to one functions for the class of reversible functions. So, here is our answer that we will start looking at the class of one to one functions.

(Refer Slide Time: 4:51)

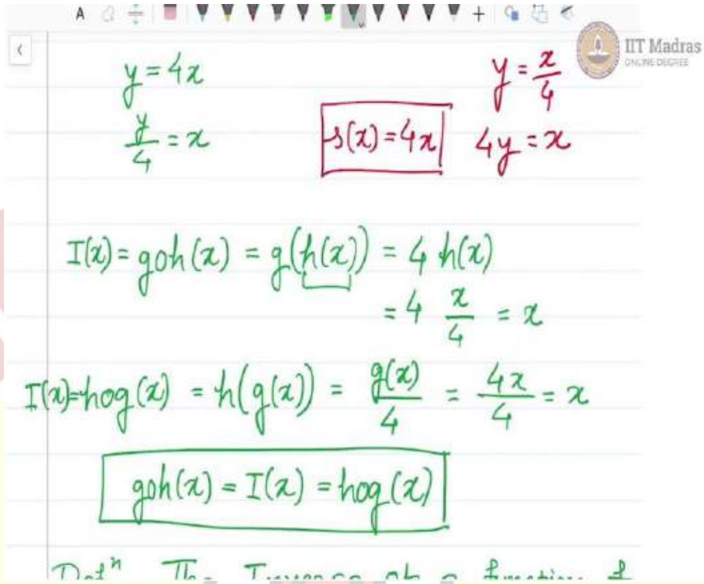


Let us look at a simple function a linear function g(x) is equal to $4x$. Is this function reversible or not? So in order to answer this question, let us look at $g(x) = 4x$. So, you can put $y = 4x$. If you look at $y = 4x$ from our basic understanding of linear equations or rather than linear equations an equation of a straight line. This is a straight line passing through origin having slope 4. So, if I want to find a point x on this $ax$is then I will simply transform this as $\frac{y}{4} = x$ and this transformation is unique. Therefore, I can write some function let us say *r(x)* as $\frac{x}{4}$. And this function will actually be giving be the inverse of this.

So, let us take this function, if this function $h(x) = \frac{x}{4}$. So, I do not need to write *r(x)*= $\frac{y}{4}$. $h(x) = \frac{x}{4}$. Now if I start with this function and I want to get value of x, what should I do? I will write, so I will write $y = \frac{x}{4}$ and in that case I will get $4y = x$. And therefore, I will get another function which is say $4x = s(x)$. So, essentially what we have seen is this g(x) and h(x) have something in common. So, let us recollect the notion of composition of two functions, and try to answer this question.

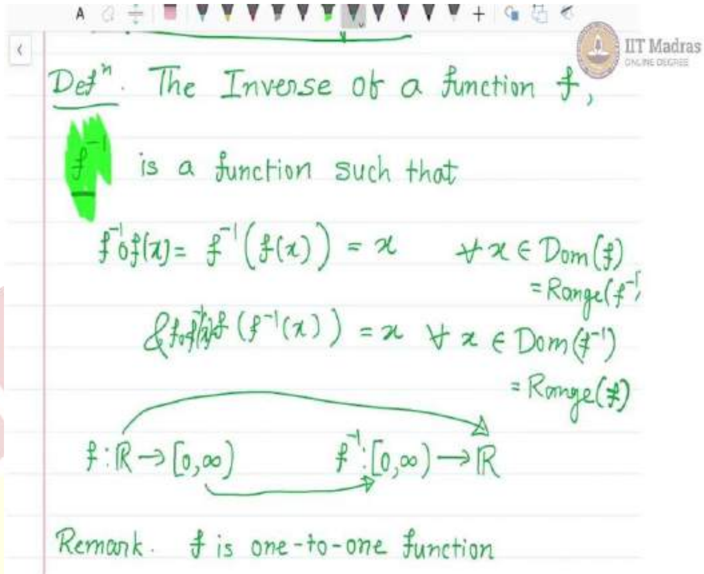
For example, if I consider the function goh(x). Now this function is again a function and it will simply operate like g of h(x). And once you start with g of h(x), what you will do is, you will treat h(x) as an argument of g and put the values of h(x) inside. So, let us try to understand this, so it is like g(*h(x)*) is actually, what is *g(x)?* 4 of x. So, it will be 4 $\times$ *h(x)*. Now what is $h(x)$? $h(x)$ is nothing but $\frac{x}{4}$ So, 4 times x by 4 which will give me x. So, what this function is, this function actually gives me identity function. $goh(x) = x$ and a similar manner

(Refer Slide Time: 8:08)



I can start thinking about *hog(x).* now in this case, if you recollect the notion of composition of functions studied in week one, *h(g(x)).* So, if h of g(x) I will simply see what is *h(x).* $\frac{g(x)}{4}$ and therefore what is g(x)? It is $4x$ therefore I will get $4x$/ 4 which is actually equal to x. And therefore this is also equal to identity function of x. So, now to summarize what I got is *G o h(x)*= $I(x) = hog(x)$. Now this becomes our definition of inverse function. And let us define it formally as the definition of inverse function.

(Refer Slide Time: 9:16)

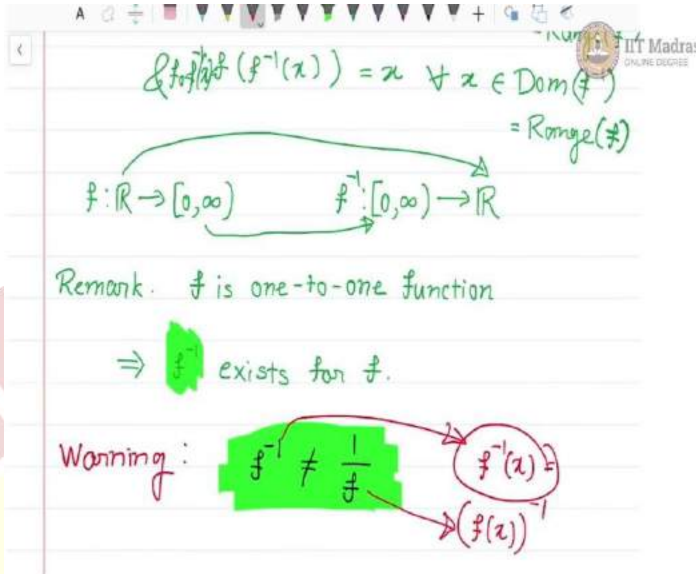


So, here is a definition of inverse function. The inverse function inverse of a function f, we denote it by .$f^{-1}$is actually a function this is our notation f^{-1} is actually a function such that.$f^{-1}f(x)$ or I can rewrite this as $f^{-1}f(x) = x$. Now here is a typical thing that comes for all x belonging to domain of f which is equal to range of f^{-1}. And $f(f^{-1}(x))$ or you can write this as $f(f^{-1}(x))$ being equal to $f(f^{-1}(x)) = x$ for all x belonging to domain of f^{-1} and range of f^{-1}

So, right now when I did this particular calculation I have assume that everything goes from real line to real line there was no such event. Because this function is define from real line to real line. And this function is also define from real line to real line. So, there was no consideration for domain and ranges. But sometimes it may so happen that your original function maybe define, let us say f is define from $\mathbb{R}$ to$[0,\infty)$. If such a definition is there, then you need to worry about the domain of a function and the range of a function. Because here the domain of f is $\mathbb{R}$ and range of f is 0 to ∞.

So, if I talk about f^{-1} of this, then naturally I cannot go over entire real line. I have to go over 0 ∞ and then I have to come to $\mathbb{R}$. So, this is how it will be define and therefore, the domain of f will become the range of f will become the domain of f and the domain of f will become the range of f^{-1}. This is the typical factor that you need to always remember. Now let us go ahead and improve our understanding about one to one functions.

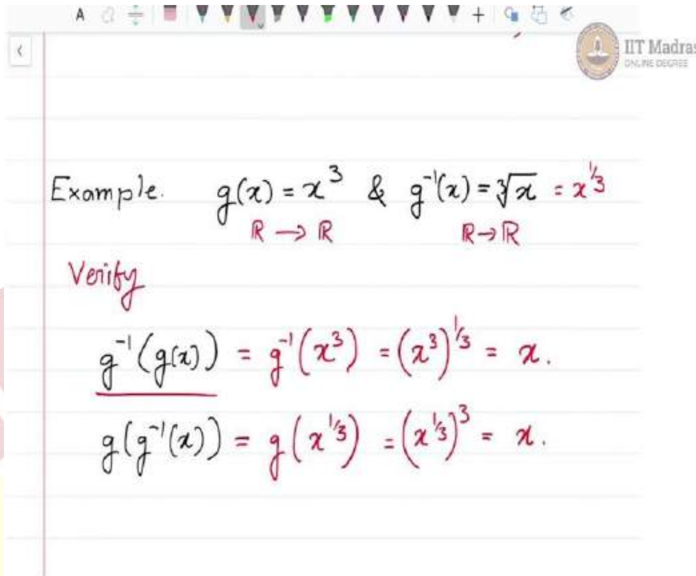
(Refer Slide Time: 12:00)



So, if the given function is one to one function then f^{-1}always exist for *f*. Now the notion may confuse you. So, let me give you one precise warning that the notion f^{-1}does not mean $\frac{1}{f}$. This is very important. Because you may quite often confuse f^{-1} with $\frac{1}{f}$. So, whenever we want to discuss in this course or in Mathematics, whenever we talk about $f^{-1}(x)$ it is simply means it an inverse function.

So, this is an inverse function and whenever you want to talk about the $\frac{1}{f}$. Then you should talk about $f(x) - f^{-1}$. So, this $f(x) - f^{-1} = \frac{1}{x}$ and this f^{-1} is actually has a meaning $f^{-1}(x)$ with this you always remember. Now if f is one to one function f^{-1} always exist for *f*. This you have to trust me. I cannot prove it right now with the current tools, so f^{-1} always exists.

(Refer Slide Time: 13:11)

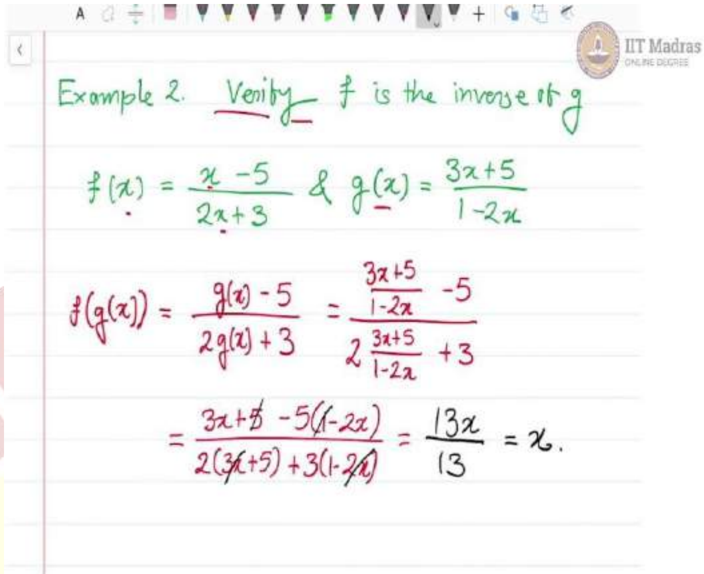


Let us take one example g(x) is equal to x cube and g^{-1} of x is $\sqrt[3]{x}$. This you can write as x raise to 1 by 3 as well. So that, this is simple to verify. So, now you want to verify that the given functions are actually inverses of each other. So, in this case let us first identify the domains, it is a real line to and range is real line. So, naturally for inverse also it's real line to real line. Now question about it. So, let us talk about g^{-1} $g(x)$.

Now if you recollect the definition of inverse function then naturally the inverse function is a function such that all this combinations, all this combinations should produce x $f^{-1}(f(x))$ or $f(f^{-1}(x))$ f. So, let us talk about $g^{-1}g(x)$). So, let us keep g^{-1}intact and put what is $g(x)$ which is x^3. Now this you substitute the function $g^{-1}$of x as $x^{1/3}$ then this becomes $(x^3)^{1/3}$. Then multiplication of indices a^{mn} applicable, so it will x. So, one way it is true.

Now the second way also you have to check. So, what you will do now, is you just write g within the box you write $g^{-1}(x)$ here, x raise to 1 by 3. And then simply put the function g so x raise to 1 by 3 the whole thing raise to 3 which again a raise to m n. So, this will also give you x domain and ranges we have already seen. So, whatever the conditions of that therefore g and $g^{-1}$are inverses of each other. So, $g^{-1}$is inverse of g .

(Refer Slide Time: 15:24)



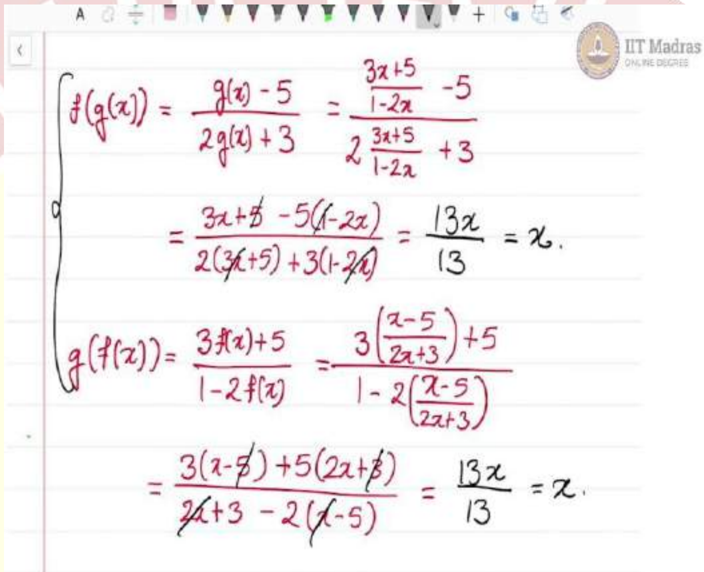
Let us take this example where we are suppose to verify, whether f and g are inverses of each other. So, let us try to verify, you can check the domain and co ranges of these functions. I will simply start with $f(g(x))$. So, if I start with f $(g(x))$ as per our notion what we will do? We will simply keep this g x in place wherever f has an argument x. So, we will take this and we will put g x wherever x is written there.

So, let us do that exercise that is, $\frac{g(x)-5}{2g(x)+3} = \frac{\frac{3x+5}{1-2x}-5}{2\times\frac{3x+5}{1-2x}+3} = \frac{3x+5-5(1-2x)}{2(3x+5)+3(1-2x)} = \frac{13x}{13} = x$. this is f o g(x). Now what is g(x)? g(x) is this so let us go ahead and substitute those values over, those functions in place of $g(x)$. So, it is $\frac{\frac{3x+5}{1-2x}-5}{2\times\frac{3x+5}{1-2x}+3}$. So, now it is a matter of your Algebra just simplify this. So, denominator both have $1-2x$ in common, so multiply the numerator by $1-2x$ and denominator by $1-2x$. So, that we will get rid of this. So, it will be $\frac{3x+5-5(1-2x)}{2(3x+5)+3(1-2x)}$.

So, what is a question, we have to verify that f is the inverse of g So, essentially I want to come up with a number with a function which is x f o g(x)= x, this is my end goal just remember this. Now you can simply (multi) simplify this $3x$ + 5 - 5 will get rid of this + 5. Let me change the color over here. So, this will get this one will get rid of this then this is - $2x$ - $10x$ and - $10x$ will become + $10x$. Because of this - sign and then 3. So, I will simply get here $13x$.

Now you look at the denominator which is 2 into $3x$ that is, $6x$ then look at the corresponding term here - $6x$. So, this x, terms corresponding to x will vanish and 3 and $2 \times 5 = 10$. So, I will give get the denominator to be equal to 13 and that will give me x as my answer. So, f o g(x) is verified. Does this complete, will this complete our verification of whether f is the inverse of g? No, because I want to check whether g is also the inverse of f. Then only the verification will be complete.

(Refer Slide Time: 19:03)

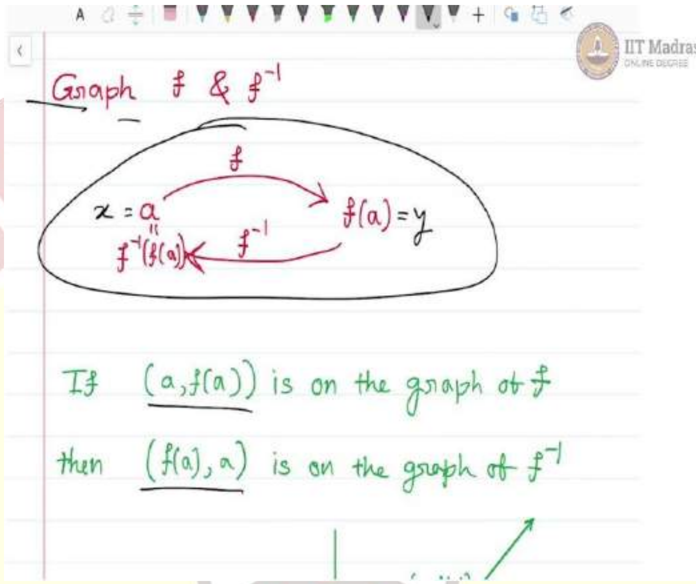


So, let us go ahead and do that, that is we will consider g f of x and that should give me x that is my end goal. So, now you look at what is a function g and put f x as it is everywhere. So, it is 3 times f x + 5 1 - 2 times f x. What is the next step take the functional form of f x and substitute it in the expression.

So, 3 into x - 5 upon $2x$ + 3 + 5 upon 1 - 2 times x - 5 upon $2x$ + 3 and then again the same logic applies multiply both sides by the $2x$ + 3 and then you will get 3 times x - 5 + 5 times $2x$ + 3 to be, upon 2 1 is there. So, $2x$ + 3 as it is - 2 into x - 5. Let us look at the simplified form let me change the color. So, $3x$ + $10x$ that will give me $13x$ here 3 into 5 - 15 + 5 into 3 + 15. So, this is taken care of vanished upon again the same logic applies $2x$ - $2x$ will vanish 2 into 5 will give me 10 and this 3 will give me 13.

Therefore, I got this domain and ranges you have already verified for yourself and therefore process is now complete because f of g(x) is x g o f of x is again x. So, we can verify that f is inverse of g as stated. So, this completes our discussion on inverse functions.

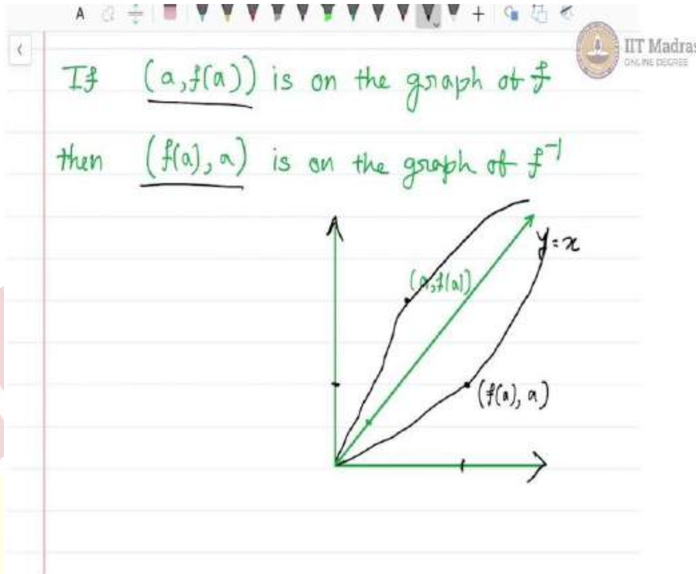
(Refer Slide Time: 21:40)



Now it is important to understand graphically what the inverse function is or how the graph of f and $f^{-1}$is changes. So, we already have a wage understanding of the graph of f and f^{-1}. Now let us look at it formally, so if I know something about f or the graph of f, then given a value of a. I am able to calculate f of a and a f of a is the payer which we call as graph of x, graph of f. You look at f^{-1}, what happens when you talk about f^{-1}, here f of a is actually on y a$x$is and a is on x a$x$is.

So, when you look at the inverse function the values on y a$x$is actually get convert into values of x a$x$is. And the values on x a$x$is will actually get converted into values on y a$x$is. So, this is the mapping that we have given. So, if you start with f of a which is y then you will talk about $f^{-1}(y)$of y and when you talk about $f^{-1}(y)$ you will actually get it to be equal to a. Because $y = f(a)$. and this is how the entire circle is complete. So, in particular if a and $f(a)$ is on the graph of f then f(a), a is the graph of f^{-1}. That is obvious.

(Refer Slide Time: 23:17)

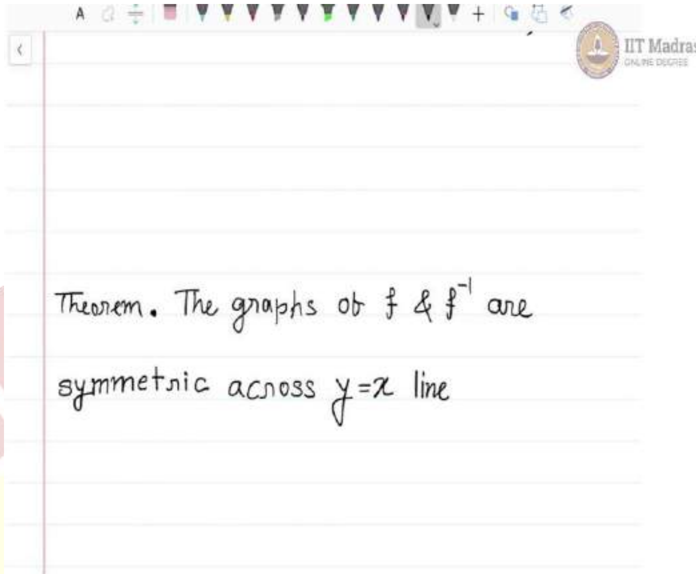


So, let us look at, let us imagine this is the graph. This is the graph of a straight line a, fa. So, you plot a line $y = x$ here. This is a line in $y = x$ and this is a point which is on the graph of $f(x)$. So, now you are saying that where will this point be, when I talk about f inverse. So, then we are also answering this question that wherever a was there it will be *f(a)* now. And wherever $f(a)$ was there now there will be a.

So, in this case you just take this distance and you plot it, you just take the distance on the y a$x$is and choose the distance on x a$x$is here and take the distance on x a$x$is for this point and put that distance over here. That means it will be somewhere here. And therefore, the point will be somewhere here and this point is actually $f(a)$. So, what we are actually doing when we are plotting the graph is actually we are reflecting our original function, in the original function is somewhat like this. Let us say, so it is somewhat like this.

Then what we are doing is we are actually reflecting it along y a$x$is and it will be very similar function. Which will look like this. So, this is how the graph of inverse function will look like it is actually a reflection along y is equal to x or reflection along a function y $f(x)=x$.

(Refer Slide Time: 25:13)

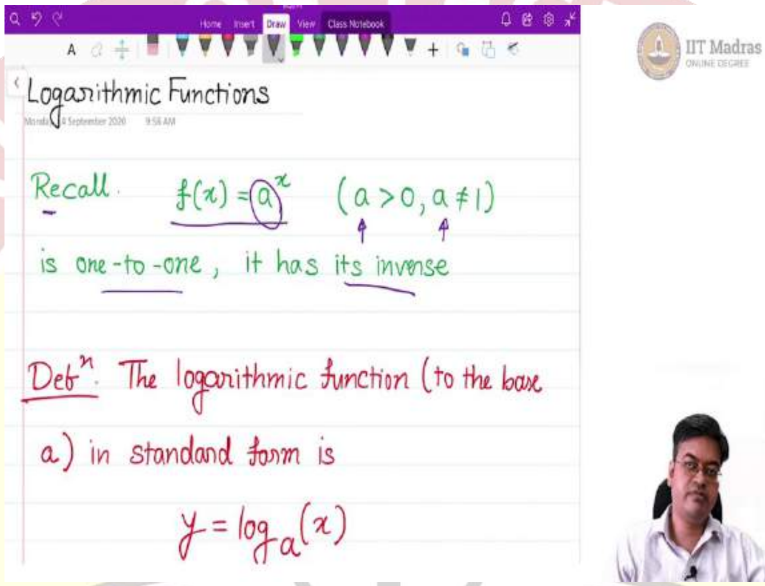


So, in particular the graph of f and f^{-1} are symmetric across the line y= x. This is what you have to remember all the time. If you want, you can prove the theorem but there is nothing it just a graphical prove that, if I want to compute this particular point and if I know that the inverse of this function e$x$ist, then you just take length on y and plot it across x and length of, length in x direction plot it across y direction. That is what I did a this is actually the prove of the theorem that the graph of f and f^{-1} are symmetric on y =x line. That completes our topic on inverse function. In the ne$x$t video we will deal with the inverse functions in a more restricted manner that is, inverse of this e$x$ponential functions.

IIT Madras
ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture No. 52
Logarithmic Functions

(Refer Slide Time: 0:14)

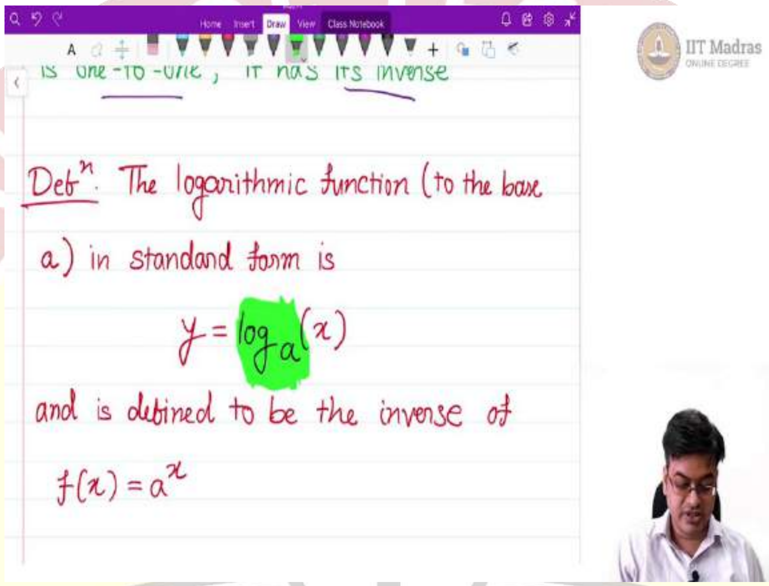


So, in this video we are going to look at the inverse of exponential function. In the last video we have seen the inverse of a general function and we have concluded that if the function is one-to-one, then the finding the inverse of a function is very easy. So, let us focus on inverse of exponential function in this video and see its properties graph or how it is graphed and a various other properties about domain and range of these inverse functions for exponential functions.

So, let us recall our notion of exponential function, we started with a function which is a function will be called as exponential function if it is written in the form $f(x) = a^x$ where there were some conditions on a, for example, a should be greater than 0 and a cannot be equal to 1, a greater than 0 is a typical condition which we need because otherwise we have to deal with complex random, complex variables which is out of scope of this course.

So, we are putting a to b greater than 0 and a≠1 is the condition because if you put a=1, then $f(x) = 1^x$ which is 1 for all of them, so it is not an interesting function to study. So, whenever these conditions are enforced we know that our exponential function $f(x) = a^x$ is one-to-one and because every one-to-one function has the inverse this function also has the inverse, there is nothing special about it. And that inverse we will define as logarithmic function. So, naturally since we are talking about exponential function with base a so we will talk about logarithmic function with base a.

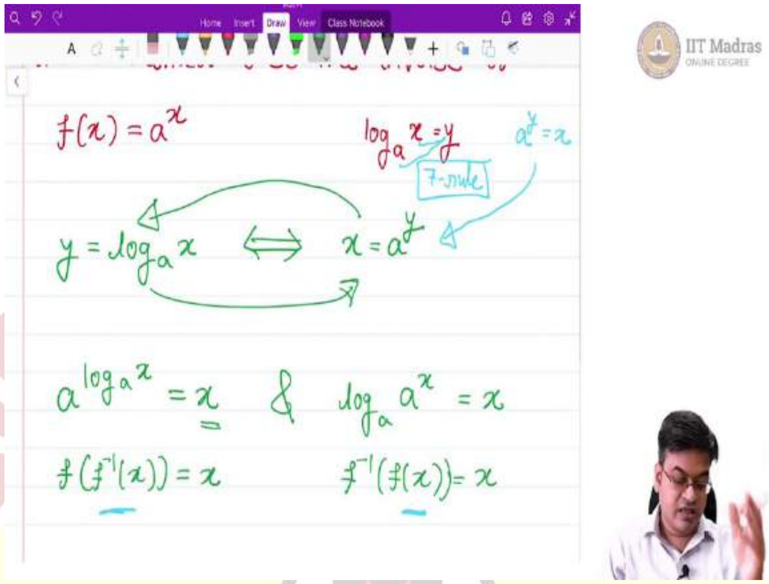
(Refer Slide Time: 2:16)



So, here is a definition of a logarithmic function. The definition says that the logarithmic function to the base a in the standard form is given by y=$\log_a x$. So, remember this function is represented by log to the base a and x is the argument of the function, so this is the definition of a function or this is replacing f, $f^{-1}(x)$and then x is the argument and we are plotting it along y axis and is defined to be the inverse of the function $f(x) = a^x$.

So, $f^{-1}(x)$is actually $\log_a x$, is this simple. So, now we need to understand what will be the domain and codomain or range of this function that is an important thing that we need to understand. So, in order to that let us try to devise some rule so that we will have a track of what is exactly happening when we are talking about logarithmic function and how it is related to exponential function.

(Refer Slide Time: 3:34)



So, there is a one to one correspondence between logarithmic function and an exponential function which is expressed by this relation y=$\log_a x$ if and only if x= a^y or for more precision you can write this as $\log_a x$ =y then you can actually virtually assume this 7 rule that is you start from the base, go to the right hand side and come back that means what we are saying is you start with a, go to the right hand side, that right hand side is raise to the power and that should give you x, that is what this rule is.

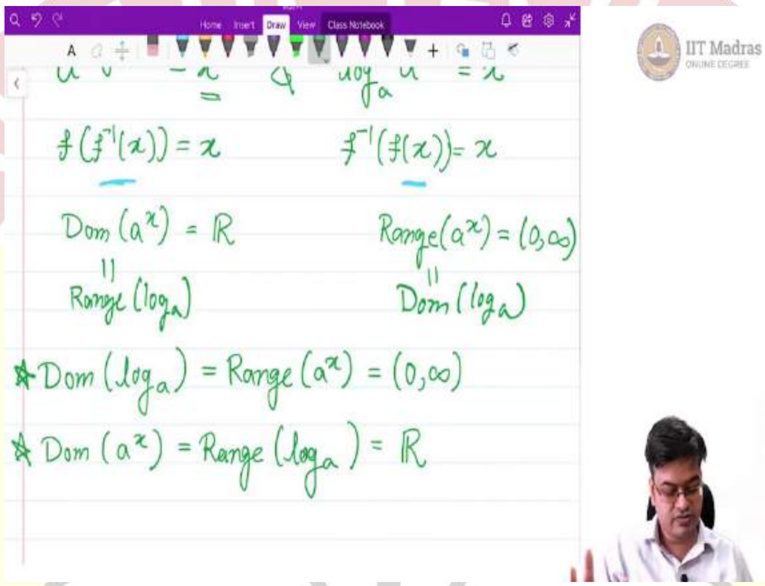
So, this is simple technique to remember known as 7 rule. So, you can use this 7 rule to memorize the one-to-one correspondence between log and the exponential function. You can easily see that by definition if I write x= a^y, then I want to know the value of y, I should be able to get it by taking the log of this function x.

So, this is the mathematical definition of our logarithmic function. To make this mathematical definition precise we need to understand some prototypes that is whether this function we have defined it to be the inverse of f but whether this function is actually the inverse of f or not that is what we need to figure out.

So, as stated earlier we can actually check these two rules $f(f^{-1}(x)) = x$ and $f^{-1}(f(x))$. So, what is$f(f^{-1}(x))$? As I mentioned earlier $f^{-1}(x)$ is nothing but $\log_a x$ and f is a^x so you just substitute $a^{f^{-1}(x)}$. What is that? $a^{\log_a x}$. Now, what this should be? You use this one to one correspondence from here to here and here to here and you will get this to be equal to x.

In a similar manner you can apply it to f of x and$f^{-1}(x)$ f inverse, so $f^{-1}(f(x))$ is $\log_a f(x)$ but what is $f(x)$? It is $a^x$and therefore $\log_a a^x = x$.

(Refer Slide Time: 6:27)



Now, in order to understand this completely I need to understand the domain of log function and range of log function and the range of log, range of exponential function and the domain of exponential function. So, let us understand this particular thing. We have already seen what is the domain of a^x, so we already know domain of a^x because x can be entire real line and then it maps this domain onto the range of a^xthat range cannot take negative values, this is what we have seen when we studied.

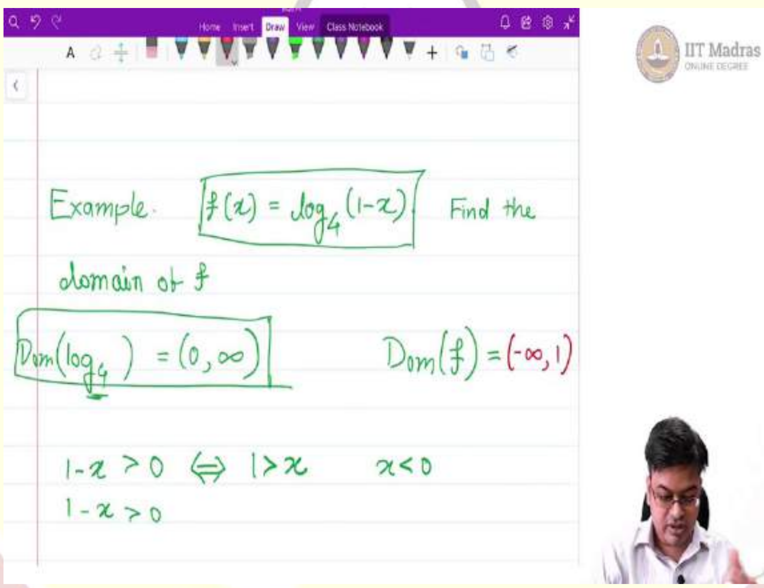
So, it was 0 to ∞, so this should be clear before going to the range of log function. So, if at all the logarithmic function is to be defined, this if you recollect this should become domain of log to the base a and this should become the range of log to the base a, so the this is the crux of the definition of inverse. So, when this is satisfied you are done.

So, essentially your log function will be defined from 0 to ∞ to real line. That means in the domain it cannot have negative values, it cannot have 0 as well and in the range it will have the entire real line that is what is written here in this case that is domain of log to the base a is actually range of

a^xwhich is 0, ∞ and domain of $a^x$is actually the range of log to the base a which is real line, the entire real line.

These are the two important points which will help you in understanding the domains of the functions which are derived from these functions that is logarithmic functions or exponential functions. So, these, all these things you should always remember the valid ranges and domains of the function. So, this completes our verification that logarithm function the way we have defined is actually an inverse of exponential function. Once the verification is complete let us dwell more and find the domain of the derived functions, derived, by derived functions means composition of basic logarithmic function.

(Refer Slide Time: 9:04)

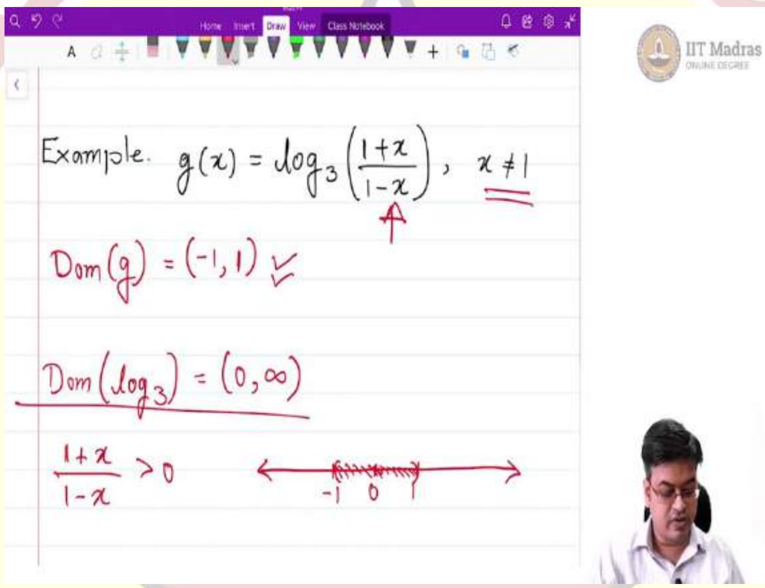


For example, let us take an example of f x which is log to the base 4 of 1 -x. Now, log to the base 4 is actually a function which has a domain. What is the domain of this function? The domain of this function is actually 0 to ∞. Now, that means the argument that is supplied to this function log to the base 4 cannot be 0, or it cannot be a negative value. So, based on this understanding from the definition of our log function you can look at this function which is f of x and look at the argument of the function 1 -x.

According to this definition 1 -x must be strictly greater than 0. This will happen if and only if my 1>x, 1>x and because 1 -x needs to be greater than 0 can x be less than 0, if you look at x to be less than 0, 1 -x will actually be greater than 0. So, the only condition that we require over here is my function should be defined that is domain of this function f should be equal to, it cannot include 1, 1 to, it is not 1 to ∞, this is how we commit mistakes.

So, domain of f is x should always be less than 1 that means the domain of this function should be here -∞ to 1 and it cannot go beyond 1 this is what our understanding is about this function. Now, let us go and enhance our understanding in finding the domain of a function which is slightly more complicated than this function.

(Refer Slide Time: 11:26)



So, our question is to find the domain of this function g. In order to find the domain of this function g, let us first understand what is the domain of the function log to the base 3. Now, this function is defined when the argument given is between 0 to ∞. So, now I want the argument of this function which is this gx to be between 0 to ∞. So, what I should do is I want this 1+x upon 1 -x trapped between 0 to ∞ that means it should be greater than 0. Now, when this can happen?

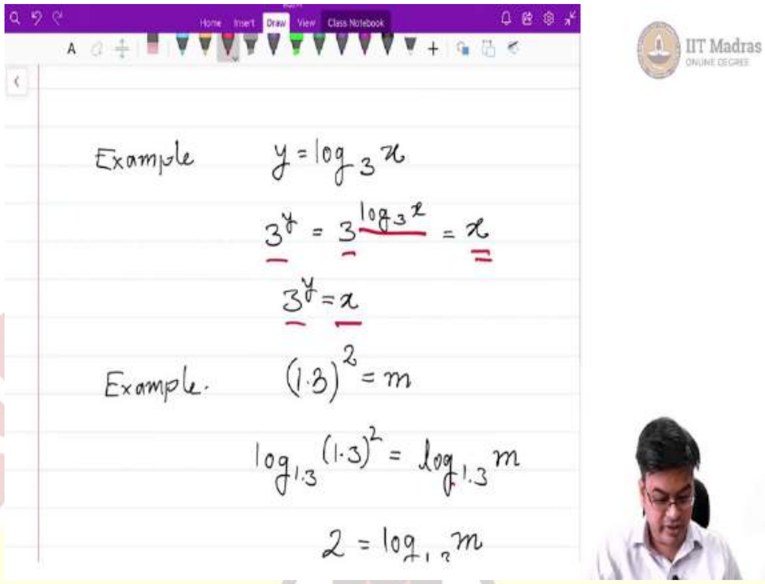
So, naturally let us split the real line into some parts x is≠1 is already given to you, so x cannot take the value 1, this is a point 0, this is a point 1, let, for safety let us put the point -1 as well here. And now x cannot be equal to 1, so this point is actually deleted, so this point cannot be there. Then, 1 -x should, if 1 -x>0 that means my x<1 the function is defined.

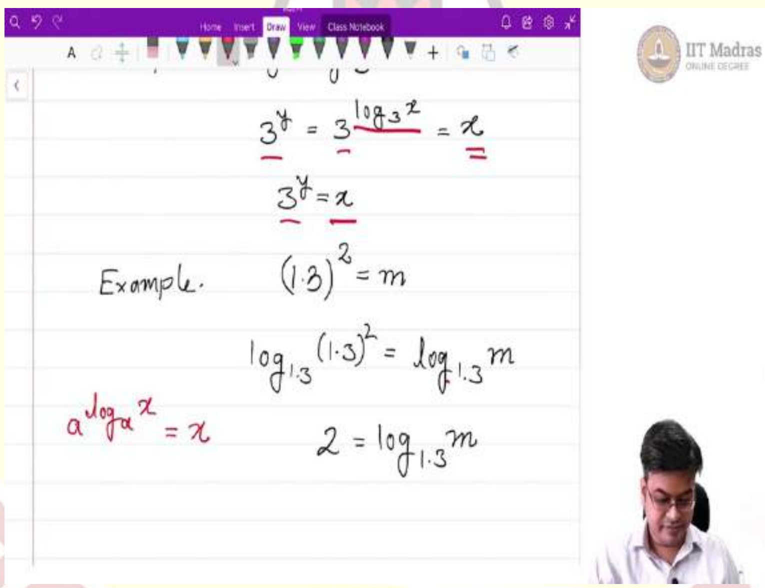
So, I have this in the similar manner -∞ to 1 but let us not go for -∞ because there is in the numerator there is 1+x, so this 1+x, it can become, it can take a negative value when x<-1 and if x<-1 this 1 -x will become positive. So, I have to rule out that part as well. So, this -1 to 1 is rule, -∞ to -1 is ruled out, -1 will give me the value 0 so -1 is also ruled out and therefore I am only left with the interval of this form which is -1 to 1.

So, based on the arguments and based on this domain I know that the domain of this function is valid only between -1 to 1. Now, you may say why not 0? 0 will not cause any problem because if you look at the function, if you substitute x=0 you will get log to the base 3 of 1 which is a positive number and therefore it is well defined. So, the domain of this function is nothing but -1 to 1, this is how we need before trying to solve any problem related to logarithms we need to first verify whether it is, what the problem that we are willing to solve is defined in a proper domain or not.

Most of the times when you try to formulate a problem the problem may not be defined in a proper domain and then solving that problem is a meaningless exercise. So, just to ensure that always your problem is defined in a valid domain. So, this ends the verification of this. Now, let us take one more example which will actually help you in understanding the reversibility of log and exponential function.

(Refer Slide Time: 15:06)



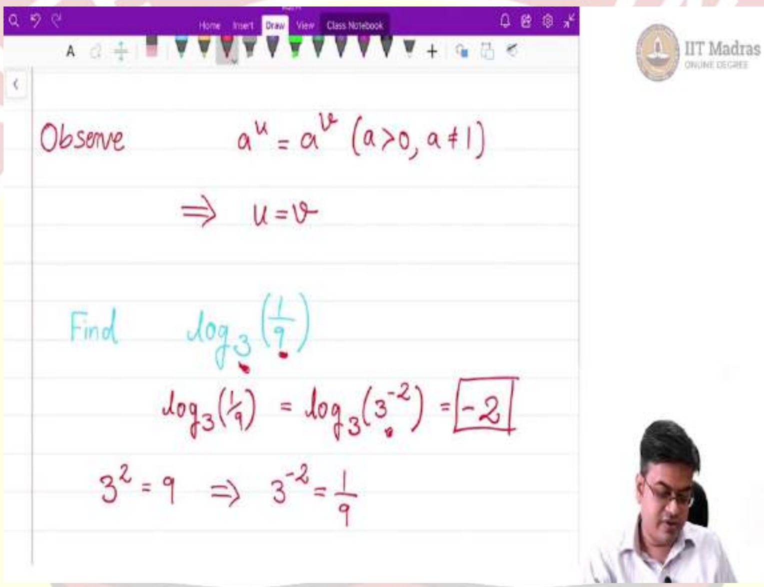


So, here is an example where we are actually demonstrating the reversibility of a log function or the inverse of a log function. So, y=log to the base 3 of x. We assume that everything is well defined and this x belongs to 0 to ∞. In that case this y will belong to the real line and if I want to write 3^ythen I will write y as $33^{\log_3 x}$.

By definition, by definition this function is the inverse of the log function. Therefore, you will get this to be equal to x and therefore your $3^y = x$. Now, how this helps in your calculations? Suppose, you know some number $1.3^2 = m$ and you want to identify this m. Then you can actually take the log of this function, log of this function which is the inverse of this and which will be equal to log to the base 1.3 of m and if you equate these two what you get here is 2 being equal to log to the base 1.3 of m.

Why is it so? Because 1.3 square we have taken the log so this is like a^x and you are simplifying it. So, a^x, $a^{\log_a x}$ is actually x. So, you will get the number 2 naturally. So, this is how the log thing helps.

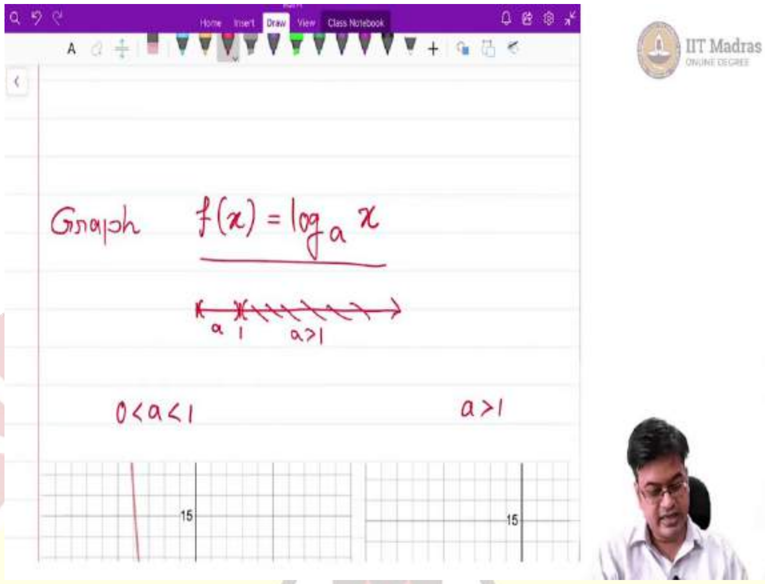
(Refer Slide Time: 17:15)



And here what the, the fact that we have used is $a^u = a^v \; for \; a > 0$ and $a \neq 1$ implies $u = v$. If you use this fact and you are asked to find the log to the base 3 of 1 by 9, then you can easily find. Let us see how. So, you start with log to the base 3 of 1 by 9. Now, you look at this 9 and 3. If you look at 3 square that will give you 9 isn't it and that also implies3^{-2} will give me $\frac{1}{9}$. So, I will simply use the fact that $\log_3 3^{-2} = \frac{1}{9}$.

So, but this is an inverse function, this is like 3 raise this particular thing is like 3^x, $\log_3 3^x$is again going to be x, so you will get -2 to be the answer, there this is how you can solve some problems very easily when you can identify the base is actually multiple of this particular argument. So, this is the use of log we will deal with it in more detail when we will solve the problems on logarithms. Now, for a moment we have identified what is the inverse function of our exponential function, it is logarithmic function to the same base as exponential function.

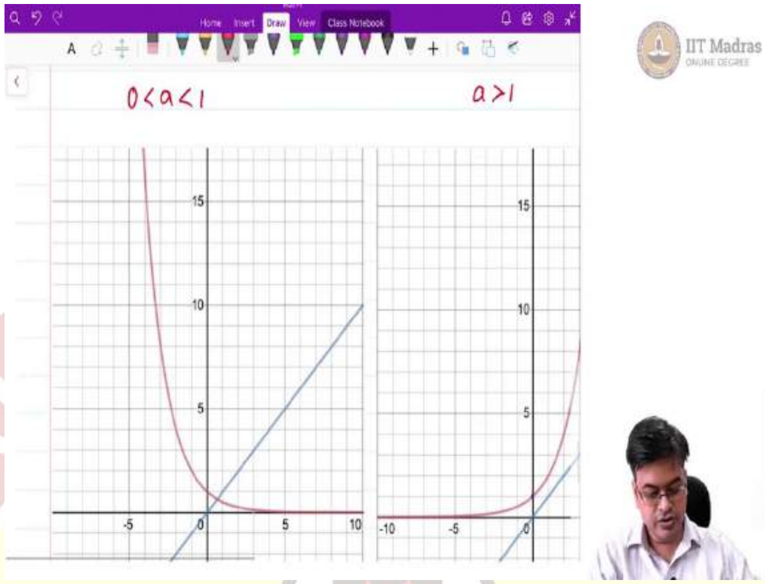
(Refer Slide Time: 19:23)

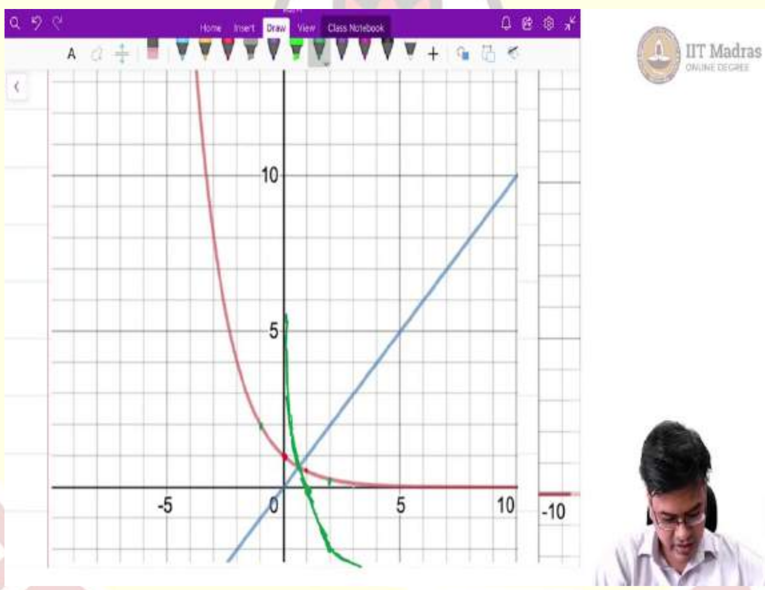


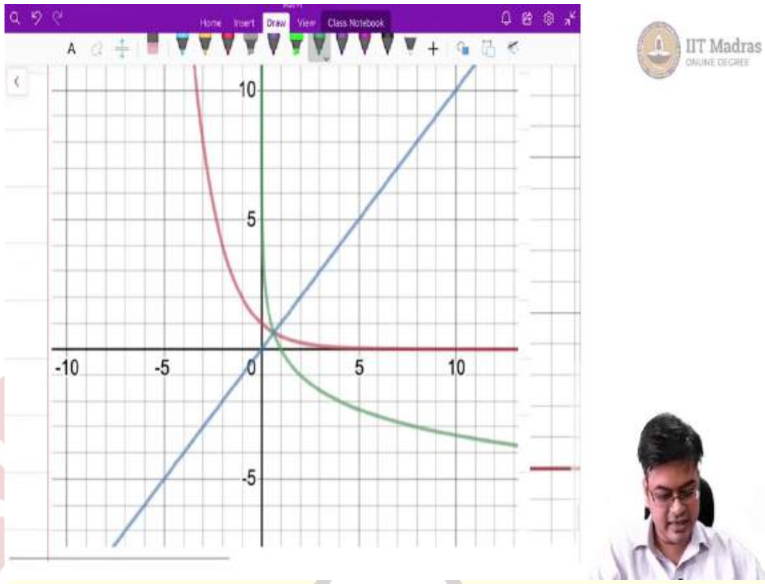
Let us try to look at the graph of the inverse function that is graph of f x equal to log to the base a of x. How will it look like? If you remember the graphs of exponential functions, the graphs of exponential functions were having two discriminations, like if you take a the line from 0 to ∞, then there was some split at 1 and from 0 to 1 when there is, the value of a lies in 0 to 1, the graph was different and from this side onwards that is a>1 the graph was different.

So, let us first imagine those graphs and let us recollect from the previous video what was the interpretation of the graph of the inverse function. If you recollect from the previous video, the graph of the inverse function is nothing but the reflection of the original function f along the line y=x or the mirror image of the function.

(Refer Slide Time: 20:25)







So, let us look at the exponential function first when 0<a<1 and a>1. So, this is the graph when 0<a<1. Now, I have made it big enough so that you can understand better and the blue line is the line y=x. Now, if I want to translate the mirror image of this function how will I translate?

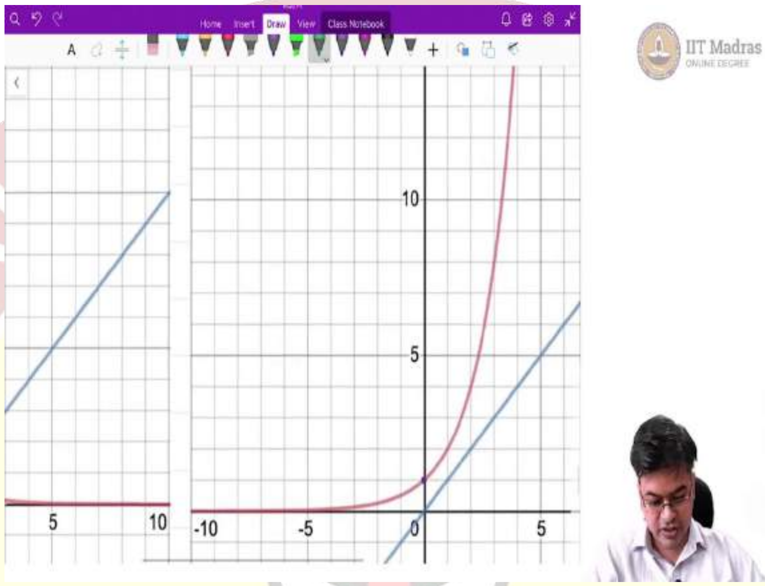
Let us take one point so let us take a point 0, 1 over here, the translation of that point will be 1, 0 over here and then take this point over here, I should not draw any point here because it may confuse you so the translation of that point in this zone is a point over here and a point over here, and similarly you go on translating and connect the two lines.

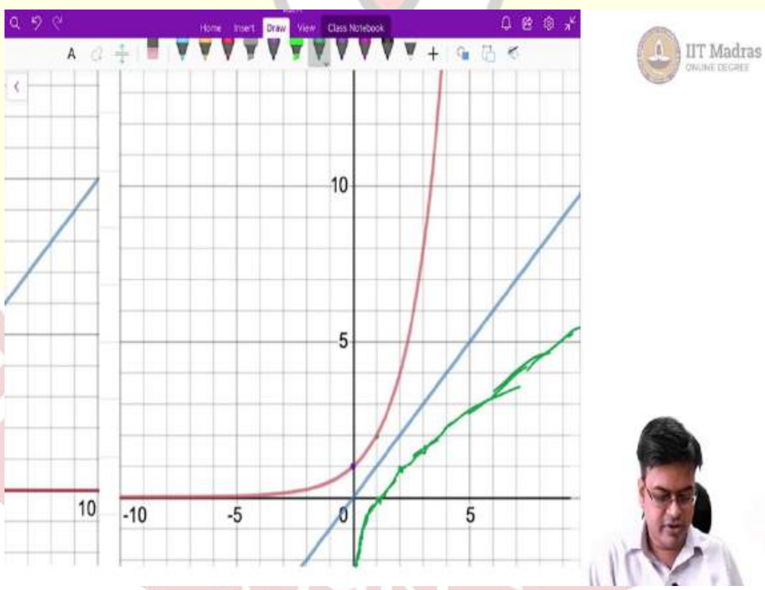
For example, here if I go on translating this point then the translation will actually go to some place over here and if you take one more point over here then the translation will actually go to the other quadrant which is 2 units below this and over here. So, the graph of this function will actually look something like this, it will pass through the same nodal point and it will pass through this and then on y axis it will be very flat, very close to the y axis and so on, so this is how the graph of the function will look like because it is a mirror image.

So, this is how it will look like, it is not an asymptote but because the graph paper is over I am not able to draw. In a similar manner this is the case when 0<a<1. So, I have drawn the graph in the next sheet which is a green line you can see this green line actually matches with this green line, I have slightly shifted the graph paper in order to have a better visibility.

Now, you can actually see this is the original function, this is the new inverse function and this is the line y=x, so you can see the correspondence of the inverse function with respect to the original function, all this is possible because our function is one-to-one.

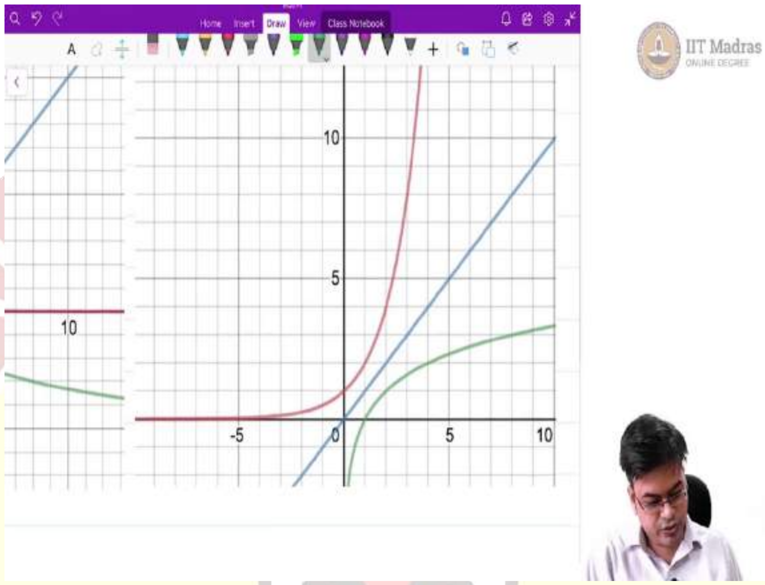
(Refer Slide Time: 23:07)

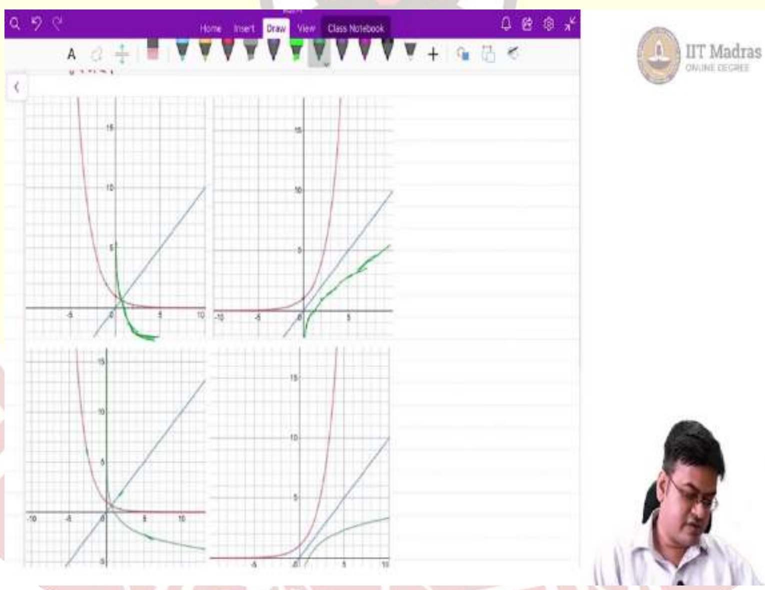




Now, if you look at, again look at the graph of a function where a>1 then this is the graph of a function here there are no overlaps, so it is relatively easy to draw the graph. For example, I can choose this point over here if I go one unit from here I should get something like this here so it is a reflection along x axis, so it will be relatively easy to draw the graph here, this point reflected here that point will be reflected here and then I can draw that, I can join the curve like this and it will be exact mirror image of the original function and it will be going close to this particular function.

(Refer Slide Time: 24:03)





So, roughly this will be the graph, I have drawn the full proof graph on the next graph paper which is here. So, now you can easily visualize the graphs of both the functions, let us zoom out and see all of them together all 4 graphs together. So, these are all 4 graphs handled together, so my graph actually looks like this graph for both the cases, so this is how it is easy to draw the graphs of inverse functions once we know the graph of the original function.

In the next, this is, that is all for this video. In the next video what we will see is we will try to use our knowledge of logarithmic functions and try to see how the formulation of a mathematical problems becomes easy when we consider logarithmic functions, even though there is a limitation that logarithmic function is defined only from 0 to ∞ not on the real line. Thank you.

IIT Madras

ONLINE DEGREE

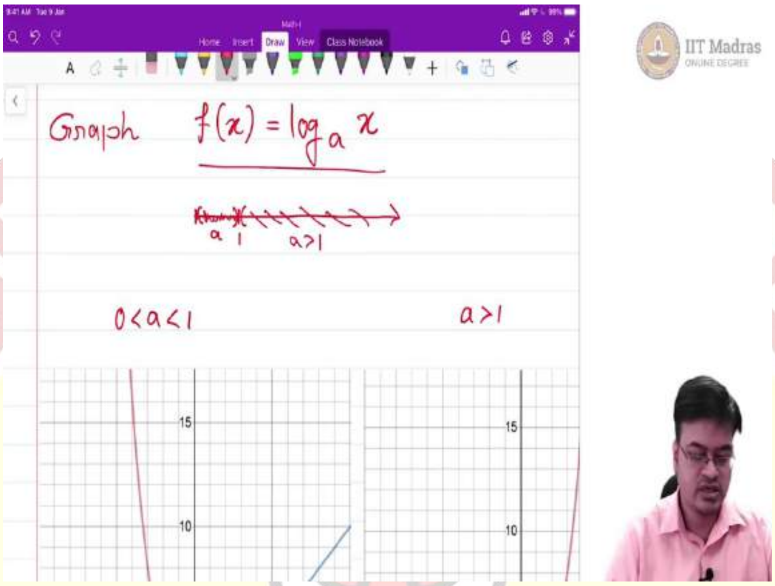
Mathematics for Data Science 1

Professor. Neelesh S Upadhye
Department of Management Studies
Indian Institute of Technology, Madras

Lecture No. 53

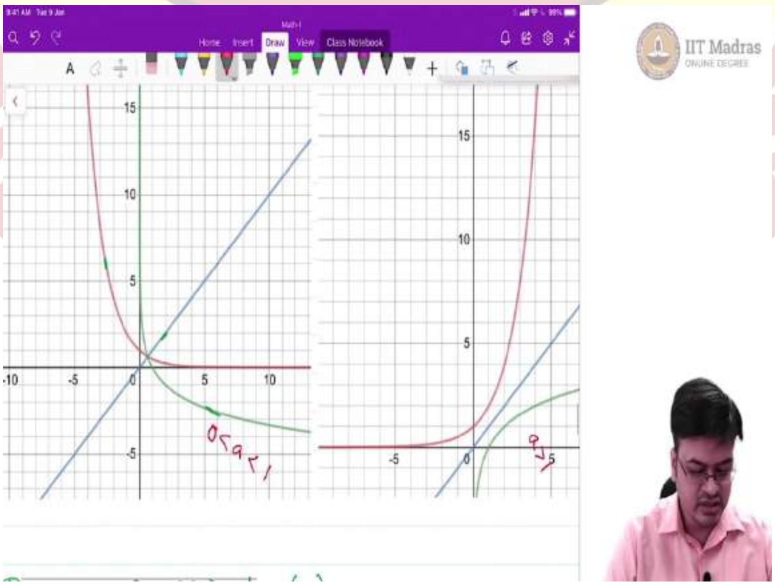
Logarithmic Functions: Graphs

(Refer Slide Time: 00:14)



Hello friends, in the last video, we have seen how to graph a function $f(x) = \log_a x$. In particular, we have seen two kinds of graphs or two kinds of divisions, when our a is between 0 and 1 the graph has one form and when a was > 1 the graph has the other form.

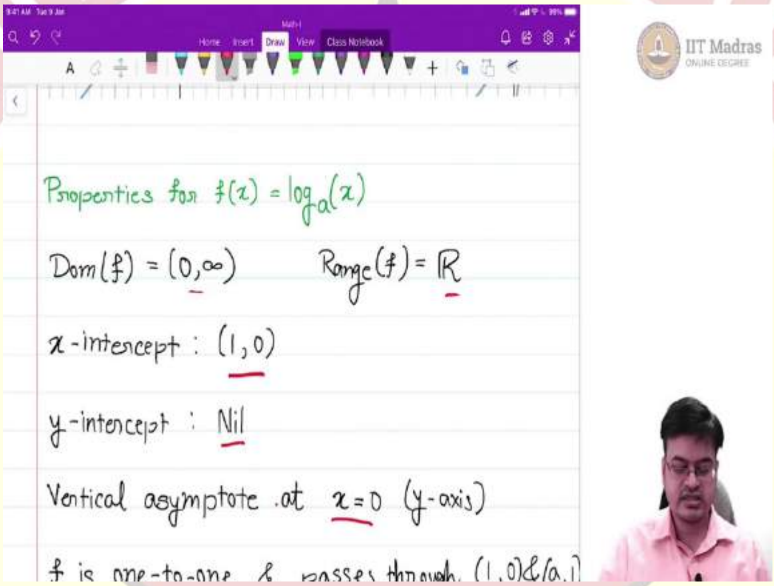
(Refer Slide Time: 0:44)



In particular, what we have seen is if the graph of exponential function for a < 1 there is the graph of a raise to x for a < 1 is given by a red line, then you can actually reflect this graph along the blue line which is y =x and get the corresponding graph for$\log_{10} x$ of x when 0 < a < 1.

In a similar manner, when a is > 1 the matters are, matter is very easy. And you see there is because there is no intersection, you can simply reflect the red line along the blue line to get the green line and the final graphs will look like this. So, this is the this is the case where my a is > 1 and this is the case where 0 < a is < 1. In particular, we have already seen some something similar in the, when we studied exponential functions. In particular, we will try to list all the properties of this graph of exponential function.

(Refer Slide Time: 01:52)



So, let us start our journey of listing the properties of graph of logarithmic function. So, first thing is the domain of the function we as it is an inverse function of exponential function is $(0, \infty)$ and the range of a function is real line, when you studied exponential function, the intercept was (0,1) here the x intercept is (1, 0) and there was no x intercept here, there is no y intercept, because simply because it is reflection along y =x line.

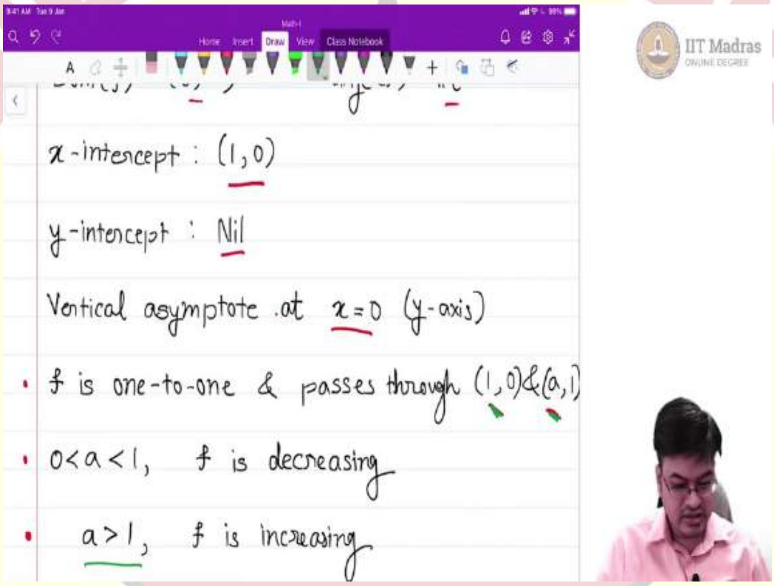
Then you had (vertic) in when you studied exponential function, you had a horizontal asymptote that is x axis was your asymptote in this case, you will have a vertical asymptote and that is x =0 is the line it is easily visible over here. For example, if you look at this green line, the vertical asymptote is towards the positive side of ∞ that is positive side of y axis. And if you look at this particular picture, it is towards negative side of y axis.

So, these are the typical features that you can you will understand when you look at the graph of a logarithmic function, then naturally this is the inverse function of a one to one function.

So, it is one to one and it passes through two points. If you recollect, the exponential function was passing through 0.01 and 1 a. So, naturally this function will pass through points 1 0 and a 1 all the time. So, these are the two static points whenever you consider a graph of a logarithmic function.

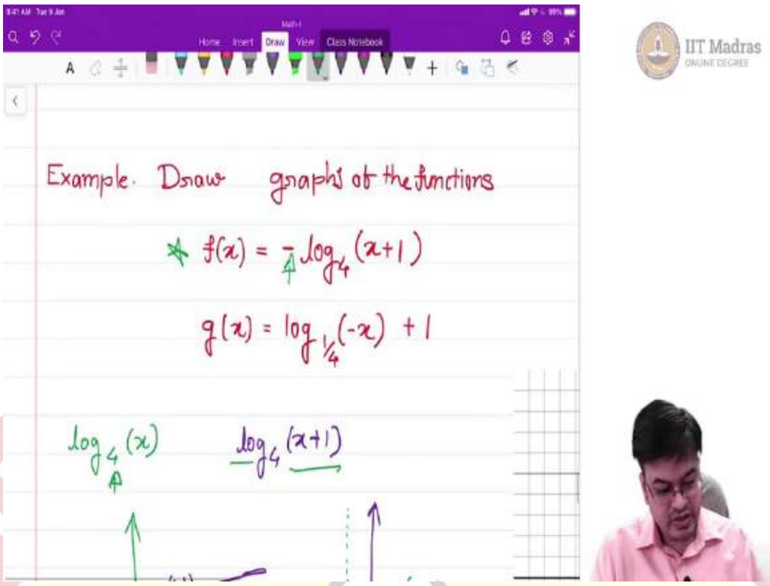
As it is visible from the graph for $0 < a < 1$ this green curve is actually a decreasing function. And for $a > 1$, this green curve is actually an increasing function. So, that those properties naturally boiled down to for $0 < a < 1$ the function is decreasing and for $a > 1$ the function is increasing.

(Refer Slide Time: 4:07)



So, these are important properties of graph of logarithmic function. So, while drawing the graphs of logarithmic function in a standard form, you should always remember whether you are satisfying these properties or not, that is a cross check whether your answer is correct or not.

(Refer Slide Time: 04:29)

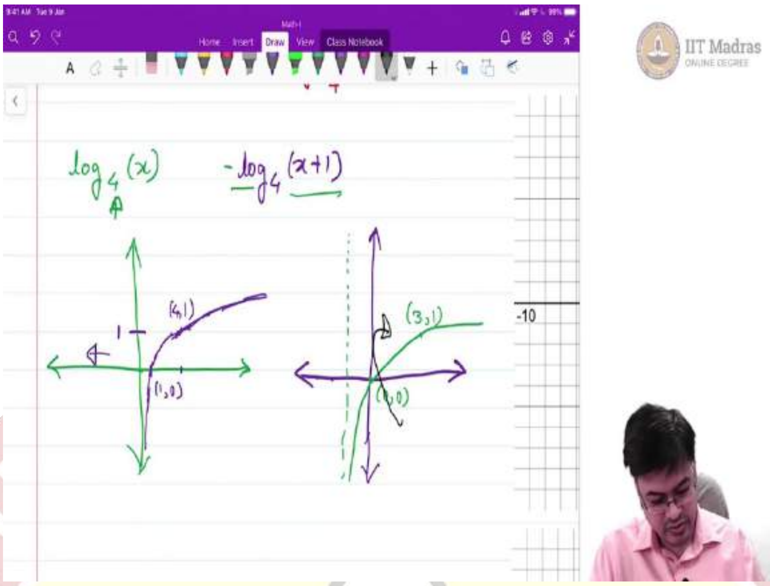


So, let us enhance our knowledge by taking an example of drawing a graphs drawing a graph is not correct, or drawing graphs of the functions that drawing the graphs of the functions,$f(x) = -\log_4(x+1)$ and $g(x)$ =$\log_{\frac{1}{4}}(-x+1)$.

Now, you remember the domains the domains of the function. So, here if I want to draw a graph, let us take the function $f(x)$ here If I want to take a graph of $f(x)$, so, first let us understand how the graph of $\log_4 x$will look like in order to understand this let us go to the properties is a the base > 1 or < 1 this is the first concern, so my base is > 1.

So, the function should be increasing okay naturally the function is one to one and my curve should pass through (1,0) and (4,1) correct and I should have vertical asymptote at x =0 naturally the function is actually increasing that means, I will come from down to up and therefore, and obviously x intercept is (1,0) domain of x is 0 to ∞ range of f is ℝ.

(Refer Slide Time: 06:02)



So, keeping all these things in mind, let me draw a quick sketch of this graph $\log_4(x)$. So, let me change the colour and it will be like this this is an asymptote, then it will pass through this point and you know what this point is and then it will be like this, this is the basic understanding it should pass through point (1,0) this is the point (1, 0) and (a, 1).

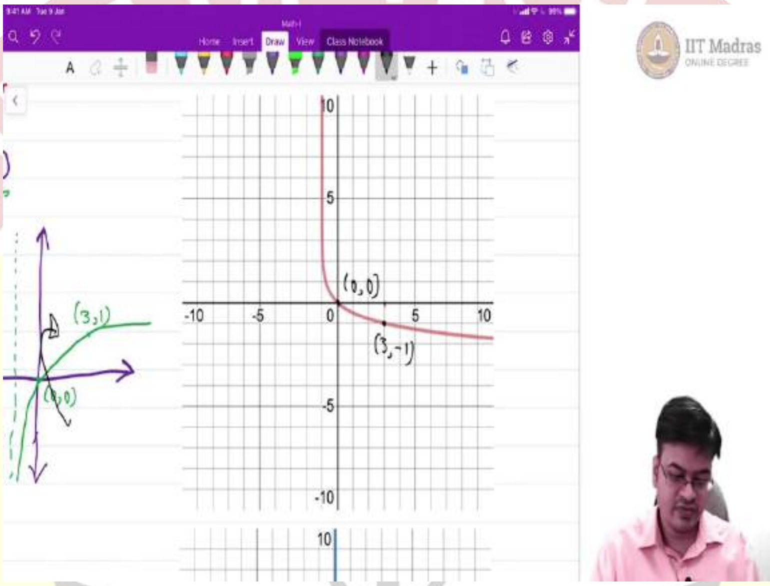
So, 4, 1 should be the point that it should pass through. So, naturally let us see this 1, 4 on x axis and this is 1 on y axis. So, as it passes through (4,1). So, this is these are 4 units this is 1 unit. So, this is 1 unit and that is 4 units on x axis this is 1 unit. So, this is how the graph will look like, but you have not been given a graph of $\log_4 x$.

So, what is happening it is going to x+1. So, now, $\log_4(x+1)$, how will it look like? That is the next question. So, what we are doing is we are shifting it on x axis. So, whatever value x was taking now, x+1 is taking that is a shift along this direction and shift along this direction of 1 unit. That means, that simply means, again you can easily quickly draw the graph, rough sketch of the graph whatever was happening for this particular thing, everything will shift and let me write use a green colour.

Now, instead of having $y = x$, instead of having this, this particular y axis as an intercept, I will have a new horizontal vertical asymptote which will be at 1 unit apart and my curve will pass through this this will be my new asymptote and my curve will pass through this and it will behave like this simple. So, instead of 1, 0, everything is translated by 1 unit. So, I will have 0, 0 the values on y axis will not change values on x axis will change. So, everything is translated by your 1 unit. So, I will have a point 0, 0 and this point where it intersects it will be 3,1. So, now this is my new graph of locked to the base 4 of x+1.

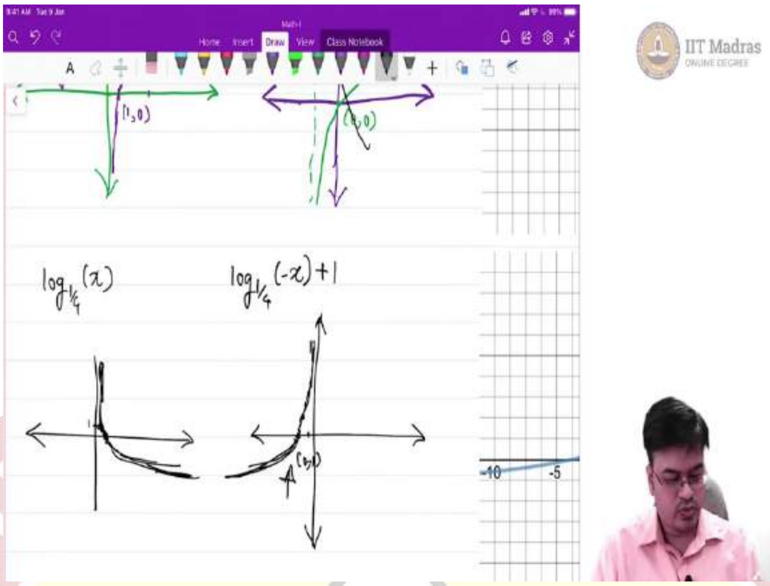
Now, the twist that is added is the-sign. So, that means it this is the graph of y =$\log_4(x+1)$. Now, if I add a-sign to this, the y will become the -y. So, now reflection along y axis so, the well what I have d1 is reflection along y axis. What I actually meant was reflection along x axis. So, the graph of this function should reflect along the x axis when I substitute $y = -y$.

(Refer Slide Time: 9:55)



Therefore, the graph will simply twist itself along the file along the line y along x axis and therefore, the upside will go down and the downside will go up and therefore, the function will look like this, that is all. So, as you can easily see, the function will look like this the point 0, 0 will remain intact and then point 3,1. So, 3 on x will become (3,-1), rest of the things will remain intact. And naturally this is how the graph will look like that is all.

(Refer Slide Time: 10:45)

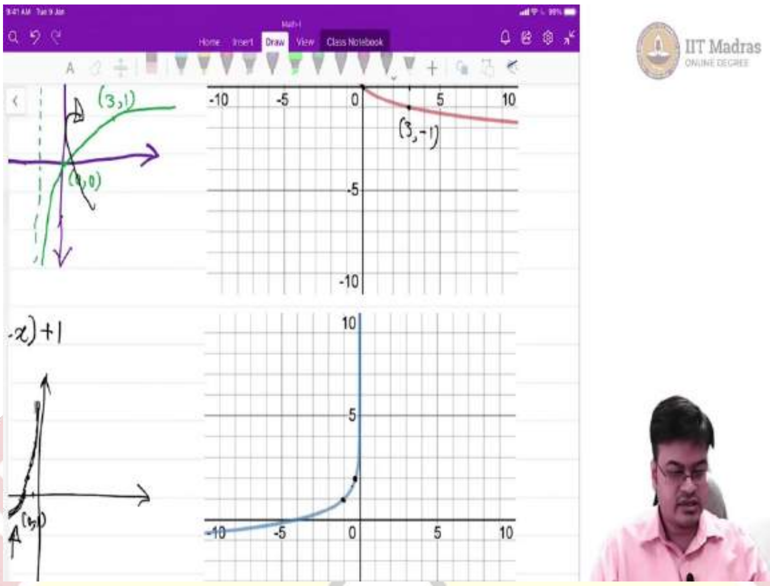


So, let us go to the next problem that is $g(x) = \log_{\frac{1}{4}}(-x+1)$ so, now we will first look at, can I draw a graph of $\log_{\frac{1}{4}}(x)$. The answer is yes I can draw and it will be simply this kind of graph. So, let me draw it a quick snapshot of that graph $\log_4(\frac{1}{x})$ by quick snapshot will give me something like this. So, this is the point and this will increase keep on increasing and it should pass through 1, 0 and a, 1. So, that point will be here somewhere. So, 1, 0 and a 1 so, $\frac{1}{4}$, 1 and 1, 0 correct. So, somewhere here it will be $\frac{1}{4}$ and 1 so this is a point 1 here, fine.

Now, the twist is this particular function is twisted with $\log_{\frac{1}{4}}(-x)$. So, wherever x is there, you are replacing the values with-x that simply changes the paradigm that is, you are taking a reflection along the y axis. So, now essentially when I substitute$-x$, this will be an asymptote this will still remain an asymptote and the graph will switch like this it is an exact mirror image of this along y axis.

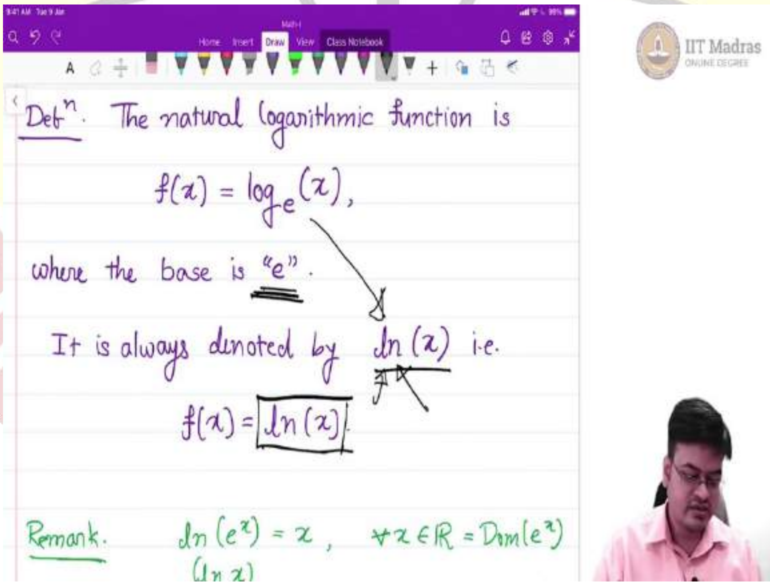
So, naturally the point 0, 1 will become 0-1, 0-1 and that $\frac{1}{4}$, 11 will become $(\frac{-1}{4}, 1)$. So, that also will be there. So, this is how the graph will look like now, you are adding a twist to the problem by adding+1. So, what will this thing this operation do. So, earlier $y = \log_{\frac{1}{4}}(-x)$. was there. Now, adding+1 will simply shift these values along y axis in upward direction 1 level up.

(Refer Slide Time: 13:10)



So, naturally the graph will look something like this. So, the point 1,0 will now be shifted to 1,1 that is this point and $\frac{1}{4}$,1 will be shifted to $\frac{1}{4}$,2 that is this point. So, these are the two points. So, we are able to map those two points and therefore, the verification of graph is complete and this is the current graph.

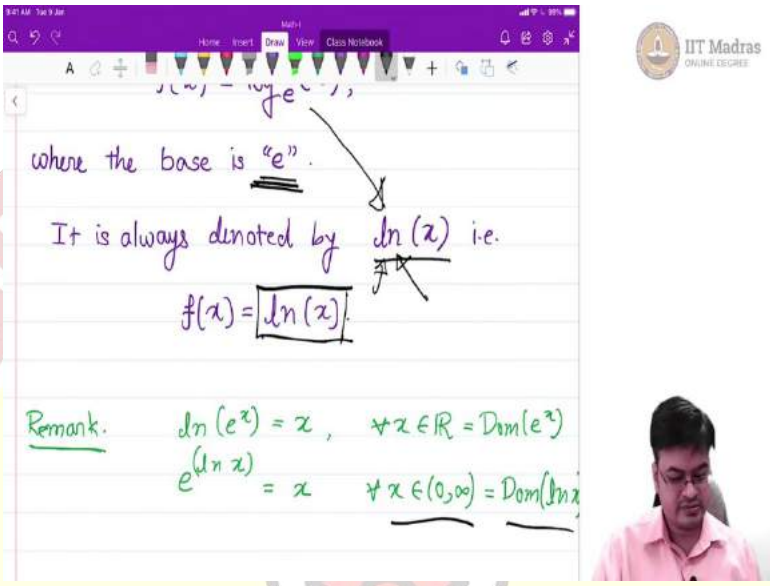
(Refer Slide Time: 13:52)



So, now in the next thing that we want to introduce to you is like logarithmic function to the general base a we have some special logarithmic function that is called natural logarithmic function which is which involves the notion of Euler's constant that is e and this is very special as the natural exponential function is special.

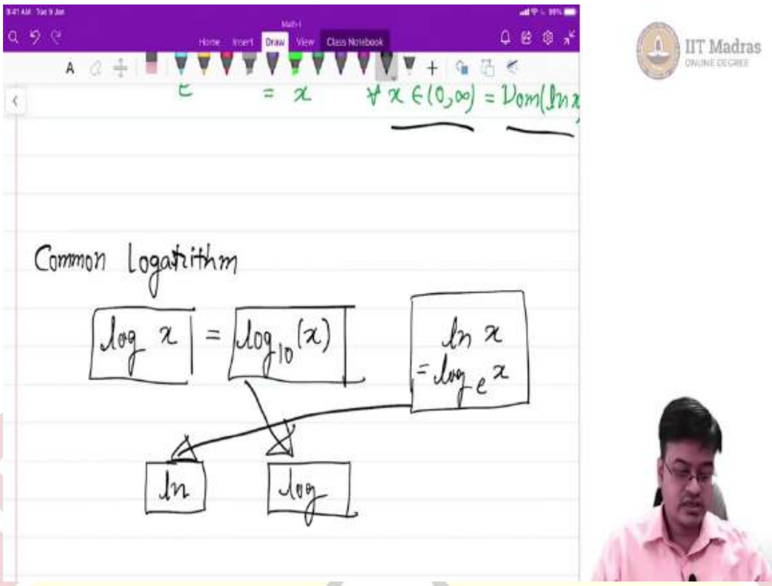
So, this function is defined in a separate way as the natural logarithmic function and it is defined as $\log_e x$, where the base is e we have already seen the importance of e in past few videos. But, to add to the speciality, we have some special notation for this $\log_e x$ it is always denoted by $\ln x$, where l stands for logarithm and n stands for natural base.

(Refer Slide Time: 14:50)



And whenever we write this as ln of x that simply means, I am talking about the natural logarithm of x that is $\log_e x$. So, hence forth whenever we discuss about natural logarithmic function, we have to use this notation ln of x it is quite standard simple verification, you can actually check ln of e raise to x is x for x belonging to $\mathbb{R}$ which is the domain of e^x and $e^{\ln x}$ is x for x belonging to positive real line, which is the domain of $\ln x$.

(Refer Slide Time: 15:39)



In a similar manner, there is some other log there are some other logarithms, what is called this called natural log something else is called common logarithm, which has to do with our decimal representation and common logarithm is actually denoted as log without any base. So, that means, it is $\log_{10} x = \log x$.

So, in general, the common in common terminology, you may consider when there is no mention of a log, you may consider this log is to the base 10 and if something like $\ln x$ is written, that is $\log_e x$. So, this is what you these are the important things you need to remember. In olden days when we used to use calculators, there were two separate keys associated with this 1 key was ln and another key was log.

So, in these keys, they were commonly referring to $\log_{10} x$if I am talking about log and log to the natural base if I am talking about ln, so just remember these are two commonly used logarithms, which we will use quite often in our daily practice. Calculators distinguish them with ln and log and we prefer that you also distinguish them with ln and log. This finishes the topic of natural logarithm, natural logarithm nothing is special. It is just a way of taking the logs with e as a base.

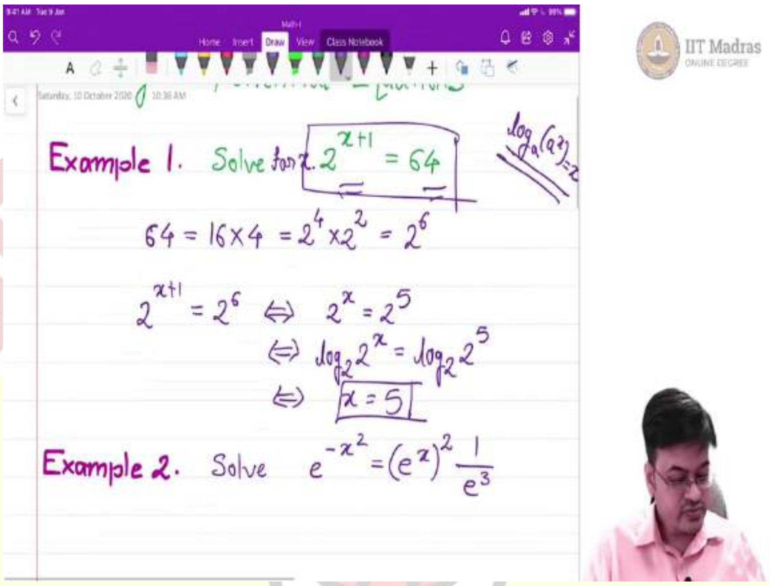
INDIAN INSTITUTE OF TECHNOLOGY MADRAS · सिद्धिर्भवति कर्मजा ·

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Management Studies
Indian Institute of Technology, Madras
Lecture No. 54
Solving Exponential Equations

(Refer Slide Time: 00:18)



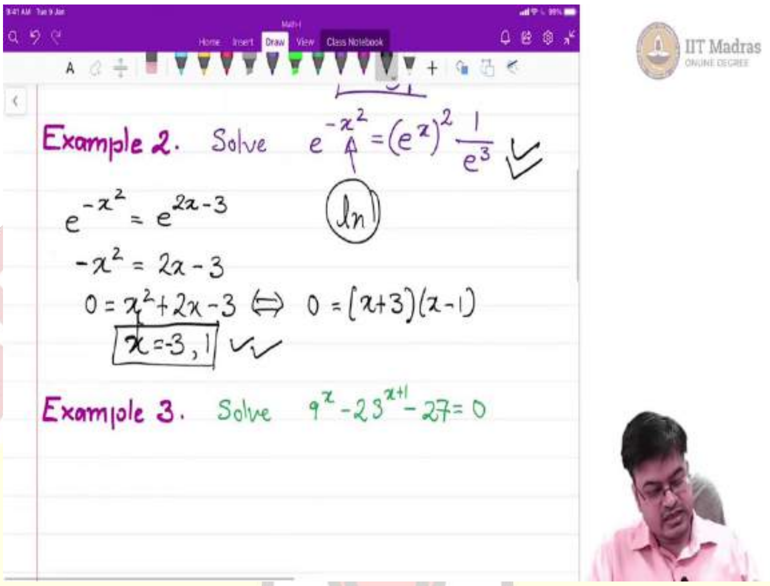
So, let us try to go back and solve our exponential equations. So, this is where the logarithms will come handy. In particular, we have seen one property of logarithm, for example, if I am talking about a^x., and if I talk about $\log_a a^x$. I get x, this property we have to emphasise and track everything in terms of this property and solve the exponential equations while solving the using logarithms.

So, let us try to get hands on these exponential equations using the logarithm. So, first you have to solve this equation naturally we have to solve for x. So, 2 raise, the equation is $2^{x+1} = 64$.. Now, as you know earlier that this is not more of logarithm, but more of inspection. So, 64 seems to be a nice number. So, you can write 64 as 16×4 is 64 but 16 is nothing but $2^4 \times 2^2$. So, what you got here is 2^6. So, my 64 can be written as 2^6.

Now, I have been asked to solve for this equation that is $2^{x+1} = 64$. So, essentially $2^{x+1} = 2^6$ if and only if, so you can take this 1 2 out $2^x = 2^5$, hit the function with a logarithm and use the property. So, use the property $\log_a x = x$ hit the function with logarithm and you know logarithm is one to one function.

So, nothing changes what should be my a, it should be 2, then I naturally I will get $\log_2 2^x = \log_2 2^5$. Using this property, you can actually write this as x =5 there is my answer, I have solved this expression. So, my answer to this question is x =5.

(Refer Slide Time: 03:08)



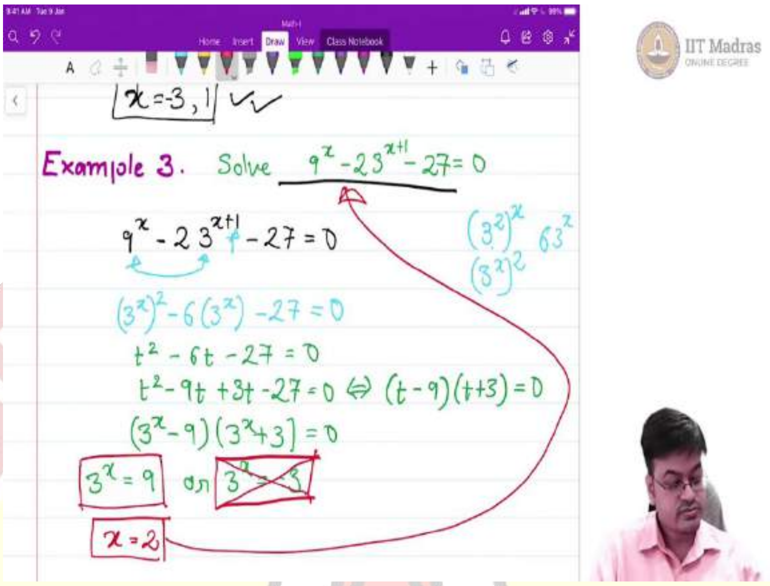
Let us look at the second example. This is also an exercise in computation where I will try to match these 2 but it becomes slightly complicated because as you can see, it involves the term containing x^2. So, again the question is to solve for x. So, let us solve for x. So, let me first simplify the right-hand side, $e^{(x^2)} = (e^x)^2$ using law of indices will become $2x$ and this 1 by e^3 will become-3.

Again, you use the natural log take ln on both sides. If you take ln on both sides using the same property that I mentioned earlier, I will get $x^2 = 2x - 3$. That simply means, I have a quadratic equation which says x^2+2x-3 =0. Use our knowledge of quadratic functions or quadratic equations. And in this case, the quadratic equation I think can be solved with by a very easy solution that is x+3x+3×x-1.

This is simply by factoring you will just look at the lectures on quadratic functions and see how to solve the equation by factoring. Actually, I will split this× $x^2 + 3x - x - 3$ that will give me-1 as common and the first factor is x+3. So, this is how I will solve done. So, now, what I have as an answer is $x \ = -3 \ and \ x \ = 1$.

Now, remember, we are solving some equations, where the domains and co domains may not be defined properly, then we may land up in infeasible solution. So, always check whether these two functions will work here or not, let us substitute this $x = -3$ over here and you can verify that it perfectly works and x =1 also perfectly works. Therefore, these both are actually the solutions to the problem.

(Refer Slide Time: 05:50)



Let us complicate the matters a bit more by putting in this kind of equation. So, let me write this equation and now, obviously, the question is to solve for x $9^{x-2} \times 3^{x+1} - 27 = 0$. So, here if you look at the equations closely before actually handling them, you can see that there is some common feature between this exponent and this exponent. What is that I have something like $3^2 = 9^x$, correct and $3^{(x+1)}$, I can take the+1 out, so that these 2 will become 6. So, 6 times 6×3^x.

Now, this thing you can rewrite as$3^{x\ 2}$. the whole square using this trick, it is very easy now to see that this particular thing is $3^{x^2} - 6 \times 3^x \quad - 27 = 0$. Now, there are 2 ways this equation is actually very similar to a quadratic equation or you can reword it as it is a quadratic inform equation. Now, using this equation, you can actually solve for 3^x not for x.

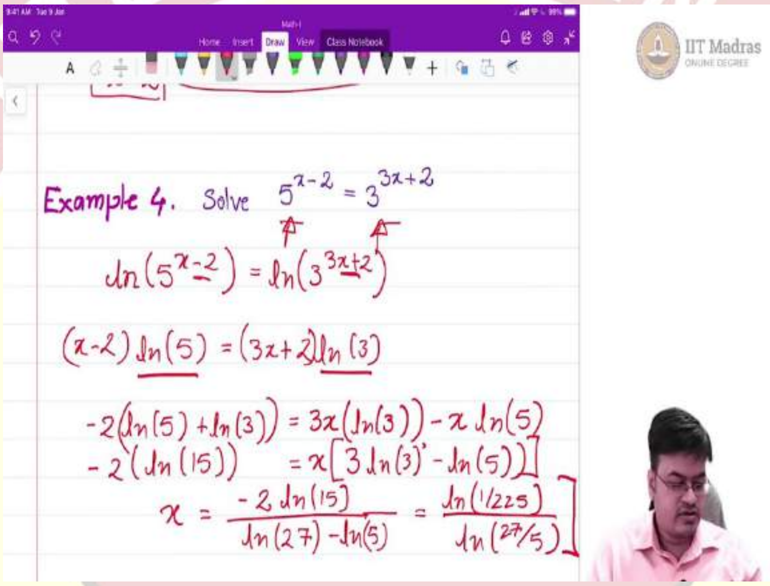
So, let us go ahead and put 3^x. as t and is $t^2 - 6t - 27 = 0$ again resort to a method which is like factoring. So, 27 can be factored×9×3, 9 3s are 27 with the-sign, so, it will be $t^2 - 9t + 3t - 27$ which is going to be equal to 0 this will give me this is if and only if t-9×t+3 equal to 0 correct. Now, the question comes t+9 =0. So, I have solved it for t what is t? t is 3^x. So, I have to resubstitute that and if I substitute that, then I get $3^{x-9} \times 3^{x+3} = 0$..

Now, you have to be a little bit careful. So, this simply means using the factor logic $3^x = 9\ or - 3$.. Now, this should give you an alarm in your head that this particular thing is not possible for any real x. So, this option is infeasible. So, I cannot solve for this, what about this?

Do I know something about this? Again, simple thing is you hit with a log you will get the answer or in this case, it is more obvious that x =2 should be the answer.

So, now once you got this you substitute this x =2 in the original equation and verify that it satisfies the equation that is that will be a good cross check. For example, $9^2$is 81, 3^2 is $3^3 =$ 27. This is 27, 27×2 is 54, 54+27 which should give you 81? Yes. So, it is a verification and we have solved this problem successfully, but remember here the occurrence of infeasible solution.

(Refer Slide Time: 10:17)



Let us take it a step further and see whether I can do something about an equation of this form, which is 5^{x-2}, 3^{x+2} unlike previous problems, the bases are not the same. So, what should we do about it is the question. So, in this case also we can actually rely on logarithm and we can blindly hit with a logarithm, but in this case, let us hit both sides of the equation with natural logarithm or you can take common logarithm it does not matter 5^{x-2} =ln 3^{3x+2}.

Now, if you have noticed that when you hit with the log of the same base the other thing gets vanished, but if you hit with a log of some other base that number will remain unperturbed, but this x-2 and 3x+2 will come in front that is x-2 ln 5 =3x+2 ln 3. Now, this ln of 5 and ln of 3 are nothing but merely some numbers which are getting multiplied with x and 2. So, they are just constants.

So, now my equation is actually a linear equation let us try to simplify this. So, here there are there is 3x here there is x so, we what we will do is we will take both parties on one side that is the parties corresponding to x on one side and parties corresponding to constant on one side.

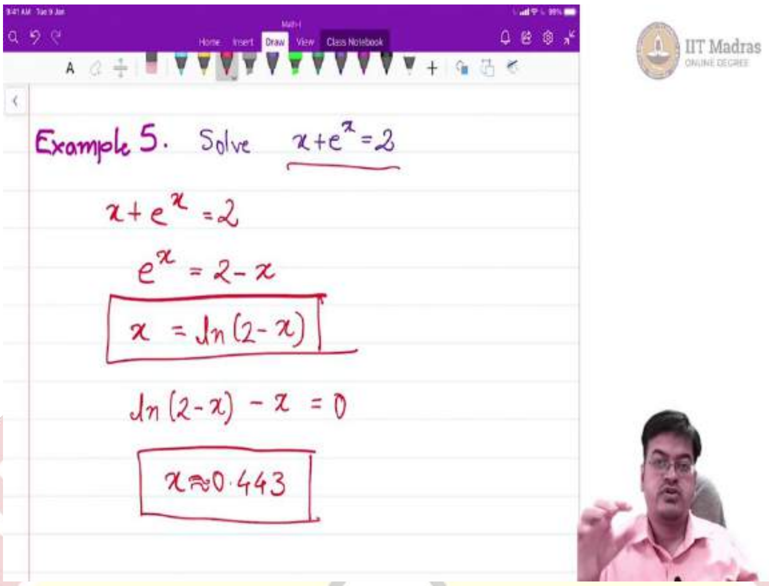
So,-2 times ln of 5 will remain as it is+2 when it comes here becomes-2 and ln of 3. So,+ln of 3 which will be equal to 3x was already there multiplied with ln of 3 and from here comes x which is-x times ln of 5.

So, now what I got here is 3x-x times ln of 5, I want x to come out common So, I can process it further which will give me x as a common factor 3 times ln of 3-ln of 5 and here it is-2 times now, ln of 5+ln of 3 you will learn ahead it will be ln of 15. And if you look at this particular expression this equation will not, this expression in the square bracket will not be equal to 0. So, I can take this expression here another thing that you may notice is 3 here, can be raised to the power of 3 over there and you can further simplify.

So, you will get in particular $x = -2\ln\frac{15}{3\ln 3} = \ln 27 - \ln 5$ which can further be simplified to, if you want you will have more precise structure later 15^2which will be 225, $\frac{1}{2}$ and because of this-sign it will be $\frac{\ln\frac{1}{225}}{\ln\frac{27}{5}}$.

So, this is again a number and this resolves the problem for x. Whether these numbers are feasible? Yes, they are very much feasible and therefore there is no visibility violation over here. This we will learn in a bit later how to do all these calculations, but as of now, this is the answer.

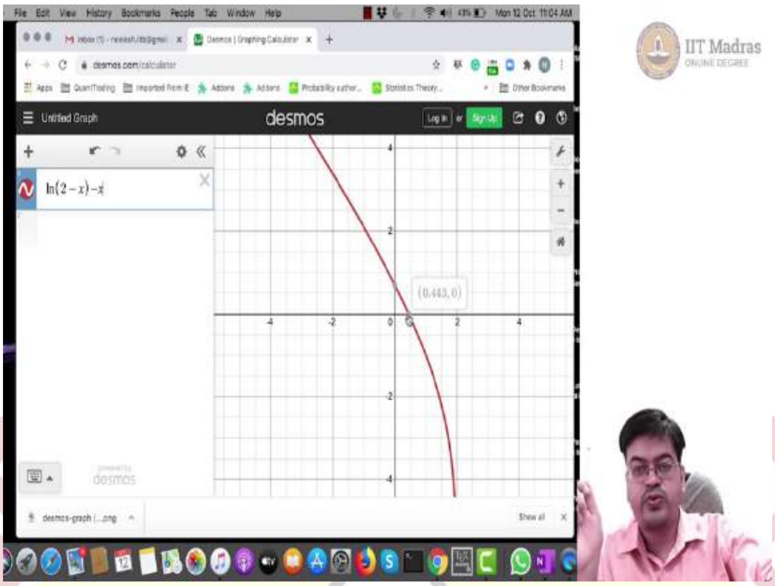
(Refer Slide Time: 15:12)



Now, let us look at this particular example, where you cannot solve on your own, but graphical solution may yield a better answer. So, for that let us look at x+$e^x = 2$. Now, if I try to hit this with a log, but before hitting this with log because I am encountering a + sign, let me put rewrite this equation as e raised to x equal to 2-x.

Now, hit this expression with a log, so you will get x =ln of 2-x and you are struck you cannot go anywhere. Now, if I want to solve this equation again the only thing that right now we are aware of is I will hit it with exponent and I will again get back the same equation. So, it is of no use. So, let us focus on this equation and let us let us try to graph this equation. So, in particular I have a function which is $ln\ 2 - x - x = 0$. Now, let us try to use some graphical tool like Desmos to graph these equations.

(Refer Slide Time: 16:50)



So, let us go ahead and use Desmos for graphing the equation and naturally the choice of log is only up to B. So, I will use ln or I will use whatever so, ln of 2-x you can see the graph-x. Now, the point where it cuts 0 is the solution to the problem and you can actually check using Desmos that the point is actually 0.443.

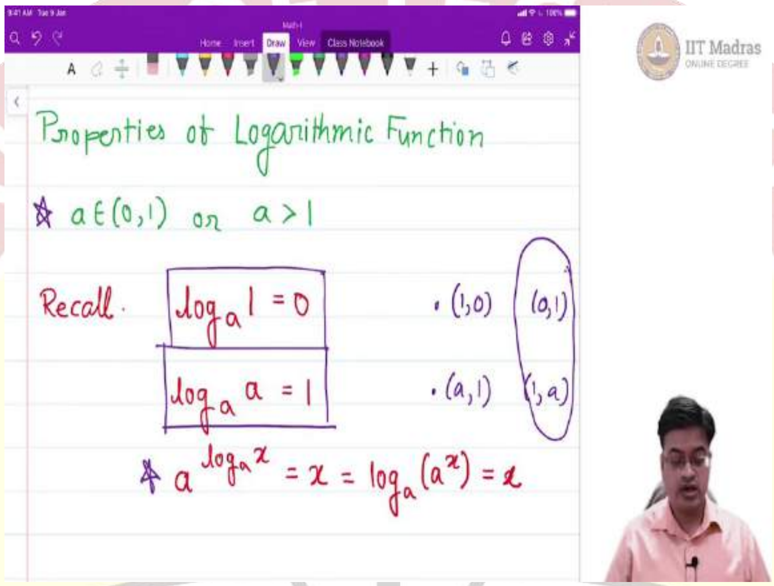
So, let us go back and draw and tell everybody that x =0.443 is the approximate solution to this problem, but in such problem this is the best that you can do. So, today we have seen how to solve exponential problems using graph or using algebraic methods. Thank you.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture No. 55
Logarithmic Functions: Properties - 1

(Refer Slide Time: 00:14)

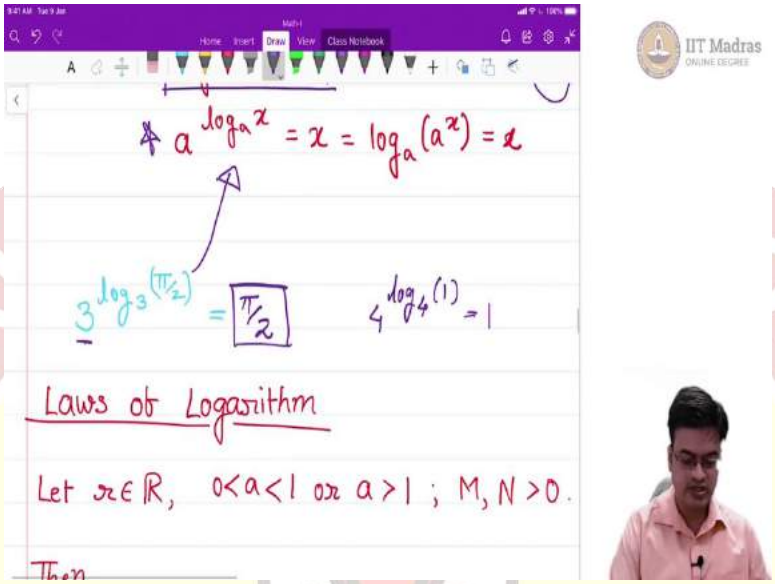


So, in this video we are going to look at further properties of logarithmic functions. So, when we look at the logarithmic functions in general we have a standard set of conditions that are imposed that is $a\epsilon(0,1)$ and a > 0. And we already know few properties of logarithmic function namely logarithmic function is actually an inverse of exponential function which is conveyed through this property that is if you take $a^{\log_a x}$, then you will get x back and if you use $a^x$and take log $a^{\log_a x}$ you will get x back.

So, essentially our logarithmic function is an inverse of exponential function. Another thing that you need to recall based on the graphical representation of logarithms is log to the base a of 1 is 0, that means we have already located that point 1, 0 is on the graph of the logarithmic function independent of whether a < 1 or a is bigger than 1. Another thing that you need to remember or recollect is log to the base a of a is equal to 1. What does this mean?

The point a1 is on the graph, which you have also seen, so when you are talking about logarithmic function, these two points are on the graph, when you are talking about exponential function because it is a reflection along y is equal to x axis 0, 1 and 1a are the points that you will look for exponential functions. So, this we have done enough. So, these are the basic logarithmic properties that we are already aware of.

(Refer Slide Time: 02:17)

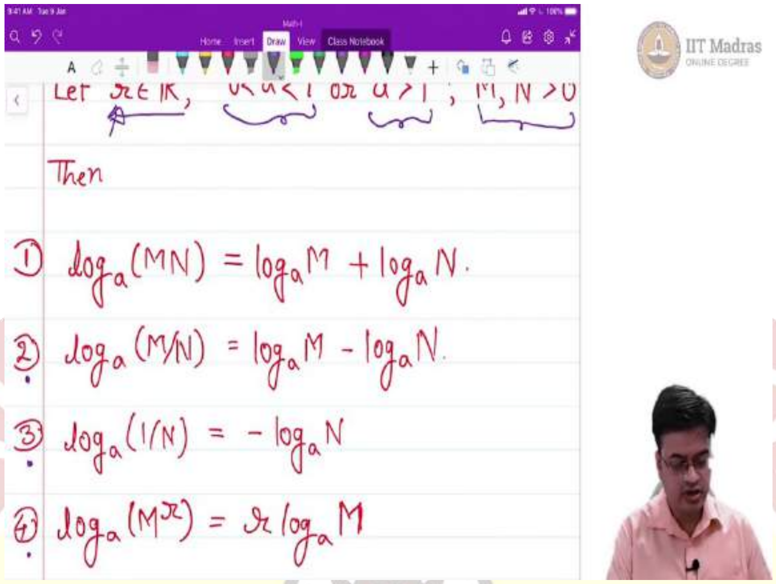


Now, let us go further and explore something which is called laws of logarithm. Prior to that now because we have mentioned log to the base a of a^x is x, you can also ask a question that what if I have been given a function like this, which is $3^{\log_3 \frac{\pi}{2}}$, so based on this you can use this particular formulation which is where a is 3 log to the base 3 of pi by 2 naturally this should be equal to pi by 2. So, the sometimes some complicated numbers may simplify in this way, it is simple demonstration of use of these properties.

Another thing that you can also see is say let us say 4 raised to log to the base 4 of 1, now you already know log to the base 4 of 1 is 0 based on this property and therefore it is 4 raised to 0 which will naturally give you 1, so all these simple simple tricks you should solve, you should solve more and more problems and gain more confidence in while using the logarithms, because while solving the problems on logarithms and exponential functions applying a log function or applying an exponential function will play a crucial role while solving the problems.

So, let us focus on some simple laws of logarithm and in fact when I, why these are called laws? Because these are the principles for which the logarithms were invented by Napier. So, let us see what are the laws of logarithms.

(Refer Slide Time: 04:07)



So, in order to define laws of logarithms first we need to restrict to the valid zone, so we will restrict to the value zone in such a way that my a is between 0 and 1 open interval or my a > 1, then because the logarithmic function is always defined on the positive side that means the argument that is applied to logarithmic function is positive, so my M and N are actually the arguments for the logarithmic function, so they are always positive.

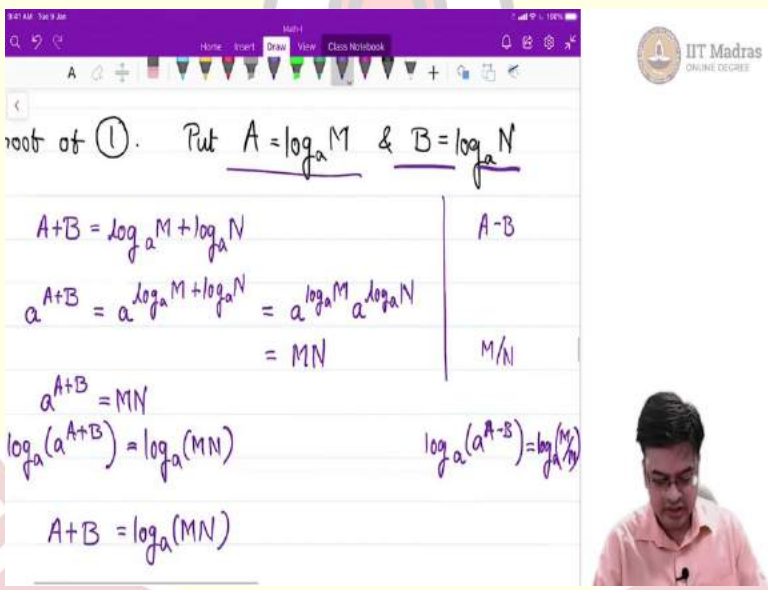
And here is one more thing that is some are real number r is given to you. If all these conditions are satisfied then there are 3 laws, we will see each of them and we will try to prove each of them. So, first law is, logarithm of a product of two positive numbers is sum of the logarithms, verbally you can state this as logarithm to the base a of product of two numbers positive numbers is nothing but the sum of the logarithms of the individual numbers.

In a similar manner if you go ahead and do some simplifications you can also come up with a second law that is, logarithm of a quotient of two positive numbers is nothing but difference between the logarithms of those two numbers. In a similar manner this is not a new law but we will state it for the sake of clarity that logarithm of reciprocal of a number is nothing but negative of a logarithm of the original number.

And the fourth one which uses this number r, which we have defined here and remember this is any real number, then the logarithm of M raised to r or any number raised to r any positive number is to r is nothing but r times logarithm of that number. So, when many astronomist where doing some calculations and they wanted to do the product of the two distances which are very high in the power of 10^{32}or something like that. And in that case the multiplication of two numbers becomes a tedious task, so in order to handle these tasks they have actually invented these logarithms.

So, if you search on the Google why the logarithm were invented you will come to know about many references from astronomy where they are successfully using logarithms. And remember they were doing this in around eighteenth century, so there was an absence of computational power. So, these laws were helpful and therefore they are governed as laws of logarithms. Now, let us try to prove each of them, one by one, let us take the first law which is logarithm of the product of two numbers positive numbers is equal to logarithm of is equal to some of the logarithm of these two numbers.

(Refer Slide Time: 07:32)



So, in order to prove this let us put A $= \log_a M$ and B is equal to log to the base B=$\log_a N$. And now what you do is you actually consider you actually consider A+B, so my A+B is nothing but log to the base a of M$\log_a M$+$\log_a N$, so what I will do now is I will simply go back and consider some properties of logarithms that I have already considered.

So, I will consider these kind of properties and let us see how this property can help me in proving this particular identity. So, I have used that particular property and let us say I have raised this as a^{A+B}. In this case, what I am using? I am using actually an exponential function is inverse of logarithmic function, so left hand side I have raised to the power a, so naturally right hand side will also be raised to the power a, so I will get this+$\log_a N$, no confusion here.

Now, you can actually see this particular thing when I am looking at this particular thing what I am getting over here is a raised to $a^{\log_a N}$. So, now what is this actually, if you look at our definition of a raised to log to the base a of M, you will get this to be equal to M and you will get this a raised to $a^{\log_a N} = N$, so you got MN.

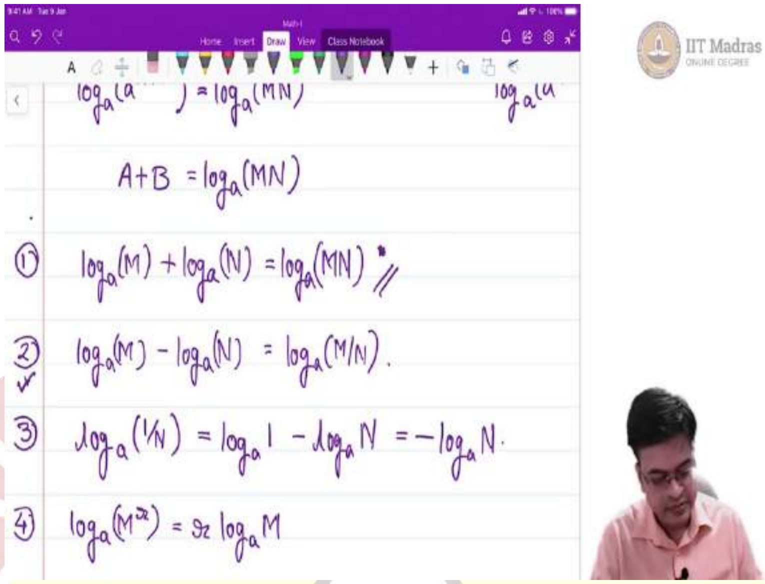
So, now what you got here is MN and $a^{A+B}$is MN, so what I have got here is $a^{A+B} = MN$. Now, if you look at what you want to prove, if you want to prove the left hand side is log to the base a of MN, so how will I get that? Again use the similar property which is given here and by using this property take logarithm on both sides. So, you take log to the base a of a raised to A+B which is equal to log to the base a of MN.

Now, my claim is we have proved this result, how? Because $\log_a a^B$ is nothing but A+B and I am saying A+B=$\log_a MN$, now what is a? Just substitute what we have put A as, log to the base a of M and B as $\log_a N$. Therefore, I can rewrite this as log to the base a of M+log to the base a of N is equal to log to the base a of MN. Clear.

So, my first result is proved first law is proof, now you can easily guess what modification do I need to make for proving the second law. So, if you look at the second law that is log to the base a of M upon N is log to the base a of M-log to the base a of N. So, if you want to prove this what modifications you need to do? You simply use this A and B, there is no change in this only thing that you will have is A-B over here.

If you have A-B over here all the things correspondingly all the things will change and instead of MN what you will get is $\frac{M}{N}$, rest of the things are just same you can practice it as an exercise and therefore you will get a similar result which says $\log_a M$log to the base a of a raise to M a raised to A-B is equal to $\log_a \frac{M}{N}$. And you are done, then you will again apply the you again use the inverse function property of logarithms and cancel with this off and you will get the second result.

(Refer Slide Time: 13:29)

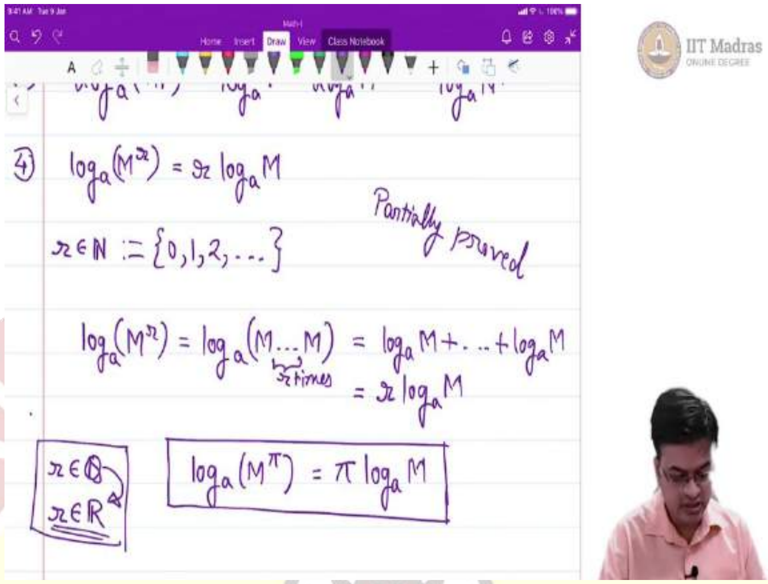


So, I have not doing it properly, but you can easily derive it from it. So, let us write the second result, this is the first result, first law of logarithm, second law of logarithm actually follows by replacing+sign with a-sign. So, $\log_a M - \log_a M = \log_a \frac{M}{N}$.

Now, you look at the third result which actually talks about log to the base a of 1 upon N. now, in this case you simply apply the second rule and you simply apply the second law logarithm that we have just proved and where the M is equal to 1. So, naturally I will get log to the base a of 1-log to the base a of N. Now, what is log to the base a of 1? We have already seen that 1, 0 is on the graph of a log, so based on this particular property we can easily conclude that this particular thing is going to be 0, so my final answer should be-log to the base a of N, that is all.

So, on let us look at what is the fourth result, $\log_a M^r = r \log_a M$. How will you prove this? Again you will apply the same modus operandi, that is you will first isolate some term and then you look at the term.

(Refer Slide Time: 15:33)



So, let us first look at this r belonging to set of natural numbers. If r belong to set of natural numbers, how will you proceed? The answer is very easy, what is set of natural numbers? In our case in our course we have defined set of natural numbers to be equal to 0, 1, 2 and so on. So, in this case our if r belongs to this particular set then you can easily see that I can write this log of M raised to r as log of log to the base a of M into M into M this is done r times, this is done r times and then I can apply the law of or the multiplication rule of the logarithm which is this.

And I can simply get this as $\log_a M^r = \log_a(M \dots\dots M)[r\ times] = r\log_a M$. And therefore you will get the answer to be equal to this is equal to log to the base a of M r times. Now, if r is set of rational numbers, this is a situation becomes tricky still can be managed, but this will not prove, we will assume for the sake of convenience and naturally the proof or the answer lies when you study calculus when r belongs to set of real numbers, you can actually construct a sequence of rationals which will converge to real number.

So, what I have done here is I have partially proved this is partially proved the law of logarithm when r belongs to set of natural numbers. When you study the math 2, or math for data science 2, in that you will come to know how to handle these particular objects. But for us what is important is if you give me $\log_a M^\pi = \pi\log_a M$.

This is this particular property of log will be very handy while solving the problems, you imagine an irrational number which was in the index of some positive number is taken into the multiplication with respect to log, so this simplifies the calculations significantly and as I already mentioned when we discuss exponential function how the number e has arrived natural exponential function how the number e has arisen, the same logic applies when we will prove the result for a set of real numbers. So, let us not get into those details right now, but for our purposes this law is very crucial and we will use this law left and right.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

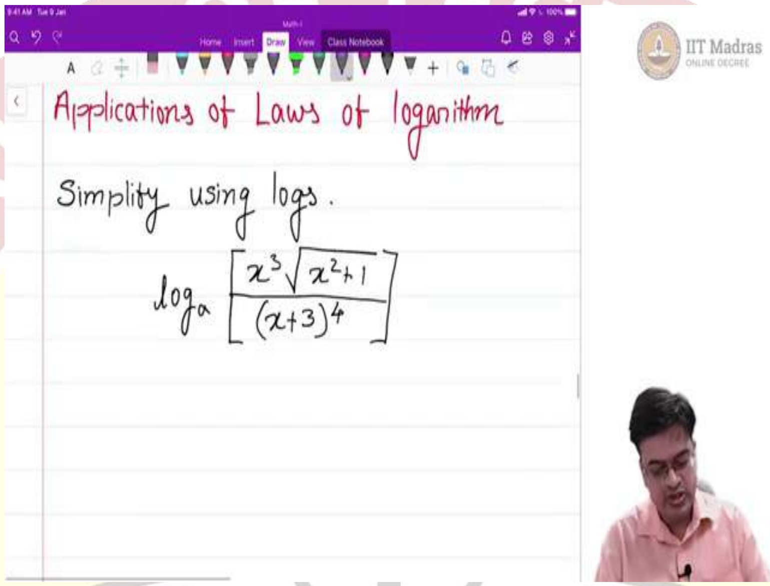
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

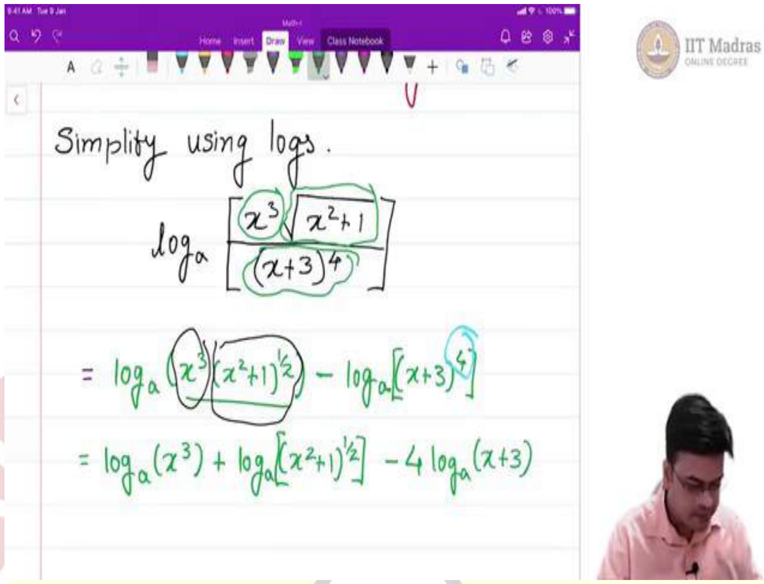
Mathematics for Data Science
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture No. 56
Logarithmic Functions: Applications

(Refer Slide Time: 00:15)



So, let us now go ahead and use these laws of logarithms and try to see some simple problems, how the problems can be simplified using logs or how the problems can be made complicated using logs.

(Refer Slide Time: 0:31)



So, let us see the first problem, it looks an ugly sum or ugly product and we have taken log and how the process is simplified when you use the properties of logarithm that you have studied just now or the laws of logarithm that you have studied just now. So, let us go ahead and do that.

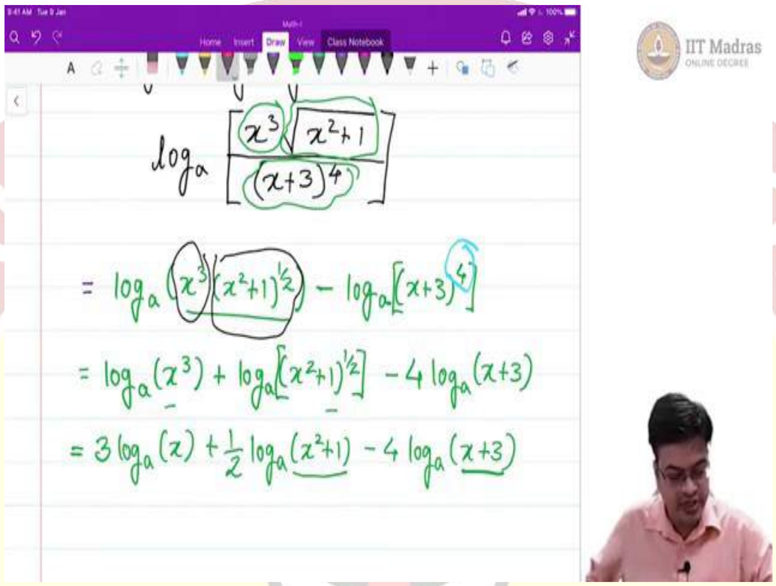
So, this particular thing let me write this to be equal to, so first you identify or isolate the terms that you identify can be separated. So, first term that I can separate out this is x cube, the second term that I can separate out is this square root and the third term is numerator. So, essentially this particular term if you look at can be split into 3 components, so I want end result that I want to get after simplification should have 3 components.

So, first I will apply the quotient rule that is $\log_a \frac{M}{N}$, so in this case this fetches me $\log_a x^3.(x^2+1)^{\frac{1}{2}} - \log_a (x+3)^4$. So, here I have not used any other rule so this is simply the quotient rule that I have used that is the second rule that we have derived.

Let us go ahead, and see what we can do with the first term that is this term. So, now you can see is simply see these terms can be characterize into two terms this is the first term and this is the second term. So, I can write this particular thing as$\log_a x^3 + \log_a (x^2+1)^{\frac{1}{2}}$. So, to identify the raise to half is inside the log I have put it this way then minus you look at this term again.

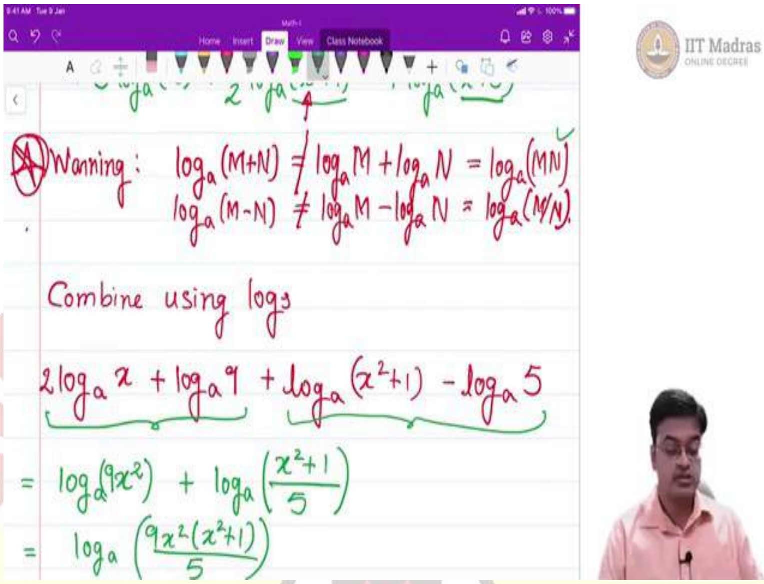
Now in this term something is raise to the power 4. Do I have any rule for indices of the law, indices within the argument of logarithm? Yes, we have just now proved it for set of natural numbers. So, you can use this that rule and say that this is equal to $-4\log_a(x+3)$.

(Refer Slide Time: 3:20)



Let us go ahead and do a similar thing for the other two terms that are listed here then we will get the final answer that is $3\log_a(x+3)$. 3 times log to the base a of x plus half times here I am using it for rational numbers which I have not proved $0.5\log_a(x^2+1) - 4\log_a(x+3)$. So, this is how I can simplify. Now in general, when you study logarithms people generally get confused between the product becoming the sum. So, here we have some terms like this. Now these terms cannot be handled with log.

(Refer Slide Time: 04:16)



So, right now let me give you a note of caution or a warning so to speak that is let me write this as warning. Generally though it is obvious but while doing the calculation people used this rule; $\log_a M + N \neq \log_a M + \log_a N$. So, this is what people use and this try to solve the problem so that they think they will simplify but remember this is not equal to that, why? Because we have just now proved that this is nothing but $\log_a MN$. So, these two things are different.

In a similar manner you can have a quotient rule that is $\log_a M - N \neq \log_a M - \log_a N$. Because this is actually equal to $\log_a \frac{M}{N}$. So, just remember this warning because generally in the when you are in the fighting spirit you are trying to solve the problem you tend to make these mistakes and which will ruin your entire answer. So, this is with extra star marked I am emphasizing that these two are not equal.

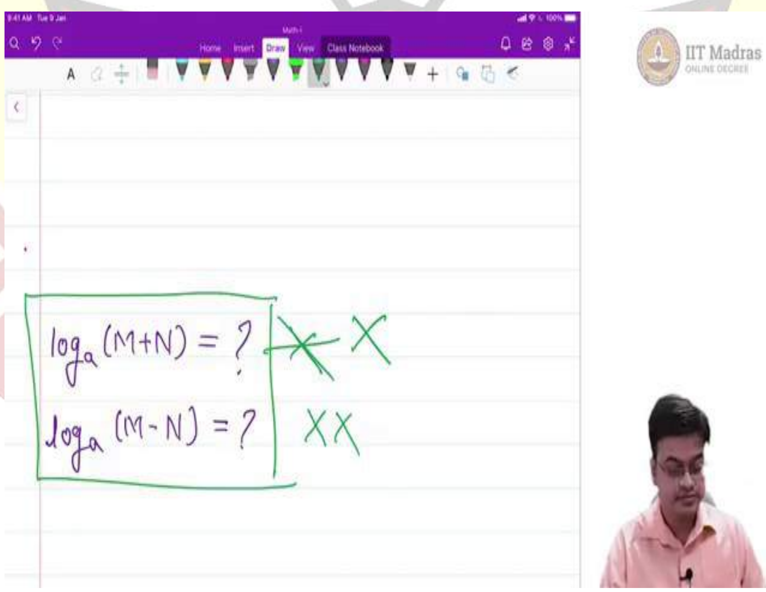
Now let us try to see how we can simplify, sorry we have here seen how we can simplify our life using logarithms where everything is now almost linear terms except for this quadratic term. Now the next question that can be asked is can you combine using logs? The answer is yes, so if you do not want to see such a big expression or you want to have a nice compact expression, the question is can you combine? The answer is yes, and now let us handle the terms one by one and merge the terms.

So, first term let us take these first two terms, what are these two terms? One is $2\log_a x$. So, you have already seen $\log_a x^r = r\log_a x$. Apply that in reverse so you will get this particular term as $\log_a x^2$, do not stop there you just apply the product rule now. $2\log_a x + \log_a 9$can actually be merged as we can use this rule and say this is equal to $9x^2$.

Now let us look at the next term which is this, next two terms in fact and there is negative sign so naturally a quotient rule will come and you will have something like $\log_a \frac{x^2+1}{5}$. Can you combine these two? Again apply the product rule and you will get this to be equal to $\log_a \frac{9x^2(x^2+1)}{5}$. So, this is how you can simplify your life while studying logarithms by giving a combined expression.

So, you can use this to simplify your life give a long expression with positive signs when it will help when you have lot of large numbers to be calculated. You can also combine the logarithmic terms and combine the expression in a compact form when this will help when you have lot of small-small terms that are unnecessarily occupying the space. So, when extremely large terms the simplification will help when extremely small terms will come the matter will be simplified by combining the terms. So, these are two simple avenues where you can actually do something.

(Refer Slide Time: 08:41)



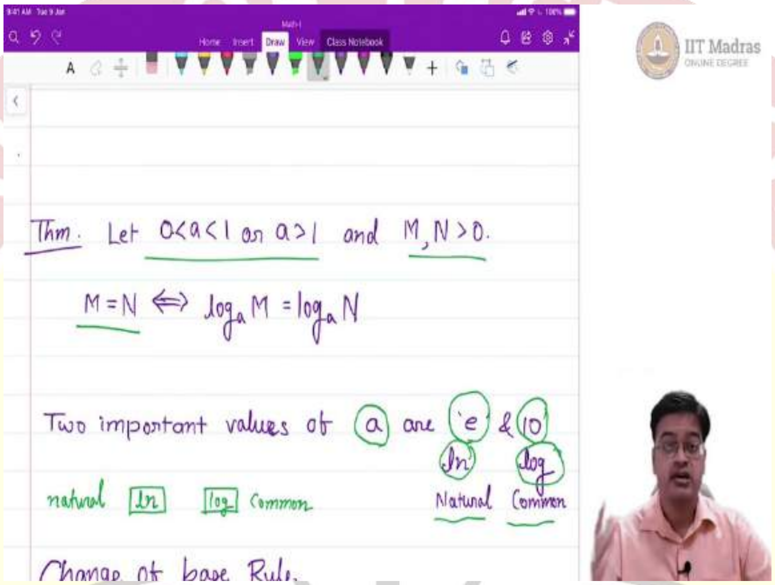
So, I have already warned you about this $\log_a M + N$, this is nothing you cannot get anything out of this, this is nothing. In a similar manner $\log_a M + N$ you cannot use reverse exponential or any other form to get something out of this while solving the problems of logarithms. So, just whenever these such terms come be aware and do not apply the things blindly.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology Madras
Lecture No. 57
Logarithm Function: Properties -2

(Refer Slide Time: 00:14)



Now, we already know some facts, but it is better to revise them when we are actually handling the problems. So, here is one such theorem that says that for a between 0 and 1, open interval 0 and 1 and for a > 1 and M and N > 0. So, these are essentially taking care of the conditions that we are in the valid domain of logarithm.

Then, if M = N we know that $\log_a M = \log_a N$. How this is derived? Because if you recollect we have already proved that our function logarithmic function is one to one function that means for every element in the domain there exists a single element in the co-domain or the range.

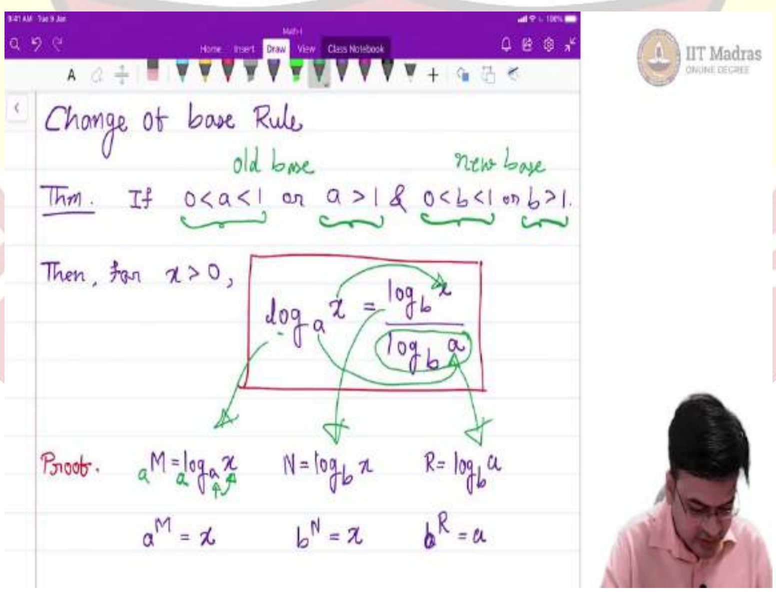
So, based on that you can actually see that if M = N $\log_a M = \log_a N$. Then, another thing that we want to see is two important values of logarithms. That is log, whenever we are talking about log to the base a, this a is typically limited to or restricted to when we are solving any engineering problem or any any stuff that involves some kind of engineering discipline, we generally restrict to two numbers that are e that is Euler's constant, which we have mentioned and 10.

So, when whenever you are talking about these two numbers when you are talking about $\log_e M$ we will give a notation ln and when we are talking about log to the base 10, we will give a notation in log without any base. Why we are doing this? These two have special names also natural logarithm and common logarithm. Now, this natural logarithm is as I already mentioned, it comes naturally in the theory of calculus and therefore it has some special value and if you are using a scientific calculator, then you will identify this natural logarithm as ln.

So, they are there are specific provisions given in scientific calculator because it is a natural logarithm and then if you look at the scientific calculator you can open on your computer as well, you will have some symbol of this kind, log this is log to the base 10 which represents our decimal system. Because in decimal system, everything is given in powers of 10. So, therefore this is called a common logarithm because we commonly used our decimal system and this is called a natural logarithm.

So, these are the two important logarithms that we can study in the entire life. Now, the problem that comes is why only these two why not others. So, here is my bold claim that only these two logarithms will suffice for studying logarithms and why?

(Refer Slide Time: 03:53)



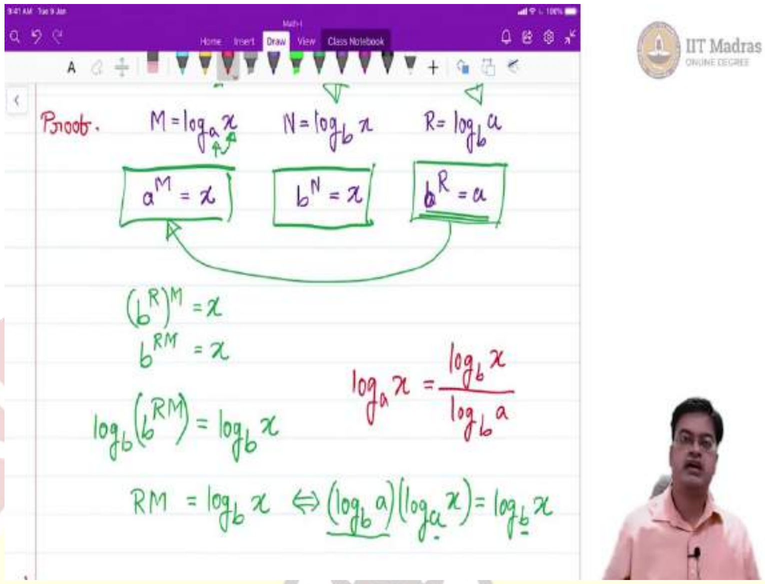
The answer to why is given by this particular theorem that is change of base rule. What let us, understand, what does this rule says, this is also called another law of logarithms. So, if you are handling x that is > 0 and if you are a is between 0 and 1 or if you are a is > 1. Remember this is a base of a logarithm. So, I am talking about change of base. So, I will talk about another base which is b, which is between 0 and 1 and b is > 1 this is called, this will be called as old base and this will be called as new base.

Now, if I want all the calculations to be done with respect to new base. If I want all the calculations to be done with respect to new base, then this theorem gives the answer, how should I go about this. So, the answer is for any $x > 0$ you consider log to the base a of x then what do you do, choice the new base that you want simply write the argument of the function over here in the numerator and the base in the denominator and write logs with appropriate base that you prefer.

So, in particular if you give me if this theorem is valid, we will prove this theorem is valid if this theorem is valid, then what we are actually talking about is you give me logarithm to any base, I will convert that logarithm into logarithm to the base 10 or logarithm to the base e and I will compute accordingly, that is the beauty of this theorem and therefore we have two important logarithms. So, we will stick to only those two important logarithms one is to the base e which is called natural logarithm. Another one is to the base 10, which is called common logarithm.

So, therefore, with this assumption, even our scientific calculators are designed to have only two keys that are ln and log they do not generally talk about log to the any base. If you have advanced scientific calculator, then it may talk about log to any base, but these two logarithms will suffice in general. Because what you are actually talking about is simple change in the basis is just a multiplication by a constant that log that constant will be given by this particular number, 1 upon this particular number. So, let us go and start proving this result.

(Refer Slide Time: 06:51)



So, for proving it is very easy if you understand the basic logic in all these proofs for proving laws of all laws of logarithm, what we are using is a key factor that is exponential function is inverse of logarithmic function that is all. So, in particular if I want to prove this, let us say M = $\log_a x$ and N =$\log_b x$, R = $\log_a b$.

So, what I am doing is I am mapping all these terms in terms of M, N so this particular term is N and this particular term is R. Now, because it is this, you look at the other description of logarithmic function. So, when you are talking about logarithmic function you are basically asking a question, if I have been given a base a, to what power I should raise this base a so that I will get x this is the question that we ask and answer to that question is given by M. So, essentially this means if I apply an exponential function, which is a raise to M, $a^{\log_a x}$, then I should get back a raise to M = x this is what I should get back.

And in a similar manner I will use this and I will I will use it for the second term and I will get b^N= x and I will get b^R = a. So, now what is happening is, if your number is a can be written in the form of b^R, if you are number a can be written in the form of b^R, then you can as well put this particular thing into this expression, that means I can the write a raise to M as b^{RM} = x that is justified.

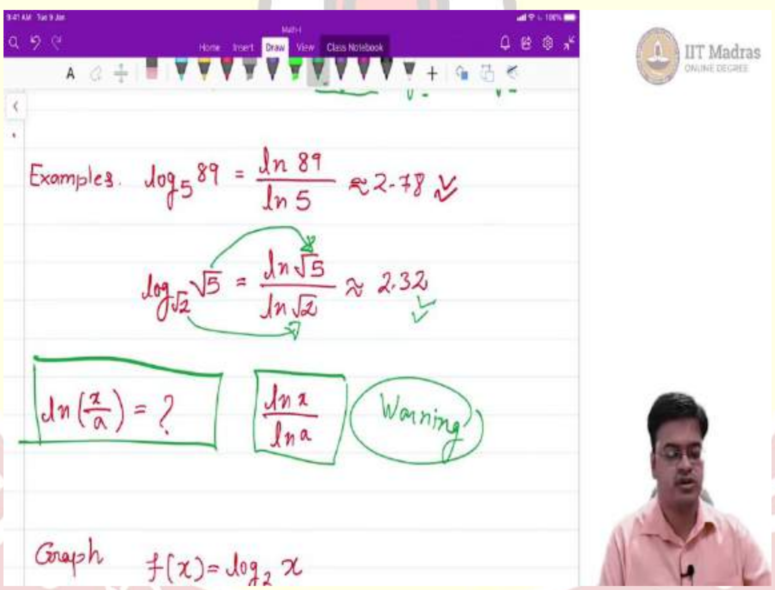
Now, what you have actually done you actually started with what. Let us, simplify this and this is actually equal to b^{RM}= x. Now, based on our understanding you simply hit this function with

$\log_b b^{RM} = \log_b x$, no confusion in this but this is actually inverse of this log function and is the exponential unction.

So, I will get back RM= $\log_b x$. Let us, substitute what is R and M and then we will resolve. So, this will happen if and only if, what is R? R is $\log_b a$ and this is $\log_a M$ is log to the base a of x solve. Let me erase it and rewrite it again. Which = $\log_b x$. So, I am I have justified what I stated.

So, when I want to switch from log to the base a or log to the base b, it is simply a multiplication by a constant that is $\log_b a$and therefore, our result is actually proved and what is what is our result that my $\log_a x$ is actually $\log_b x$which goes in the numerator and $\log_b a$, which is a proportionality constant comes in the denominator, that is all. So, now you forget about all other bases and you simply try working with natural logarithm or common logarithm that is the key idea that we will follow. So, let us use this fact and try to prove certain things like this.

(Refer Slide Time: 11:58)



For example. Now, if you have been asked some question like, what is $\log_5 89$? So, you do not have to go into much detail that what is log to the base 5? You can simply use natural logarithm and use the change of base formula and get the answer to be $\frac{\ln 89}{\ln 5}$ use your calculator it has a natural key, which is ln, compute it you will get the answer to be equal to 2.78.

Somebody gives you any absurd number some irrational number $\log_{\sqrt{2}} \sqrt{5}$. Still you do not have to worry just apply $\ln \sqrt{5}$, $\ln \sqrt{2}$ change of base formula you will get $\frac{\ln \sqrt{5}}{\ln \sqrt{2}}$ that is this thing is going to the numerator argument is going to the numerator base is going to the denominator and you know ln is nothing but natural base log with natural base. So, this will become 2.32 that is all. So again, you can use scientific calculator and get the answer. So, this is how they were calculation simplifies no matter what base is given to you, you can easily solve all the problems.

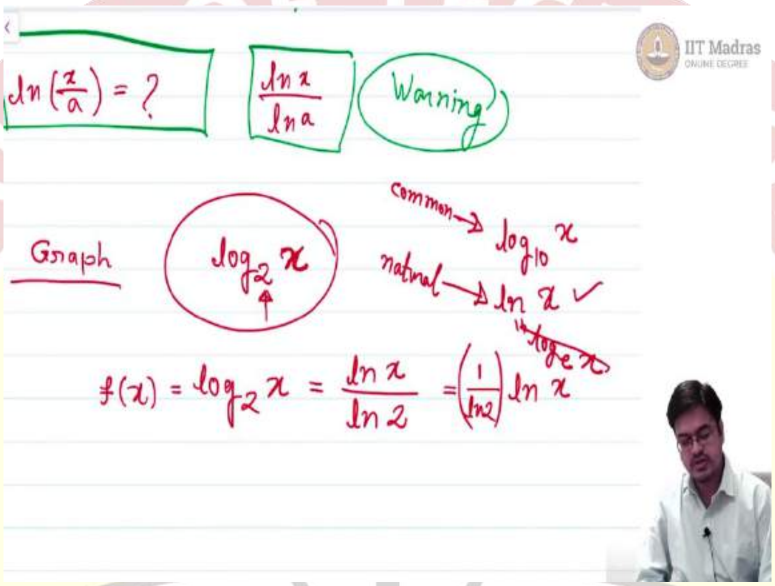
Now, sometimes people confuse with this kind of identity $\ln\frac{x}{a}$ they simply write it as $\frac{\ln x}{\ln a}$, which is not true. So, just this is just a warning that I want to give in particular. So, this identity is not at all true. And therefore, you have to be careful while solving the problems. See in the verge of solving the problems you made tend to do these kind of mistakes, which will ruin your entire answer. So, in the next video, we will come up with better versions of graphing the logarithmic function to any base using natural logs, or some other variations. Right now we will stop here.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Neelesh S Upadhye
Department of Mathematics
Indian Institute of Technology, Madras
Lecture No. 58
Logarithmic Equations

(Refer Slide Time: 0:14)



Hello students, in this video what we are going to do is, we are trying to look at the logarithmic function and how to solve equations using logarithmic functions. So, this is the goal of this particular lecture, so let us first get a simple πcture, we have already seen how to plot a logarithmic function. Suppose now you have been asked to plot a graph of a function which is $\log_2 x$.

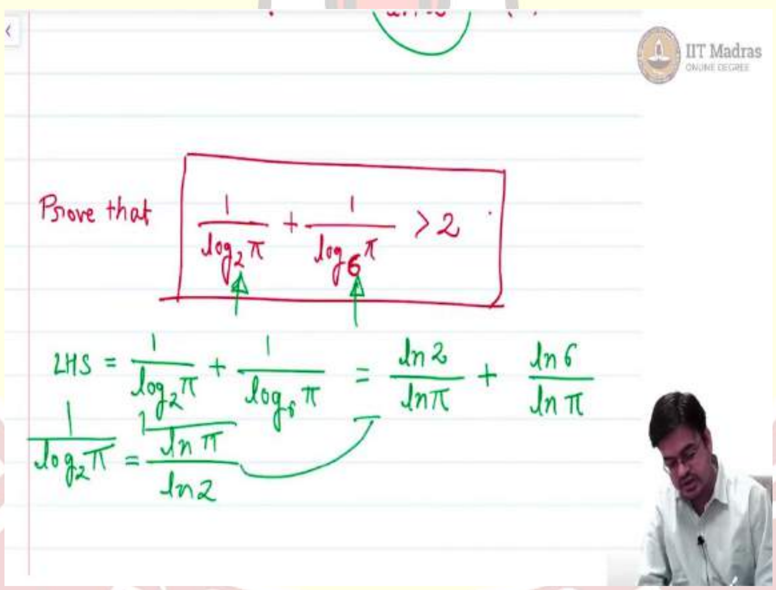
If you have a graphing calculator, sometimes the old version of graphing calculators do not allow $\log_2 x$ to be taken and in that case what you have is either $\log_{10} x$or $\log_e x$which we have denoted by $\ln x$, this is not e, $\ln x$. So, only these two things are available to you and you can plot these two things, then can you plot the function log to the base 2 of x? This is what the first exercise that we will do when we try to solve the problems, so f(x) is log to the base 2 of x.

Now using the change of base formula which we derived in the last class, you can easily convert this function × a function with log to the base 10 or log to the base e. For convenience I will choose log to the base e. So in this case, I can simply convert using my change of base formula, this is $\frac{\ln x}{\ln 2}$ 2. That simply means, what I am doing is, I actually need to plot $\ln x$ and scale it appropriately with a constant which is $\frac{1}{\ln 2}$. We can easily evaluate the value of $\frac{1}{\ln 2}$ using any calculator and this is just multiplied with $\ln x$.

So, the graph will more or less have similar features of $\ln x$, only thing is it is scaled appropriately. You can try your hand in plotting this graph. This is a clear cut demonstration of the usability of the change of base formula. So, now you need not have to bother about, to what base the function is given. You can simply convert the function × log to the base e or log to the base 10.

This is I have roughly, I have already told this is called common logarithm and this is called natural logarithm. That is why the name ln and once again I reiterate ln means this is $\log_e x$. I will differ from this notation and I will use this notation for convenience and it is a standard convention in mathematics to write this as ln. So, this is a simplest application of graphing any log to the any base of a function.

(Refer Slide Time: 3:29)

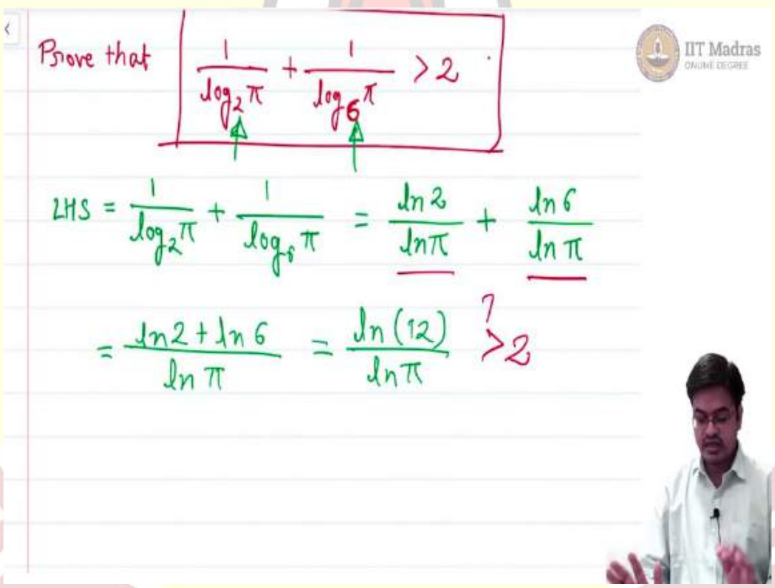


Let us now come to somewhat tricky application of this, that you have to prove that log to the base 2 of π + 1 by log to the base 6 of π is > 2. How will you prove this? Now the key tool while solving all the logarithmic equations or inequalities is exponentiation. So, we will use that tool over here. So, let us start and try to solve this problem. First of all, if you notice in this particular problem, one base is 2, another base is 6, I do not want this. I want everything to the same base.

So, I will use the same principle that I have used here, that is change of base formula. So, let us try to take the left hand side, LHS, which is nothing but $\frac{1}{\log_2 \pi} + \frac{1}{\log_6 \pi}$, apply the change of base formula. So, $\log_2 \pi$ can naturally become $\frac{\log_e \pi}{\log_e 2}$log to the natural base e of π upon log to the natural base e of 2, which essentially gets converted × $\frac{\ln 2}{\ln \pi}$. This is again simple application of change of base formula.

In a similar manner, I can write this as $\frac{\ln 6}{\ln \pi}$l, to be very precise what I did is, I substituted $\log_2 \pi = \frac{\ln \pi}{\ln 2}$ and because this thing was in the denominator because the original fraction was 1 over this, so this is 1 over that and therefore, it will give rise to this particular number. Fine. So, this is done, you do not have to worry about this.

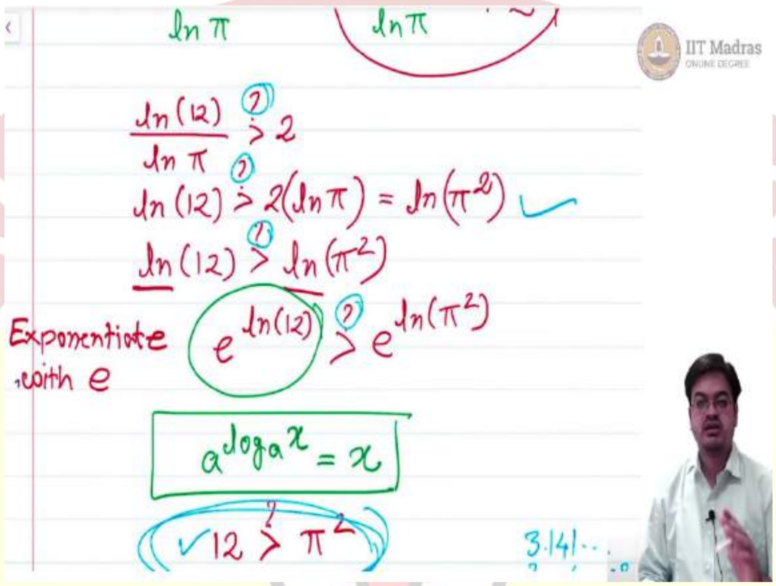
(Refer Slide Time: 5:43)



So, now you can see something amazing has happened. The denominator now is actually $\ln \pi$, denominator is same. So, I can rewrite this expression as, let us rightly write this expression as $\frac{\ln 2+\ln 6}{\ln \pi}$. Wonderful. Now do I know something about the laws of logarithm, this is $log\ a + log\ b$, we have already solved this. This is $log\ ab$, so I know this is $\ln(2 \times 6) = \frac{\ln 12}{\ln \pi}$.

Now this is about LHS. So, LHS actually simplify it to $\frac{\ln 12}{\ln \pi}$. Now the question is whether this thing is > 2, I do not know still. Let us say, assume this thing holds true, then how will you proceed? So, let us try to do it in this fashion.

(Refer Slide Time: 6:55)



I will again consider this particular inequality and try to prove it, try to see whether this is true or not, l$\frac{\ln 12}{\ln \pi} > 2$ I do not know whether this is true or not. But I know $\ln \pi$ is a non-zero number. So, I can simply take $\ln \pi$, is ln π positive or negative? It is positive, so I can simply multiply throughout by ln π, so ln 12 > 2 × ln π.

Now I have one law again to my aid that is multiplication rule, so log of x^a is a×log of x, that rule I will use and this in fact will become=ln π^2. So, now I am checking ln 12 is > ln π^2d or not. This is the question that we are asking again. So, all these are questions, we have not yet proved anything. Now both side logs are to the same base, so I can exponentiate this, I can exponentiate this with e, Euler's number, Euler's number.

So, if I exponentiate this, then because exponentiation, the operation of exponentiation is nothing but applying exponential function to a particular argument that is monotone, it is monotonically increasing. So, I will have inequalities intact, therefore $e^{\ln 12} \geq e^{(\ln \pi)^2}$.

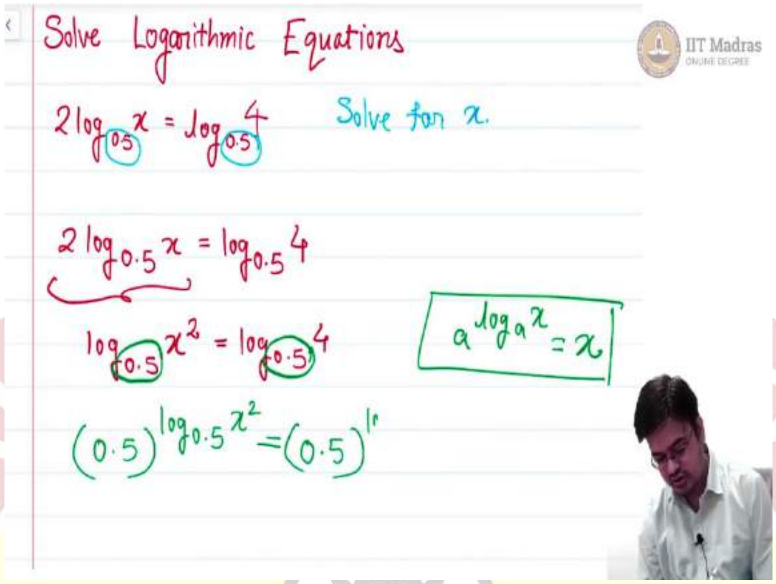
So, now you can simply go ahead and try to solve this particular problem. What is $e^{\ln 12}$ this particular thing, $e^{\ln 12}$? $e^{\log_e 12}$, so I already have proved that $a^{\log_a x}$ is nothing but x, this we already know using our inverse function definition. So, I will apply this definition over here and therefore, my $e^{\ln 12}$ will simply convert to 12 which is $>\pi^2$ π or not? Is it true or not, we have to check.

So, now you look at the way we started, we started with some complicated inequality which is given here, the inequality, if you include this, some complicated term on the right hand side some number was there and now we made it tailored to our understanding which is, whether 12 is $>=\pi^2$ or not? Now how to prove this? Very easy, what is π basically? It is 3.141 something, something. This number is strictly > 3.15, strictly smaller than 3.15.

So, the π^2, the π is smaller than this, therefore, π^2 will also be smaller than 3.15^2. And 3.15^2 will not exceed 10, you can check for yourself, it will not exceed 10. Therefore, this inequality which is under question is certainly true because this π^2 will always be less than 10, which is less than 12 and therefore, naturally this inequality holds true. So, I do not need to put a question mark over here, I do not need to put a question mark over here, and all these inequalities are true.

And therefore, we have proved that this particular inequality in particular is true. This is the way we will solve a question which is using logarithms. Now you can simply identify what technique we have used, we have used 3 laws of logarithm; first law of logarithm that we used is change of base, second law is the multiplication rule, next the multiplication rule repeated again, here multiplication rule repeated again and then the third law, it is not a law, it is actually the definition of inverse function that we have used. Using these 3 together we were able to find our answer.

(Refer Slide Time: 12:01)

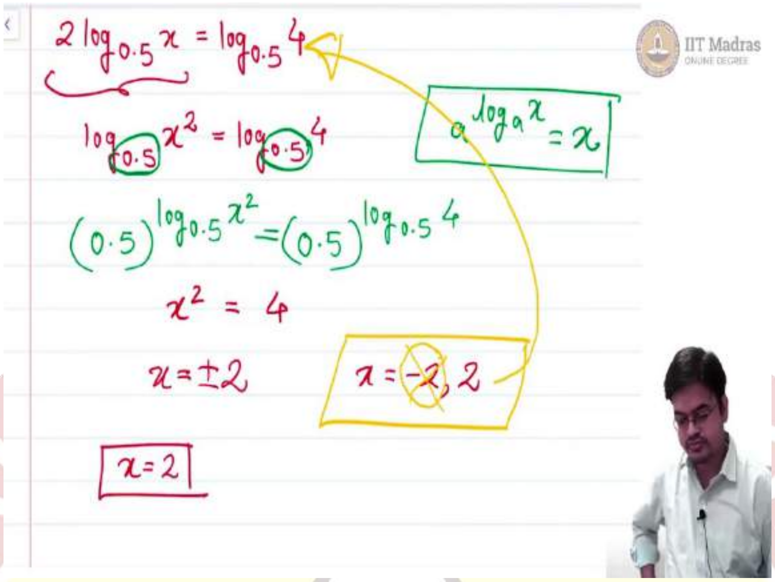


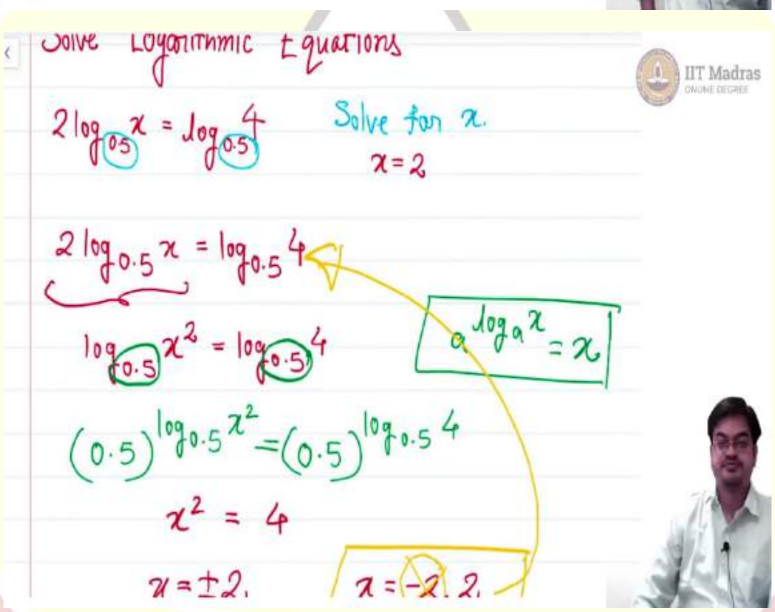
Let us try to solve some slightly complicated problem like this. Let us say you are asked to find, solve log to the base 0.5 of x = the base 0.5 of 4. So, what do I mean by solve, this is actually solve for x. So, you are interested in finding the feasible values of x which will satisfy this equation. Now at the beginning you may be worried about this 0.5 in the denominator because if you remember for 0 less than a, if a is the base, a less than 1, the behavior of log function was somewhat different.

Do I really need to worry about it, is the first question? Before worrying about anything, let us try to simplify this expression, so what is this equation saying, let us write this 2×log to the base 0.5 of x = log to the base 0.5 of 4. Now first thing, if you look at the left hand side, is there any rule that I can apply? Yes, I can apply that power law, power rule. So, I can simply convert this × $\log_{0.5} x^2$ and then let it be as it is, so it is $\log_{0.5} 4$.

Now if you look at this particular expression and if you look at the denominator which is 0.5, here also 0.5, not denominator, base. So, if you look at this base, you can easily say that the base is common. Therefore, the exponentiation trick which we have floated in the last class that is $a^{\log_a x}$ is x, so this trick will work. And therefore, I will exponentiate this with respect to 0.5. So, I will write $0.5^{\log_{0.5} x^2} = 0.5^{\log_{0.5} 4}$. What does this mean?

(Refer Slide Time: 14:31)



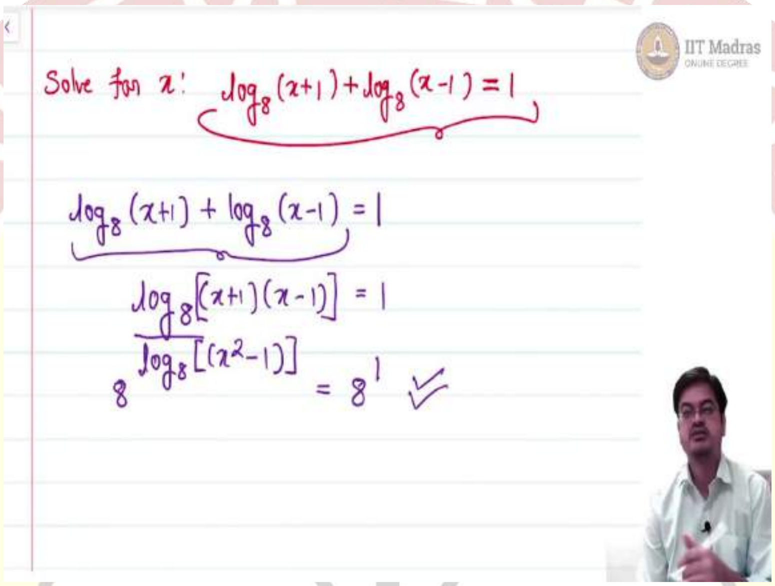


This simply means I will get here $x^2 = 4$, $x^2 = 4$ and now because $x = 4$, based on my knowledge about the quadratic function I know $x^2 - 4 = 0$ has two roots, that is $x = +$ or $- 2$, these are the two roots that are available. Now that means I have two solutions to this particular problem $x = - 2$ and $x = + 2$.

The next question that you should ask is are these both solutions feasible when I substitute them × this expression? So, what is log to the base 0.5 of - 2? When you look at the logarithmic function, it is defined only on the positive real line, it is not defined on negative real line. So, log to the base 0.5 of - 2 is indeterminate, you cannot determine the value, the function is not defined, it is outside the domain. So, this value you can easily chuck off.

And therefore, your solution, solution to your problem is $x = 2$. This is the solution for the logarithmic equation which we are finding, so solve for x, the answer will be $x = 2$. You can simply substitute it here and check, you put it to 2^2 is 4 and log to the base 0.5 of 4 = the base 0.5 of 4. What if you put - 2? This is not valid. Correct. So, this way you have to verify once you get the final answer.

(Refer Slide Time: 16:22)



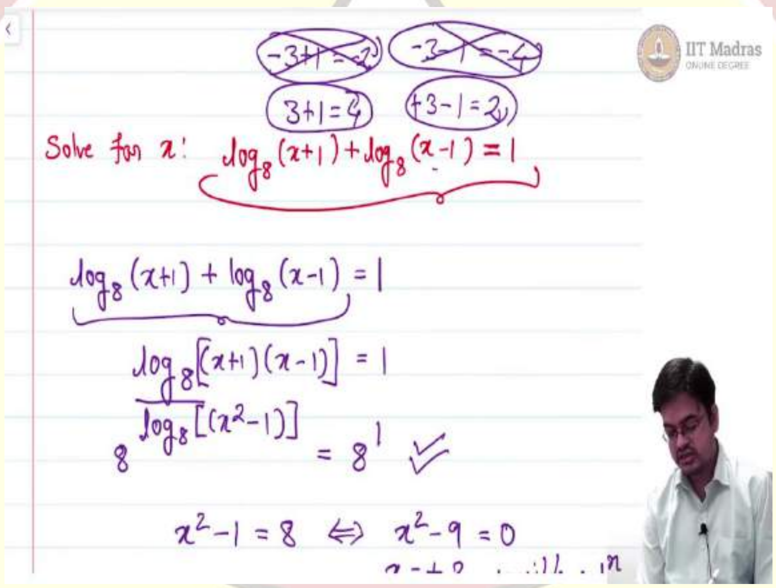
Let us handle somewhat more difficult problem which is again going in a similar line but exponentiation will again help you, but it will reveal some important traits over here. So, let us look at this particular problem where the LHS is log to the base 8 of x + 1 and log to the base 8 of x - 1. Let us try to simplify, let us start with LHS. We want to solve for x, so solving, taking LHS will not help. So, let us take the entire, entire thing that is log to the base 8 of x + 1 + log to the base 8 of x - 1 = 1.

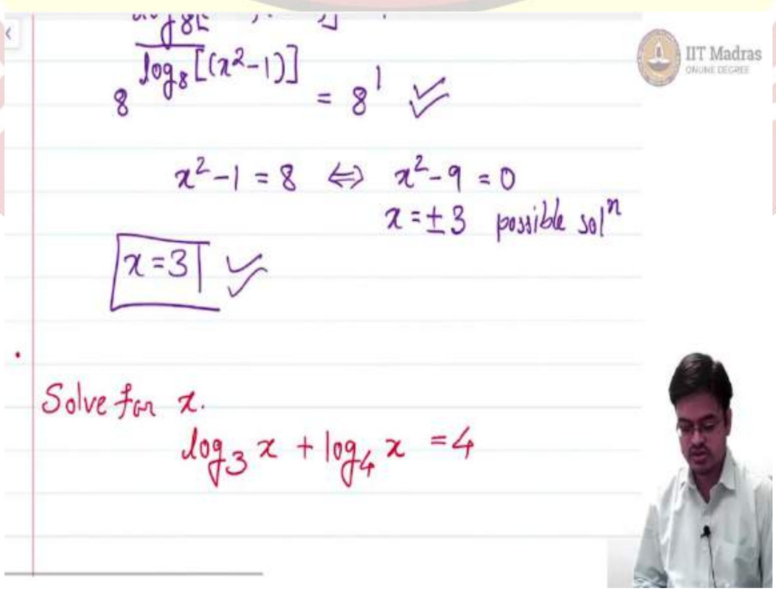
It is important to write the equation as it is, because if you write the equation as it is, you will understand the nitigrities of this equation. So, it is good practice to write once, repeat once whatever is written there, therefore I am writing this. This, do I know any rule, any law of logarithm which will help me to simplify this? Yes, I know multiplication rule, log of m + log of n = log of mn. So, I will apply that rule and I will get log to the base 8 of x + 1 × x - 1 = 1.

Now what can I do? What is the way to simplify? Now I want to get rid of factor of law, so how will I get rid of this log factor? I will exponentiate. So, what I will do is, I will to the, what is the base 8, so $8^{\log_8}$ of this particular factor, I can rewrite this as x^2 - 1, which is easy for me to do and then this = 8^1. Remember if you do not write this step, you may miss out on this, you may write this to be=1. So, just write all these things.

Then 8 log to the base 8 will vanish and therefore you will get $x^2 - 1 = 8$. Do I know something like this? We can, I know that $x^2 - 9 = 0$. That means x = + or - 3 are the possible solutions. Now is any, so in the earlier case when x = + or - 3 were the solutions is there any, so for example here in this case - 2 was eliminated. So, here also you need to do a similar check, if you put x = + 3, if you put x = + 3, then what will happen?

(Refer Slide Time: 19:25)

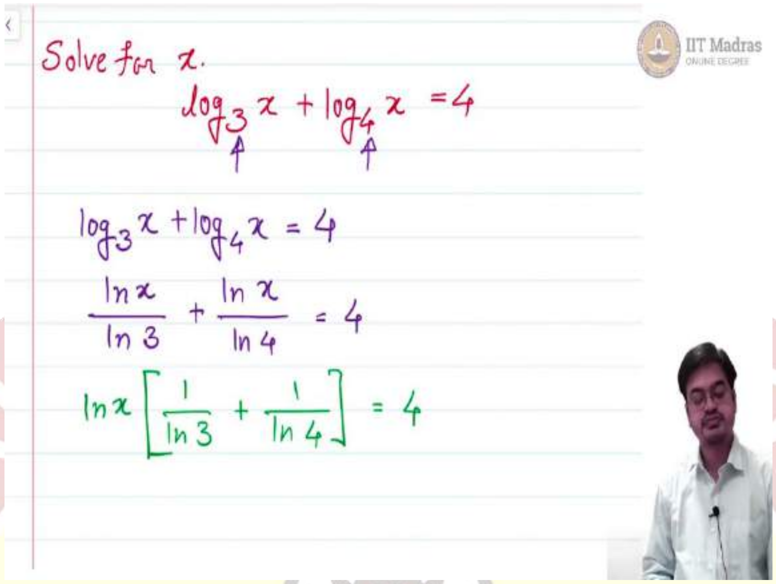




This 3 + 1, 4 and - 3 - 1, sorry + 3 - 1 which is 2 and therefore, this is 4, and therefore it is a valid expression. So, it is in the domain. If you put x = - 3 and - 3 + is - 2. This is not a log, what happens here - 3 - is - 4, which is also not a log for putting × log. So, - 3 cannot be in the domain.

So, only thing that is possible is x = 3 should be there in the domain, again when you have solved the equation, it is better to verify, so you write here $\log_8$ 3 + 1 which is 4 and 3 - 1 which is 2 = 1. This is valid, so you have good answer. So, x = 3 is the answer to this question. Let us move ahead and solve one question which has different bases.

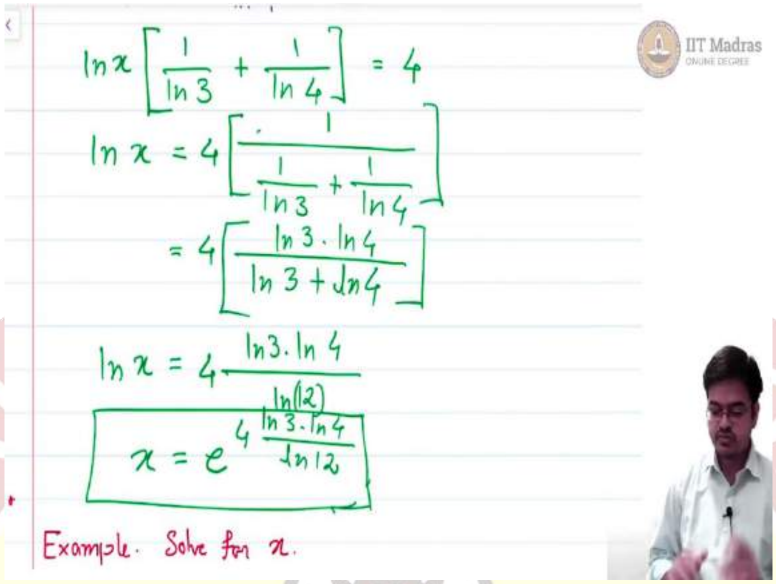
(Refer Slide Time: 20:32)



So, here is a question which is log to the base 3 of x + log to the base 4 of x = 4. Now what can you do? Because you cannot use your trick of exponentiation, because this bases are different. So, what should you do? The first thing is to make the bases equal, how will you make it? One formula is change of base formula, so you apply change of base formula, so as I mentioned it is better to write the expression once more, log to the base 3 of x + log to the base 4 of x = 4.

If I want to apply change of base formula, I will simply use this as log to the natural base, so $\frac{\ln x}{\ln 3}+\frac{\ln x}{\ln 4} = 4$. So, $\ln x$ is taken in common and therefore I have $\frac{1}{\ln 3}+\frac{1}{\ln 4} = 4$. Now you can notice that this particular within the brackets is a number which is non-zero.

(Refer Slide Time: 22:13)

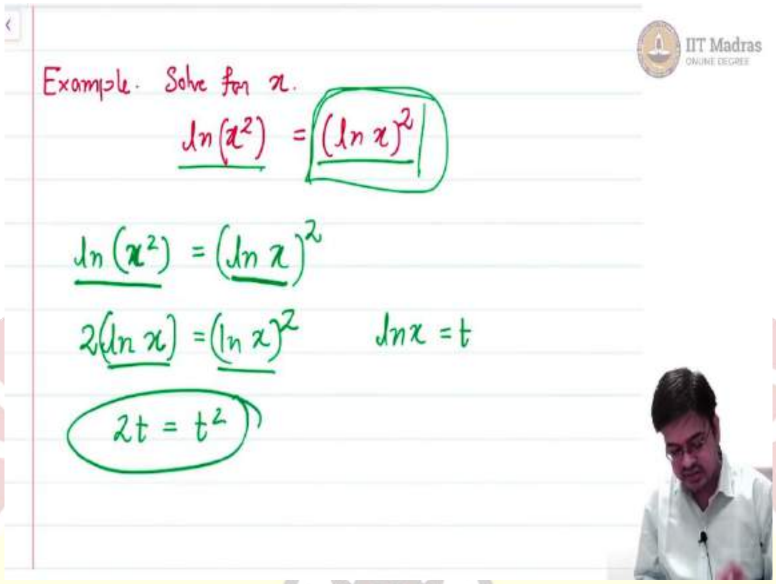


Therefore, I can actually take this number on the other side. So, let me simplify this, ln x = 4 times, when I take this on the other side it will be a reciprocal of that, so $\frac{1}{\ln 3} + \frac{1}{\ln 4}$, this is what it will be. So, can I simplify this? Yes, I can simplify this further, which will give me 4 times, so ln 3 + ln 4 will be in denominator and in the numerator it will give me ln 3×ln 4. Fine.

Then I need to do something which is let us say, I will do it in this fashion, ln 3 × ln 4 cannot be combined × anything, they will remain as it is. But what can be done is, this is ln 3 × ln 4 upon ln 12, 4 × 3, I have used the multiplication rule. So, let this, all these things remain as it is and because it is ln, I will exponentiate and I will get $x = e^4 \times \ln 3 \times \frac{\ln 4}{\ln 12}$.

If you want you can merge these 4 with one of these ln's and write, with one of these ln's and write this as $\ln 3^4$ or $\ln 4^4$, whatever you are comfortable with, but I will leave this as it is. So, this is the solution and it is a perfectly valid solution. This is how you will solve a problem which has different bases. So, this is a perfectly valid solution. Let us go ahead and try to solve one more example, which is, which looks somewhat ambiguous.

(Refer Slide Time: 24:41)

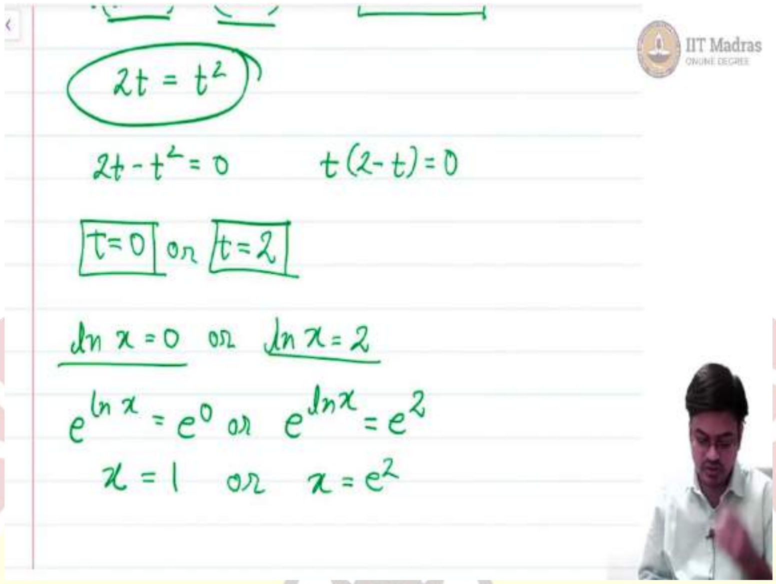


In a sense you have been given ln x^2 = (ln x) ^{2}. How will you solve this problem? That is a first question. So, we need to resolve to some methodology, let us look at this particular problem and try to simplify the things over here. So, here you typically, because this isd, you typically need some knowledge about quadratic functions in order to solve this problem. Let us try to understand this.

So, let me write ln x^2 = (ln x) ^{2}, in such a problem can you figure, the first question is, are the bases common? Yes, the bases are common, so there is no problem with this. Now what is happening is, here the term is ln x, here the term is ln x^2. Can I get the terms which both are in the form of ln x, then I can do something with it. So, I will simply ask that question and what comes to my aid is the multiplication rule or the power law, ln a^k is k×ln a. So, 2×ln x = (ln x) ^{2}.

Now comes the real power or the real strength. What is happening? It is 2×ln x, now the argument both are same, so you can treat this as composition of functions. So, what is happening is, if you write ln x = t, you are actually talking about 2t = t. So, what I am doing is, I am putting ln x = t, this I am defining. Fine. So, this particular thing is fine.

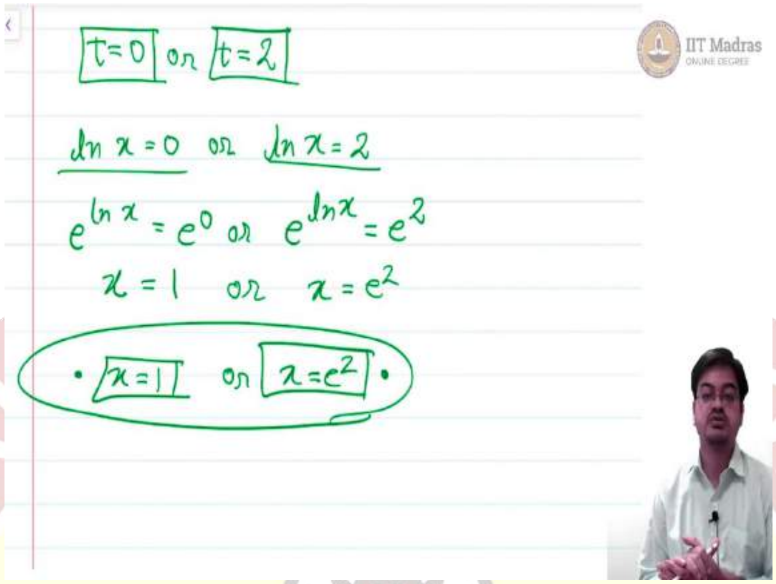
(Refer Slide Time: 26:55)



So, what I will do here, very easy then, it is 2t -, basically $2t - t^2 = 0$, take out one t common, so $t \times 2 - t = 0$ and therefore, either t will be=0 or t will be=2. These are the two possible solutions. So, now, but what is t? According to our substitution it is $\ln x$, so that means I am saying $\ln x = 0$, when is $\ln x$ 0? And $\ln x = 2$, so when is $\ln x$=2? So, these are the two questions that we have asked. So, let us say or.

And then in both cases you have a natural log, is not it? So, I can exponentiate this particular function, so $e^x = e^0$. You already know logarithmic function has a point where it passes through 0 and e^0 is, $e^{\ln x}$ will be x itself, e^0 will be 1, in this case exponentiating will give me $e^{\ln x} = e^2$ by default I wrote, so I will write it as e^2. So, that will give me $x = e^2$ or $x = 1$. This is what the answer is.

(Refer Slide Time: 28:50)



So, in this case x = 1 is one answer or x = e^2 is another answer. It is good to go to the original problem and check whether these conditions are satisfied in the original problem or not. So, first let us check x = 1, so if x = 1, ln 1 = ln 1^2. What is ln 1? You know the answer log of 1 is 0 and ln, so 0^2 is 0. And then if you put x = e^2, then ln e^2, that is 4 and ln x that is ln e^2 is 2 and 2^2 is also 4.

So, we have verified that both these are valid answers. So, what I have said just now is ln 1^2 is 1, so it is 0. 1 is 1, so ln x^2 is 0. Similarly, ln 1 is 0, so 0^2 is 0. So, the first x = 1 is the solution. In the similar manner, you substitute x = e^2 and you can plot.

Now it is good think if you can use a calculator, graphing calculator like Desmos and plot these two curves and see how, where they are intersecting. We have actually given the points of intersection of these two curves. These are the points of intersection of these two curves. You can verify for yourselves and I will see you in the next class. Thank you.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

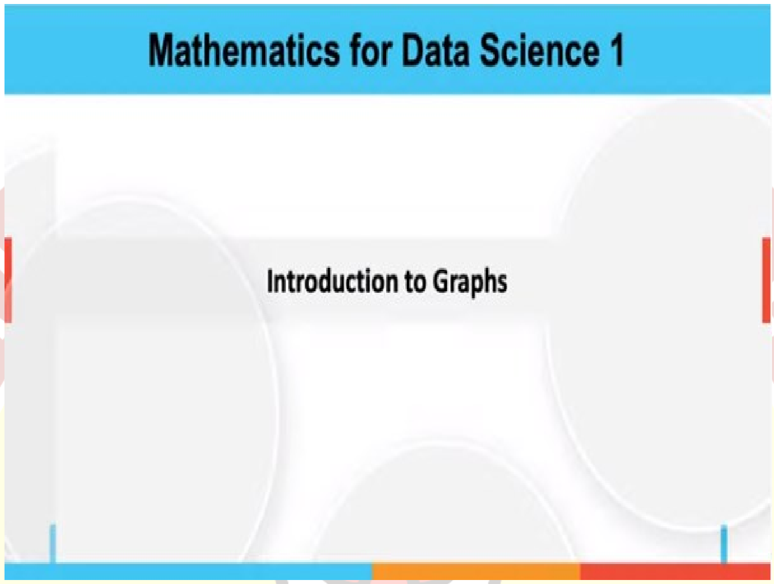
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

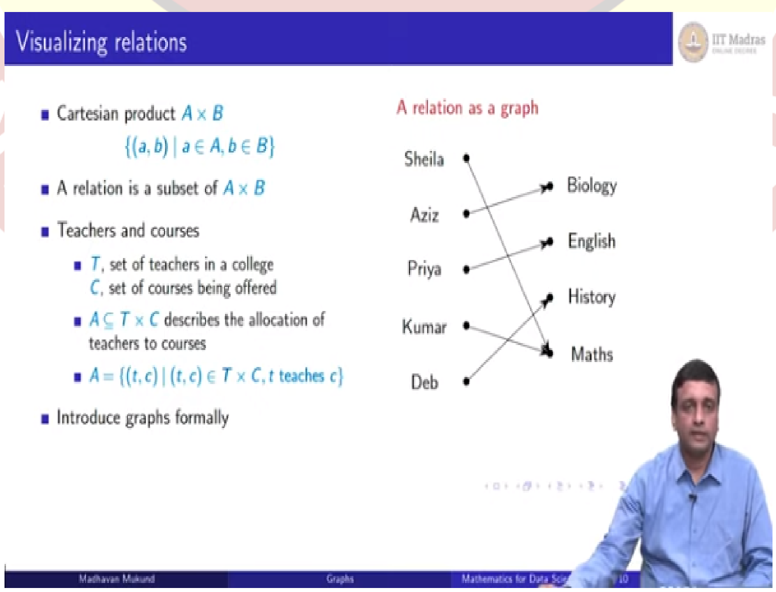
Mathematics for Data Science 1
Professor Madhavan Mukund
Indian Institute of Technology, Madras
Lecture 59
Introduction to Graphs

(Refer Slide Time: 0:09)



So, for the next few weeks we will be looking at graphs, so this will be the concluding section of this maths course, so for the next three weeks we will looking at graph.

(Refer Slide Time: 0:24)



So, we saw graphs in the first week when we were talking about relations, we said that we can take a relation, so what is the relation? We take two set they could be the same set, we take the Cartesian product all pairs. So, we take $A \times B$ and then we take some subset of that Cartesian product, so some pairs, out of the total set of pairs and we say that these pairs are related.

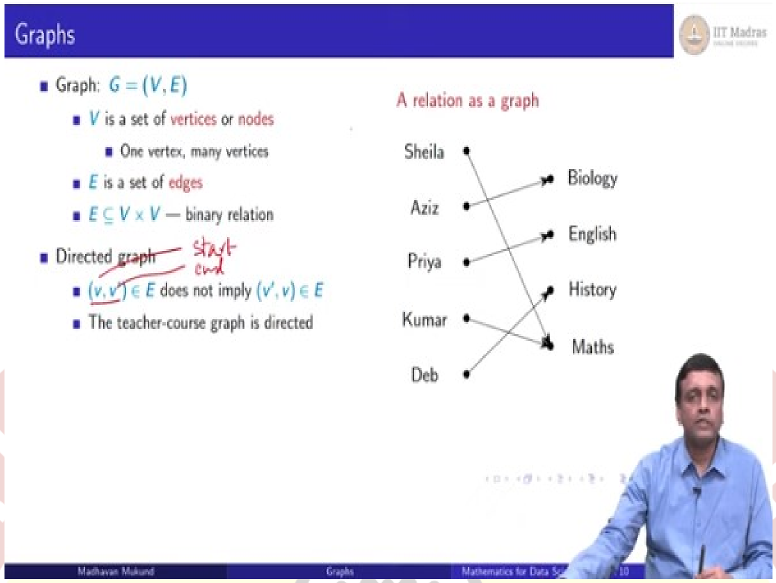
And then we said that we could visualize this, so for example, supposing our set A is a set of teachers, so let us call it T and the set C is set of courses that are being offered in the current semester, then we could have a relation which captures, which teachers are teaching which course.

So, we have $T \times C$ as a set of all possible pairs where the first element is the teacher and the second element is a course and then this relation A, which is a kind of course allocation relation describes how teachers are assigned to courses in the current semester. So, what we said was we could draw this relation as a pictorial form, so we could create these nodes or dots, representing each element in our set.

So since, there are two different types of sets, the set of features instead of courses we write them in two columns like this, so we have 5 teachers and we have 4 courses and whenever a teacher is teaching a course, we connect that teacher to the corresponding course node through an arrow, so in this case, since there are 5 teachers and 4 courses, clearly there must be at least one course which has been taught by 2 different teachers in this case you can see that maths is being taught by Sheila and Kumar.

So, what we are going to do in the next three weeks is to look at this picture that we have drawn earlier just as visualization of a relation, we are going to look at these pictures, these pictures are called graphs and we are going to formally analyze what we can do with these graphs.

(Refer Slide Time: 2:19)



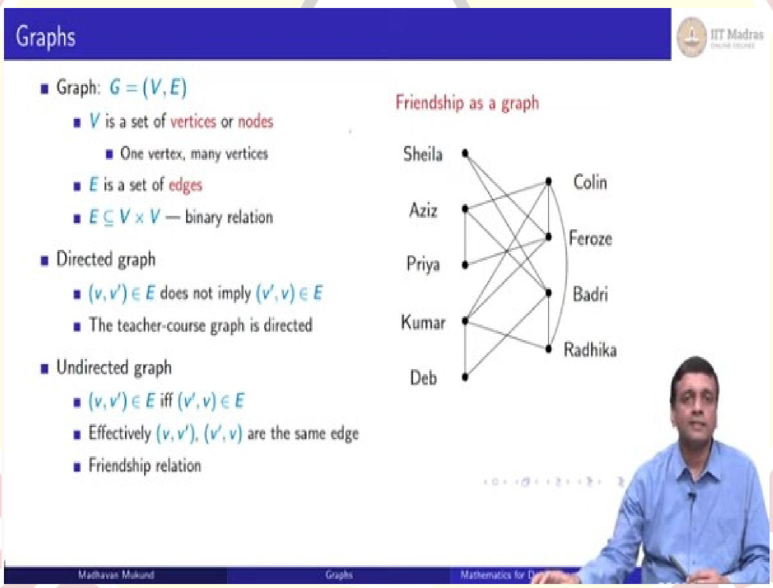
So, to begin with a graph consists of a set of nodes or vertices and edges between them. So typically, a graph therefore, has two components a set of vertices and a set of edges, so vertices is the plural of vertex. So, we have one vertex, many vertices, so we use interchangeably, either the node or vertex, as a name of for these elements and then what edges do is that they connect these vertices. So, notice in this graph that we had drawn before we had earlier two set, we had the set of teachers and we had the set of courses, and we were taking a relation which is a subset of T×C.

But once we put it into the graph we lose or we do not really care about the distinction between T and C, T and C together form the set V of vertices of the corresponding graph, so now there are nine vertices, there is no real separation between the five that came from the teacher set and the four that came from the course set and then the edges were those that the original relations represented namely, those teachers which are teaching the courses.

But in general, E is just a binary relation on the vertices so, it connect some vertices to some other vertices. So, this graph, for example, has a direction, right a teacher teaches a course, we do not have a corresponding edge from a course back to a teacher. So, if v, v′ is an edge it does not necessarily mean that v′, v is an edge in this particular relation it does not even make sense for v′v to be an edge but there could be other relations where it does make sense as we will see.

So, this kind of a graph is called a directed graph, so we have an order, we have a starting vertex for each edge and we have an ending vertex for each edge and you are suppose to go from the start to the end, you cannot go backwards, so think of it like a one way road. So, there is a one-way road from the start vertex to the end vertex, so every edge is of this form, right, it is a pair, so there is a start and there is end, so this is how you should think of an edge. So pictorially we just draw it as a line with an arrow but mathematically, it is an element of E × E so it is a pair, so the teacher course graph is directed.

(Refer Slide Time: 4:40)

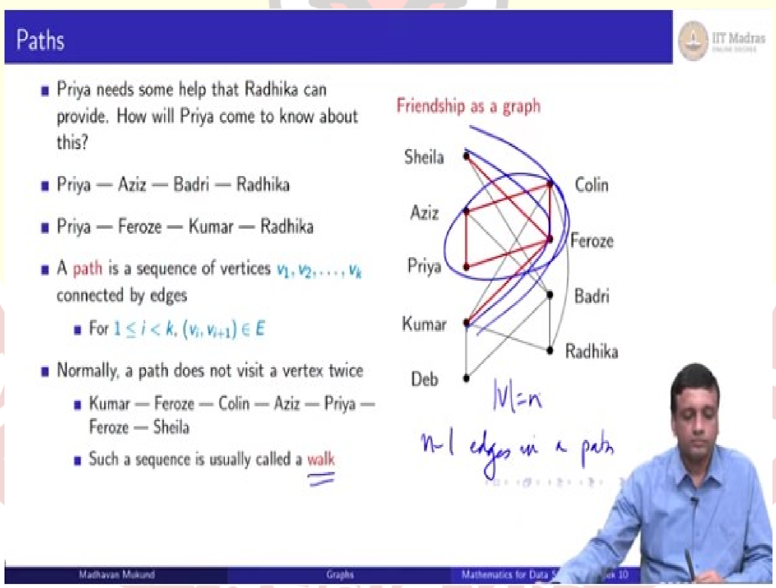


On the other hand, supposing we are just looking at a bunch of people and we are trying to capture, which of them are friends of each other. So, this now becomes an undirected graphs so presumably if Sheila is a friend of Badri, then Badri is also a friend of Sheila, because friendship is not a one directional thing, you cannot be my friend if I am not your friend.

So, in this case, we do not have an arrow, we just have pairs, which represent in some sense a symmetric relation, if you remember we talked about symmetric relations. If a,b is in the relation, then b,a is also in the relation. So, this is what we have here, we have a symmetric graph, which says that if v v′is an edge then v′v is also an edge.

So, here we have seen that there are two types of graphs that we can have, both of them are defined in terms of vertices and edges. In the first case, if the edge relation is not symmetric then we specifically record the direction and we call it a directed graph. If the edge relation is symmetric and this happens very often, so it is useful to think of this as a separate case, of course, we could always represent this by having edges in both directions, nothing to stop us from creating an edge saying that Sheila is a friend of Badri and Badri is a friend of Sheila and having an edge going from Sheila to Badri and one going back, but is much more convenient to just draw a single edge with no arrows indicating that this is symmetric. So, since this is an important special case, this is usually treated separately in graphs and it is called an undirected graph.

(Refer Slide Time: 5:58)



So, what can we do with this graph, other than just visualizing the pictorial relationship between the people? Supposing, we have a situation where Priya needs some help and it turns out that actually Radhika is in a position to provide this help, but as you can see from the graph, there is no direct connection between Priya and Radhika, so Priya and Radhika are not friends, so Priya may not even be aware of the fact that Radhika can be a source of help.

So, what do we do in real life? In real life when we have a problem we reach out to our friends and we say I have a problem do you know somebody who could help me. So, in this case Priya could reach out to her friends who are in this particular graph Aziz and Feroze and then one of them, presumably can reach out to their friends or both of them and so on and eventually somebody will hit upon Radhika. So, one possible scenario is that Priya told Aziz about her problem, Aziz told Badri about this problem that Priya has and Badri says 'Oh, I know that Radhika can solve this problem, so let me put Priya in touch with Radhika'.

So, what we have constructed through the friend relation is a path. It is a sequence, connecting Priya to Radhika even though there is no direct relationship between Priya and Radhika. On the other hand, if Priya had asked Feroze, then Feroze might have propagated this question to Kumar and independently Kumar is also a friend of Radhika, so Kumar would have found out the same thing and told Feroze for why does not Priya contact Radhika. So, this is something that you can do once you have the graphical representation of the relationship, you can look for these long-distance connections, which are called paths.

So formally in a graph, a path is a sequence of vertices, you have a starting vertex, say v1 and an ending vertex say vk and what you want to do is go from v1 to vk by following a sequence of connected edges. So, v1, v2 should be an edge. So, in this case v2 will be a friend of v1, but then v2, v3 is also an edge, so v3 is a friend of v2 and so on. So, you have k minus 1 edge connecting these k vertices, so you can go from 1 to k following these edges and this is called a path.

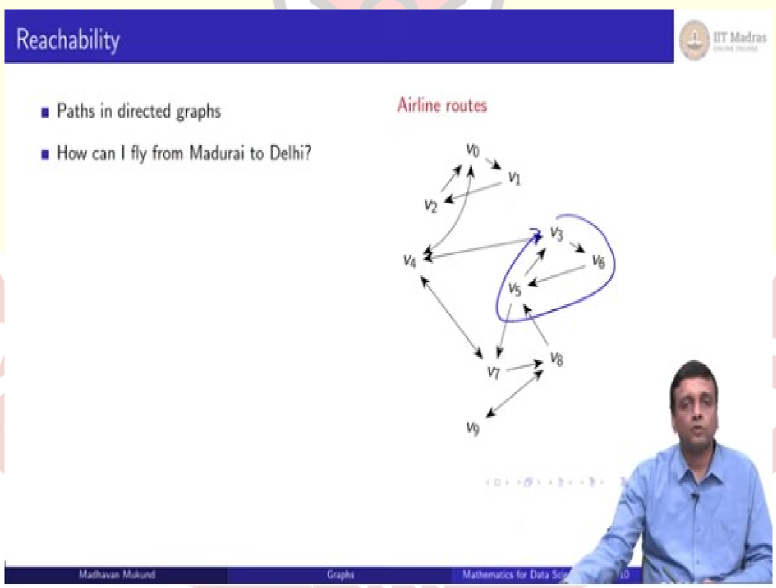
So, one thing is that there is no description in the previous definition as to whether vi in the path can be the same as a later vj in the path. In other words, you go to someplace in the graph and then you come back to that place and then proceed. Now, of course, you can imagine that this is never necessary.

So here is an example. I want to connect Kumar to Sheila but instead of going directly as I would here, right, so this would be a direct connection, so let me draw that in a different color maybe so instead of drawing a direct connection from Kumar to Sheila, I actually took this roundabout route of going around and going. So technically, in graph theory, this is not a path.

So, path should not repeat a vertex, if you have a sequence which starts at the vertex and ends with another vertex and possibly repeats vertices along the way, the graph theoretic for that is a walk, so walk is a more general type of path, in a path we usually assume that there are no repeated vertices and if there are no vertices and there are only some n vertices in the graph, every graph has a finite number of vertices and usually we call it n as the number, then clearly a path cannot have more than n vertices because if I have n plus 1 vertices in my path some vertices shall repeat.

So, this means also in terms of edges the longest accurate path that is path in the strict sense can have at most n minus 1 edges. So, we have at most n minus 1 edges in a path, where the size of v is n.

(Refer Slide Time: 9:44)



So, that was an example that we did in the friend graph, but of course, you can also do paths in directed graphs. So, let us not look at that previous directed graph which is a bit boring because there is no way to go from a teacher to a course and then go anywhere, you just get stuck. So, a more common, directed graph is something that represent say transportation network.

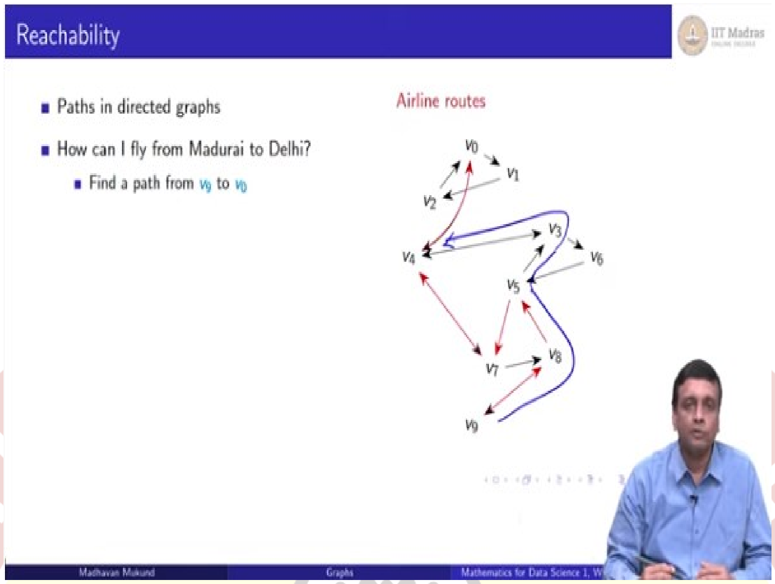
If you have ever looked at an airline map or railway map, you might find graphical representations of the routes that the airline or the railway services. So, for instance, here this might represent an abstract picture of 10 cities which are served by some airline. So, if we assume that this is super impose roughly on a map of India then v0 somewhere in the north, so let us assume maybe that v_0 represents Delhi and v9 the tenth vertex in this graph maybe represent some city in the south, let us say it is Madurai.

Now, there are arrows in this graph, indicating that not all flights go between both cities in both directions, typically between large cities like say between Delhi and Mumbai or between Bangalore and Chennai if you have a flight going in one direction by an airline you also have a flight in the reverse direction, so you can assume that if you can go in one direction, you can also go in another direction. But very often in smaller sectors airlines might operate these kinds of triangular routes, right.

So there is a route, which serves three cities, but it does not go back and forth between each pair of cities, it starts in one city goes to the second one, goes to the third one and returns to the first. So, for instance, in this case if you want to go from v3 to v5 you have to sit through a halt in v6.

So, now given this graph, the same question that we asked before about how Priya would discover that Radhika can help her. In this case we might ask a more direct and natural question saying that I am in Madurai in v9 and is it possible on this airline to travel from v9 to Delhi which is v_0?

(Refer Slide Time: 11:43)

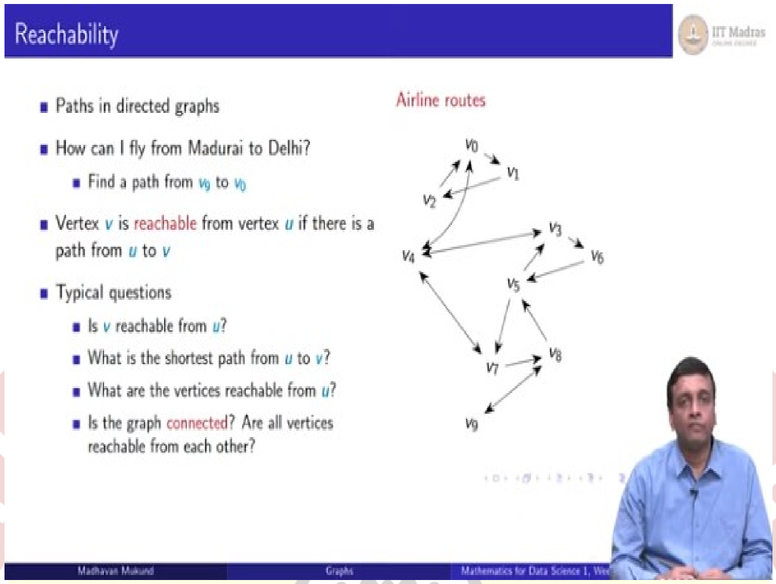


So, I need to find a path and of course, if you look at this picture here is a possible path, I can go from v_9 to v_8 and then v_5 and then to v_7 and then back to v_4 and then up to v_0. So, notice that some of these edges we drew with two arrows like this and this. So, this indicates that it is a directed graph, but we explicitly have an edge in both directions. We have an edge from v_4 to v_0 and we have an edge v_0 to v_4 where as these where there is only one arrow, one arrowhead is a unidirectional edge, that is I can go from v_3 to v_6 but I cannot go back from v_6 to v_3 in this graph.

Now, this is not the only way to go, obviously, so we could for instance, instead of doing this, we could, for instance, at the v_5 we could have followed this path up to v_5, up to there we have no choice because from v_9 we can only go to v_8 and from v_8 we can only go back to v_9 which is not very useful or we can go on to v5 but at v5 there is an option to go to v_3, right and v_3 is connected both ways to v_4, so I can come back to v4 so this is another way to reach v_4.

So, there are two ways to get to the v_4 either via v_3 or via v_7 both of them have roughly the same number of, same number of cities in between so there is not much advantage, but there are multiple paths and we can discover them.

(Refer Slide Time: 13:02)



So, in graph theory what we say is that a vertex is reachable from another vertex, if we can find a path, so we say that v is reachable from u, if there is a path from u to v. So, some of the questions that we might be interested in a graph, of course, the first question is, is a vertex v reachable from u?

This is the kind of question that we asked just now about Madurai and Delhi or about Priya and Radhika, so if Radhika can help, is there a way that Priya can find out about it through her friends networks, so this is a reachability question for a specific pair of vertices.

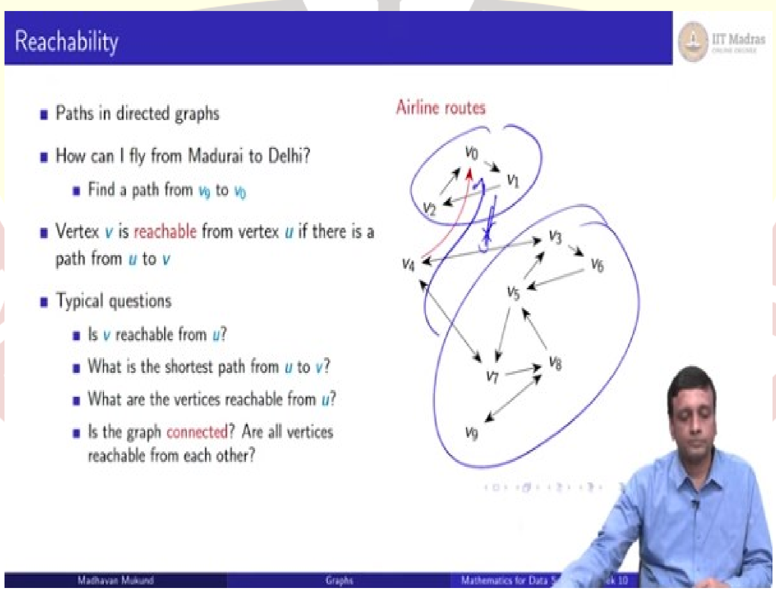
Now, given that this is possible, you might still want to find out the best possible way to do it in terms of say the shortest number of flights. So, when you log into something like MakeMyTrip or any of these travel websites, it will offer you a number of flights direct flight or one hop flight, a two-hop flight and you might prefer to go by a flight which has fewer stops, so that you do not have to waste your time waiting while the plane is on the ground.

So, we might ask for the shortest path, now shortest path for us right now is just in terms of number of edges or number of intermediate vertices, but later on we will see that we could also associate some kind of distance or time with each leg and then we could ask for the shortest path not in terms of the number of hops but in terms of some quantity that we are measuring as we travel, say the time or the distance.

Now, we could also ask a more general question which is that if I started u were all can I reach? So, in particular, if I know this, if I know where all I can reach, then I can answer the first question because if v is one of those vertices that I can reach, then v is reachable from u, so this is a more general questions then asking whether a specific vertex is reachable is asking where all can I go from a given starting vertex.

And then we can ask more general questions about the graph as a whole, so is the graph connected, can I go from everywhere to everywhere? Now, in an undirected graph this basically means that if I start at one vertex, I can reach every other vertex, in a directed graph is a little more complicated, I may be able to go from one vertex to every other vertex but I may not be able to come back. So, I need to go from every vertex to every other vertex.

(Refer Slide Time: 15:15)



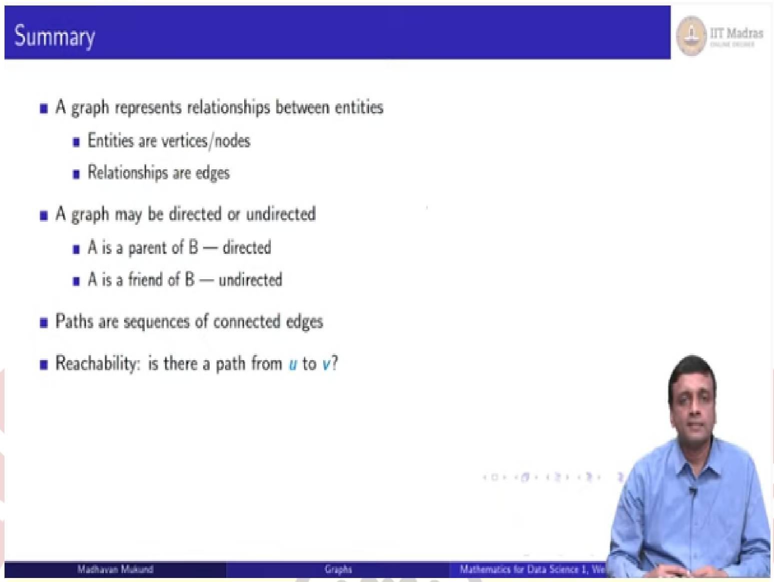
So, let at look at this graph for instance. Supposing, we take this v4 to v3 flight and make it to one directional flight. Now, is it still possible to go? In the earlier graph actually you can check that you can go from everywhere to everywhere, because the crucial edges so we have this section where you can go from everywhere to everywhere, you have this section which you can go from everywhere to everywhere, you have this section where you can go from everywhere to everywhere and all these sections are connected by these bi directional things, so from each of these components you can go to every other component via v4 and then within that component, you can go around in a circle of three and reach any place you want.

So, this original graph is definitely connected, now what happens if you break this bi directional connectivity to v_4 by saying by saying you can only go from v_4 to v_3? You cannot go back. Now, earlier we could take this as an escape route from v_3 to go from this triangle to any other triangle but now we can still go down from v_5 to v_7 and then proceed.

So, even if we make v_4 to v_3 a one directional edge, there is no problem, this graph is connected in the sense that from every city we can reach every other city. However, if we take another edge from v_4, say v_4 to v_0 and make that a single direction then we have an issue, because go we can go from everywhere here, we can go up through that edge, if we are at the top, we cannot come back down, right there is no way to leave that v_0, v_1, v_2 triangle, because there are no edges leaving them. So, now the graph has become not connected in one specific way which is the vertices v_0, v_1, v_2 cannot reach the other vertices.

Everything else can reach, so notice that is a symmetric, the other vertices can reach. So, we are not solving these questions, we are just posing these as typical questions that you might want to ask once you have a graph presented to you as a representation or some information so that information, we started with the motivation that it came from a relation, but it could also be some natural information like this like friendships, or it could be like airline routes.

(Refer Slide Time: 17:19)



So, to summarize a graph represents a relationship between entities, so in the graph these entities are represented as nodes or vertices, and the relationship that we are trying to capture is represented by edges between these and these edges might be directed or undirected. So as a directed graph, we saw one example, which was the airline route. Another example involving people can be family relationships. For example, if you have a group of people and you want to connect people who are parent child. So, if A is a parent of B, you want to say that A is related to B.

Now, clearly this is asymmetric if A is parent of B, then there is no way that B can be a parent of A. So, this could be a graph that you might have seen pictorially in the form of family trees, so people sometimes represent relationships within the family, in terms of a graph where they draw edges between parents and children and then they have a way of connecting people who are married together and all that, but a family tree is a kind of a graph and that would be typically directed because parents' child is asymmetric relation.

On the other hand, as we saw if you have a friend relation it becomes an undirected graph. So, we have these two fundamental types of graphs and the problem that we have looked at right now which is of interest is within the graph to identify paths and through paths, talk about reachability and connectedness. So, we will explore these problems more systematically in the lectures to come.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

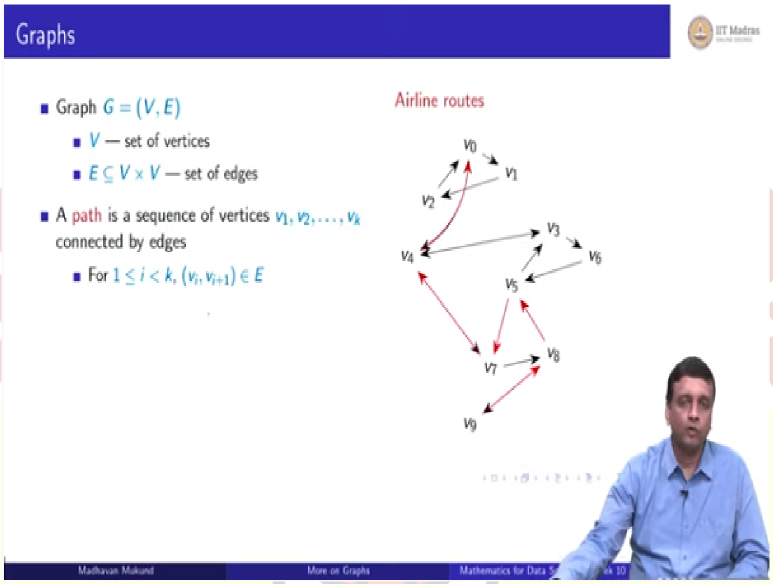
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science
Professor Madhavan Mukund
Indian Institute of Technology, Madras
Lecture 60
Some General Graph Problems

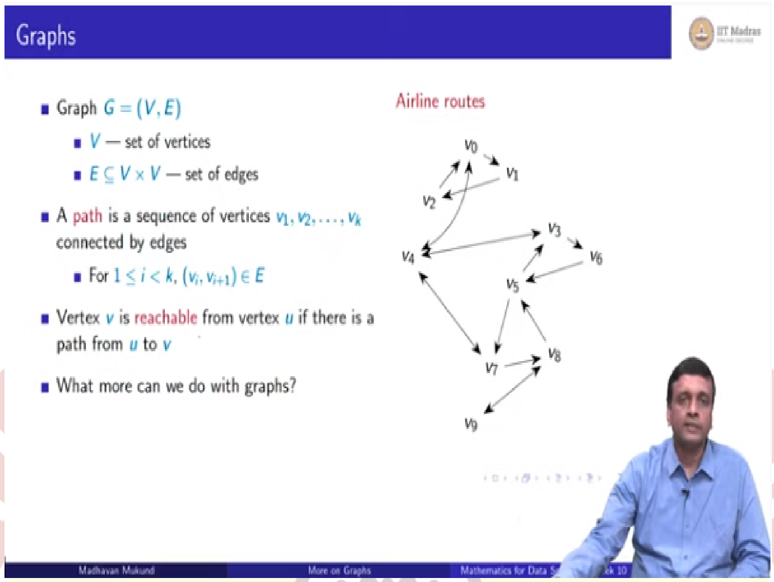
(Refer Slide Time: 0:16)



So, in our first lecture we introduced the concept of a graph, so we said that a graph consists of a set of vertices and a set of edges, so the edges are just pairs of vertices, so edge relation is a subset of v×v. So, for example, we had this directed graph on the right which represents airline roots.

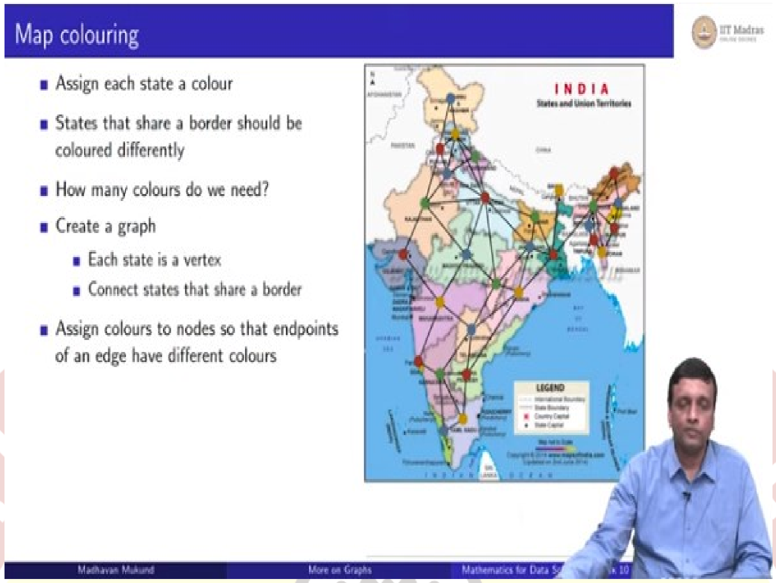
And then we said that a path in a graph is a sequence of edges leading from one vertex to another vertex without any vertex being repeated in between. So, here we see a path from v_9 to v_0, so each edge must be an extension of the previous edge, so we go from v_9 to v_8, then we go from v_8 to v_5 and so on, right.

(Refer Slide Time: 0:55)



And then we talked about reachability saying that we might want to ask whether a vertex u, starting from a vertex u we can reach a vertex v by finding a path. So, at this point the only problem that we have really looked at in graphs is reachability and this is really one problem which we will spend some time on but before we get into more details about how to calculate reachability in a graph. I would like to show you that graphs have much more interesting problems than just reachability associated to them. So, having a graph representation of a problem allows you to deal with very many different scenarios.

(Refer Slide Time: 1:35)



So, let us start with a problem which does not appear to be connected to graphs at all, that is how to color a map. So, typically when you see a political map of the world or of a country like India you will see that each political unit has a different color, but not all have the same have different colors for instance you might find a color that is repeated like you can see in this map for instance that some colors like light blue and green are repeated in different places.

So, the rule is that normally when you color some state in the map or a country in a map it must have a different color from all the countries or states which share a border, so there is no confusion at the border because one side is colored one way and the other side is colored the other way.

So, this is the rule for map coloring, so one question that you might ask and at the moment it seems like an ideal mathematical question is how many colors do we need? Now, clearly if I have a certain number of states, I can use a different color for every state, so I have an upper bound, right, if I have say 27 states then I need 27 colors, if I have 27 colors each of them will get a different color there is no question of two states sharing a boundary having the same color because no two states are the same color.

But maybe I do not need 27 colors, can I do better than that? So, here is where a graph comes in. So, how do we create a graph to represent this problem and what is the problem that we are trying to solve on the graph. So, to create a graph what we do is create these vertices and what are our vertices in this case? The vertices are the states, okay or if it is a map of the world or the countries.

Now, what is the edge relation? The edge relation is going to represent when two states share a border, when they are neighbors and must be colored differently. So, we connect all states which share a border and then we get these black lines connecting these black dots. So, for every state there is a black dot and every state is connected to its neighboring state black dots, this is our underlying graph.

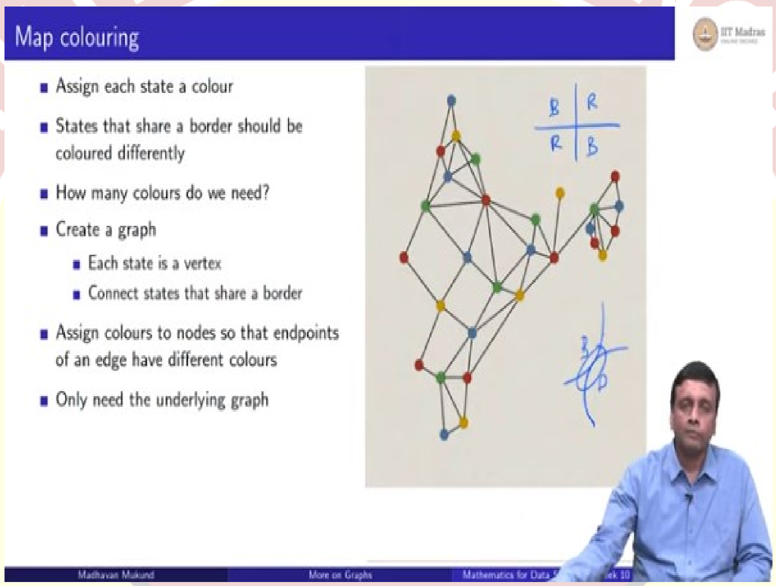
Now, our goal is to associate a color with every state on the map which is the same as associating a color with every black dot in this graph. So, we start maybe by assigning a color in this case red to Uttar Pradesh, now the rule for map coloring tells us that if Uttar Pradesh is red then all the neighboring states must be a different color other than red. So, anything which is connected to Uttar Pradesh in the graph, any node that is connected to Uttar Pradesh must have a different color.

So, we start with this color for Uttar Pradesh and then we can start coloring its neighbors, so for instance we might choose a different color green for Uttarakhand and we might choose blue in this case for Haryana. So, proceeding in this way we go to the neighbors of these and we use a different color but notice now that once we go from Haryana to Rajasthan, we can reuse the color green because green has been used for Uttarakhand and Uttarakhand and Haryana and Rajasthan do not share a border, so there is no confusion, so we could reuse a color if it is not being used for any of the neighboring states.

So, we proceed in this way so we could again now reuse red for Punjab because Punjab is not connected to Uttar Pradesh, remember Uttar Pradesh was originally red, so we could not take any neighbor of Uttar Pradesh but since Uttar Pradesh was not connected to Punjab we do not have to worry, we can reuse it but finally now when we come to Himachal, we have a problem because Himachal is now surrounded by three neighbors which have already been assigned three different colors, Punjab is red, Haryana is blue and Uttarakhand is green, so we have to choose a fourth color, say yellow for this.

So, we keep proceeding in this way and we can now whenever we need a new color, we use it, whenever we can reuse a color, we reuse it. So, we can come up with a more say you know less expensive coloring in terms of number of colors for this, by coloring these nodes in this way.

(Refer Slide Time: 5:51)



Now, notice that we do not really need that map anymore once we have constructed this graph which describes the connectivity of the states in terms of which ones share a border we could always as well start coloring the graph using this rule, that if I color a node with one color, I cannot color any of the neighboring nodes, any edges connected to it must lead to different colors.

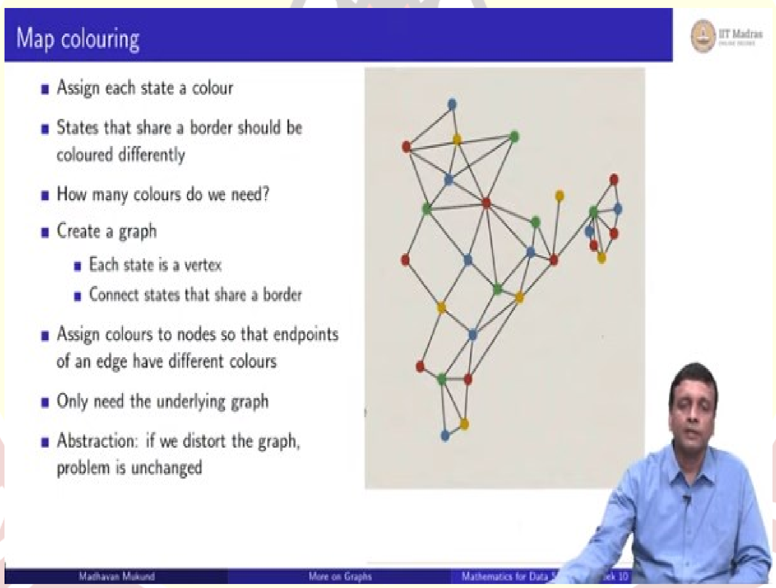
Now, here you can see one advantage which is when we are working with the physical map we have to stare at the borders and it depends on how well the map has been drawn to be able to distinguish because sometimes we have borders which meet at a corner and technically across a corner this coloring rule does not apply.

For example, if four states which happens actually in the United States, if four states meet like this. Then I can actually use the same color, I can use say red, red, blue, blue this is legal as far as map coloring goes. So, if two states or two countries touch only at one point then they are not considered to be sharing a body.

So, this might depend on how the border is drawn maybe that is the picture that you see but actually if you blow it up it actually looks like this, and then I have a problem, because now I cannot use blue, blue because there is a segment of border which is common to these two states.

But once I have transferred this information to the graph then I do not have this confusion anymore and in fact I can take this graph and I can even distort it, right. So, this is the original graph somewhat faithfully representing the geometry of the underlying country.

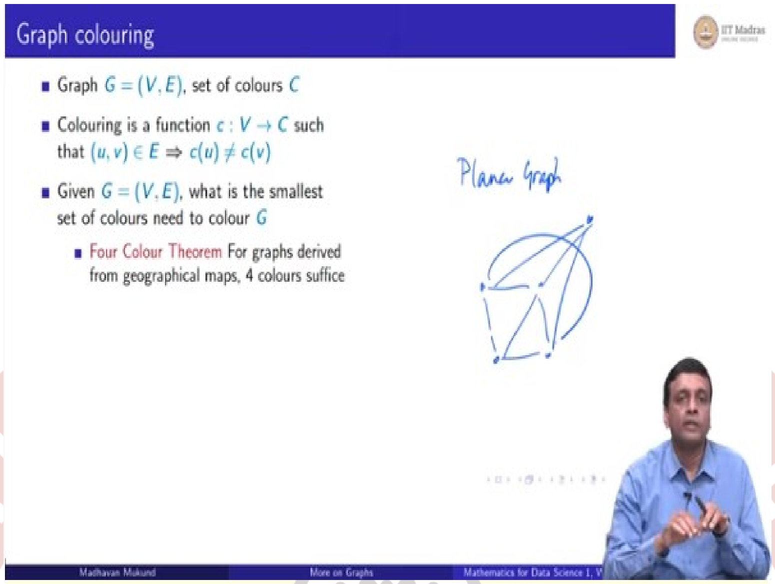
(Refer Slide Time: 7:16)



And now I can take some of these nodes which are bunched up and move them far apart so that I can draw my coloring better. So, this is one advantage of moving to the graph which is that the graph abstractly represents the relevant information, so we can now work with whatever format of that information is convenient in the graph without worrying about the original format in which the information came.

So, here in this case the geography or the geometry of the actual state boundaries is no longer important, we just need the connectivity saying which states are neighbors of which state.

(Refer Slide Time: 7:50)



So, abstractly we have transformed our map coloring problem to what is called a graph coloring problem, so in a graph coloring problem we have a graph which consists of some vertices and edges as before and separately we have a set of colors and we want to do a coloring, so what is the coloring?

A coloring is a function which assigns to every vertex a color from the set C and the rule is that if I have a pair of vertices connected by an edge their end points should have different colors. So, u, v is an edge, the color of u should be different from the color of v, this is what graph coloring demands.

And the question that we were asking is given a set a particular graph, if I do not fix the set of colors in advance, like we were doing there when we did our example, we started with one color red and then we were forced to we chose another color green and then we were forced to we choose another color blue, then when we were forced to we chose a fourth color yellow and so on.

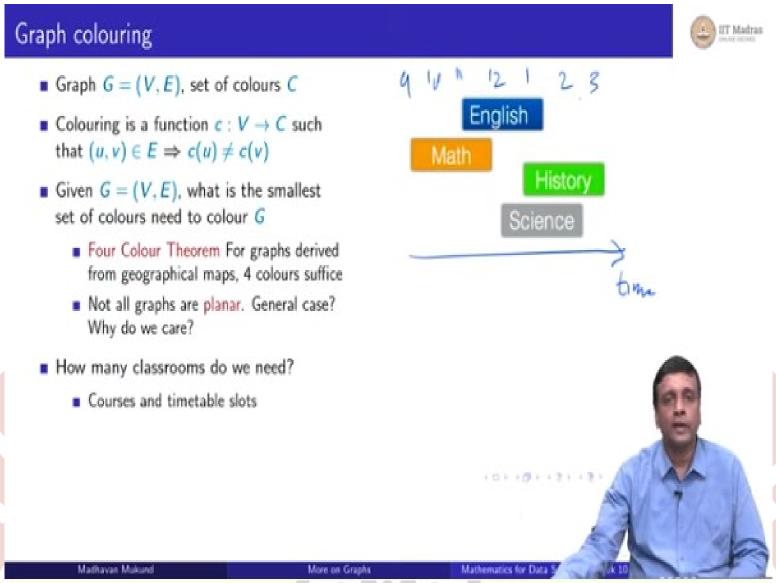
So, if I add colors only as I need them how many colors will I need, right, what is the minimum number of colors I need for this specific graph? So, it turns out that this problem has actually been well studied for these graphs which come from maps. So, there is a very well known theorem which is very hard to prove called the Four Color Theorem which says that for graphs which are derived in the way we showed from geographical maps, four colors suffice. So, technically these are what are called planar graphs. So, planar graph is something where if I draw the edges, right, they will not cross, so this is a planar graph.

If I draw this edge for instance, then these two edges are crossing but this is not necessarily a problem because I can actually take this edge is crossing, right and I can actually draw it around. So, this is still a planar graph but now if I try to connect for instance some a third thing which is outside here and I try to connect it across these then I will have a problem, right. So, some graphs cannot be drawn on a sheet of paper without edges crossing and these are called non-planar graphs.

Now, it turns out when you have a map laid out and you start connecting them obviously the map cannot, you cannot have share a border with something which is far away, so therefore, a map will always be a planar graph, right. So, not all graphs are planar, so the question that graph coloring asks is the general case, right.

Now, from our perspective the question is why do we care, I mean map coloring itself seems to be maybe a particularly specialized application where we want to look at this coloring problem and translate it to graphs but why should we care in general, how many colors we need for a graph? So, it turns out that graph coloring actually is very useful in a number of different cases.

(Refer Slide Time: 10:39)

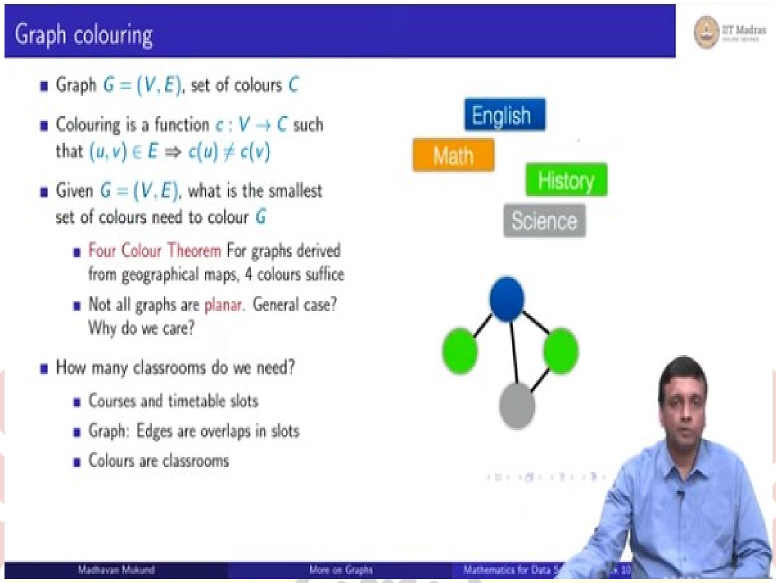


So, one typical case is in classroom scheduling, so supposing we are running a school and we need to determine how many classrooms we need in order to run our classes. Now, this depends on the time table, right. So, we have some time tables and we have some courses and let us assume that this is a graph, not a chart representing, so this is the time of the day, right. So, this might be like 9 o'clock, 10 o'clock, 11 o'clock, 12 o'clock, 1 pm, 2 pm, 3 pm. So, across the day we have these different lecture slots and we have different courses which occupy different slots in our time table.

Now, clearly if maths is running from 9 to 12 and English starts at 11, then English must be in a different classroom from maths. Similarly, if history is running from 1 to 3 and science already started at 12 and goes on till 2, history and science cannot be in the same classroom. So, if we have overlapping slots then the corresponding classes need different classrooms.

So, the question is what is the minimum number of classrooms I need in order to run all these classes without having any scheduling conflicts? Now, like before we could assign a separate classroom, we could have a separate English classroom, we can have a separate math classroom, we can have a separate history classroom, a separate science classroom and the problem is solved but we want to optimize, we do not need to want to necessarily allocate a separate class, for every classroom for every course.

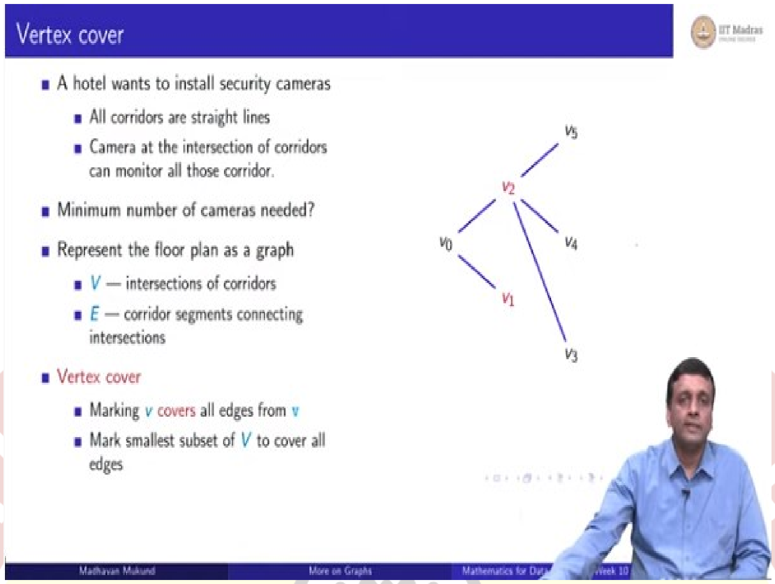
(Refer Slide Time: 12:01)



So, as before now we can draw a graph, so in this graph we have nodes which are our courses and now the edge relation represents overlaps. So, an overlap says that these two courses share a time slot and therefore, they cannot be both scheduled in the same classroom. So, for us now colors are classrooms, so here is a situation that we have four different colors assigned to these four different nodes saying that we are going to have every class running in a different classroom.

But now we observe that maths and history do not overlap. So, since maths and history do not overlap, using graph coloring we can see that the same color can be assigned to both the maths node and the history node. So, in this way many scheduling problems can be actually converted to graph coloring problems.

(Refer Slide Time: 12:50)



Now, here is another problem. Supposing a hotel wants to install security cameras, so they want to put off cameras in the corridors of a hotel and as you know in many hotels corridors are very neatly aligned, they are all straight lines but there might be a maze of corridors which meets at different corners. So, let us assume that if you put a security camera at a particular corner, it can monitor every corridor that meets at that intersection.

So, now the question is what is the minimum number of cameras that you need to monitor all the corridors on the floor? So, once again we can go to graph theory, so we can represent the floor plan of the hotel for that floor as a graph, so here the points of interest are these intersections because clearly it is to our advantage to put a camera at an intersection because if I put it at an intersection that camera can monitor multiple corridors and I want to cover all the corridors.

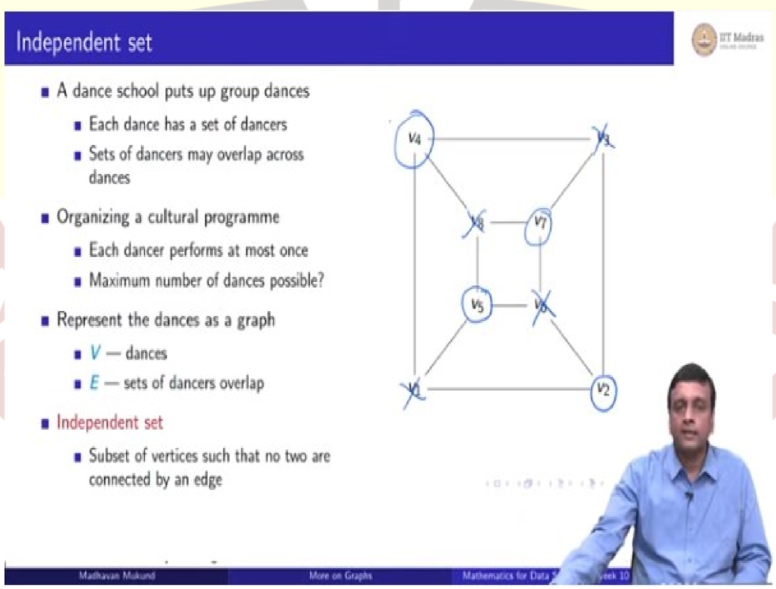
So, I put vertices and intersections and my edges now connect these intersections, there are segments of the corridor which run from one intersection to another intersection. And now my question is one of called, something called a vertex cover which is that I want to choose a subset of these intersections, such that if I put a camera at this subset then every corridor segment is covered. So, in graph theory this is called a vertex cover question, so vertex cover basically says that I want to choose a subset of vertices such that if I choose that subset then those vertices cover all the edges in my graph.

So, let us look at this graph on the right, so it has 6 vertices named v_0 to v_5, so maybe this represents the corridors in our hotel, so maybe I choose to put a camera at v_2. So, if I choose to put a camera at v_2, then this covers these 4 corridor segments but it does not cover the segment from v_0 to v_1, so I have to put one more camera, I can choose to put it at v_0 or at v_1, so let me say I put it at v_1 and now I have a vertex cover.

So, my vertex cover is v_1 and v_2, if I choose v_1 and v_2 as my locations for my cameras it covers all the corridors, so this is a problem in graph theory which you can solve independent of the source. So, this is one motivation but you can come up with other situations where the solution that you require is a vertex cover.

A similar situation could be for instance if you want to locate ambulances at intersections, so that they can reach all localities fast, so if you want to cover all localities with ambulances, where should you place the ambulances in your city map so that every locality is served within a reasonable amount of time.

(Refer Slide Time: 15:33)



Here is yet another problem. So, supposing there is a school, a famous school of dance and they are going to put up a show which consists of a number of group dances, so in a group dance obviously there are a number of dancers who participate but the school as a whole has over a period of time rehearsed many such dances and some dancers take part in more than one dance. So, if I look at all the dances that the school could possibly put up there are overlaps between the dances in terms of which dancers are required.

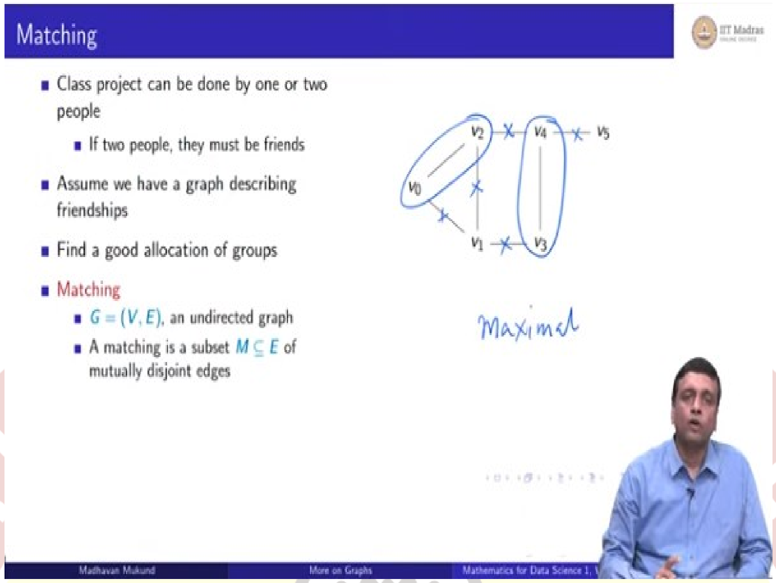
So, now the problem is to organize a cultural program and in this cultural program because of costume changes and other constraints we would like each dancer to participate in at most one dance, so we do not want a dancer to take part in a dance and then to go back have to change and come back and take part in another dance with a different costume. So, given that we have some information about the dances and which dancers are required for each dance, can we come up with a large set of dancer which we can put in this cultural program, so that no dancer has to dance twice during that program.

So, in this particular case the graph will consist of vertices which represent the dances and now an edge will represent an overlap in terms of the dance group between two dances, so if two dances share a common dancer, then we cannot put both dances in the program because one dancer would have to dance in both dances and that is not allowed by the rules that we have just stated. So, this is our graph now and now we want to find what is called an independent set, so we want to find a set of vertices such that there is no edge between any two vertices in that set.

So, we want to pick a set of dances so that no two dances in the set that we have chosen for the program share a dancer and therefore, require a dancer to dance twice. So, here is an example with 8 nodes, so now supposing we pick v_2 to be in our independent set, right, then because I cannot use anything which is connected to it, it means that v_6 cannot be part of my independent set, I can no longer use dance in v_6, I cannot use the dance in v_3, I cannot use the dance in v_1. So, what can I do? Maybe I can choose v_5, now v_5 already rules out v_6 and v_1 which I have already gone but it also rules out v8 but I can now do v7 which has no further constraints and finally I can do v_4.

So, in this particular scenario I could choose actually four of these vertices such that there is no edge between any of them, so this is what is called an independent set, right. So, the independent set here, one independent set is v_2, v_5, v_7, v_4 of course, there is a symmetric independent set I can treat the black edges also, black vertices also as an independent set and so on.

(Refer Slide Time: 18:31)



So, final example of this kinds of problems that we can do with graphs, let us look at a problem which is called matching. So, supposing we are assigning class projects and the teacher allows the project to be done by either one person, student individually or by two people but there is a constraint that if two people participate in a project then they must be friends because if they do not get along with each other then the project would not get done.

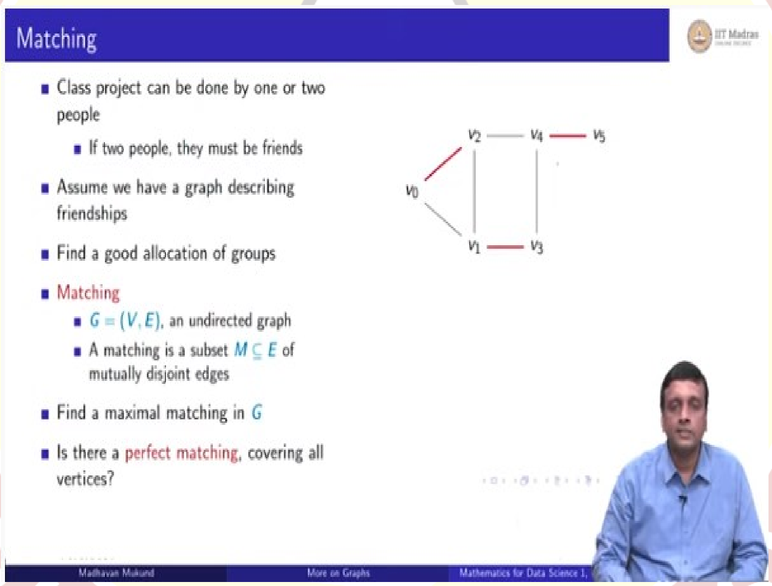
So, let us assume that like we had before we have a graph which describes friendships, what we want to do is given this graph of friends, we know who's friends with whom, we want to find a good allocation of groups, right, we want to find pairs but these must be pairs. So, if I have three people who are all friends of each other a, b and c are all friends of each other, if I make a, b a pair then b cannot be a partner of c, c has to find a different partner, c cannot partner with a, c cannot partner with b, right. So, this is what is called a matching.

So, a matching is a subset of edges which is mutually disjoint, that is if I pick one edge and I pick another edge they do not share any vertex, right. So, here for instance if I pick this edge, right then I cannot pick the edges here, here or here because all of them either touch v_0 or they touch v_2, so I can touch, I can pick any of these three, okay. So, this is the problem that we have and this is called a matching.

So, for instance you might ask for what is called a maximal matching? So, maximal matching like we started this one for instance, right. So, at this point when I have done, when I have chosen this one, right I rule out some vertices but there are some edges but there are still some edges which are permitted, so I can pick one of them and create one more pair. So, for instance I might pick this one but now this rules out this vertex, this edge and this rules out this edge also because now those edges have common vertices v3 and v4.

So, now at this point all the edges have been ruled out or included and I cannot proceed, so this is what I call a maximal matching, right. A maximal matching is one which cannot be extended by adding any more pair without violating some condition.

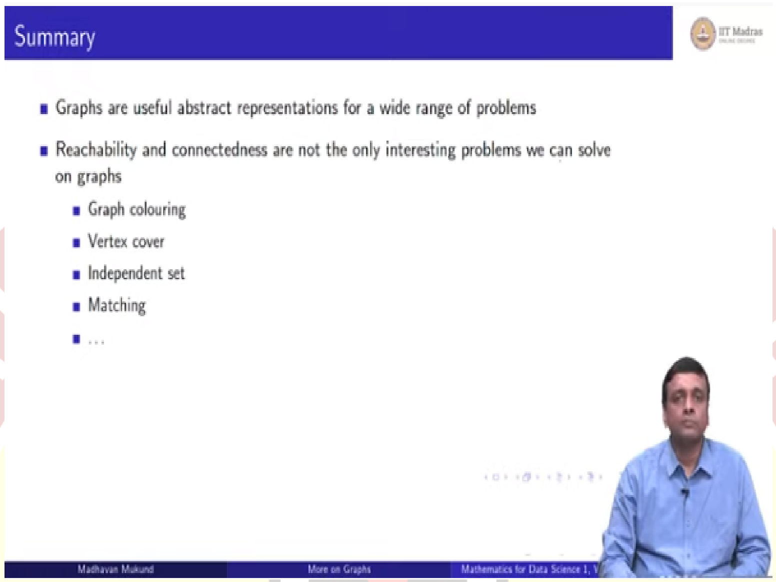
(Refer Slide Time: 20:48)



So, here for instance I what I drew as a maximal matching but here is another maximal matching, right. If I take v_0, v_1 then this has knocked off these edges, right and now if I take v_2, v_4 it has knocked off these edges because v4 is connected to both of them. So, I cannot add any more edges and I am stuck making only two pairs among these six students. Now, ideally if there were 6 students, I should hopefully be able to make three pairs and that is what we call a perfect matching.

So, perfect matching is one which is a matching but also connects every vertex in the graph, so every vertex is part of some pair. So, here is a perfect matching on this graph there are six edges, I mean six vertices and there are three edges and these three edges are mutually disjoint. So, this is the third type of, fourth type of problem that we can see.

(Refer Slide Time: 21:35)



So, what we want to emphasize is that graphs are not just about connectivity and reachability. So, reachability and connectivity are actually very important problems in graphs but that is not the beginning and end of graphs, there are very many interesting problems that you can frame once you put your problem in a graph theoretic sense.

So, we saw graph coloring which we saw an example with scheduling, we saw vertex cover where we saw an example with allocation of say security cameras, we saw this independent set problem which in our case was about having a maximum number of dances where only one dancer can only dance in one dance during the program and then we saw this matching problem of allocating groups within a class.

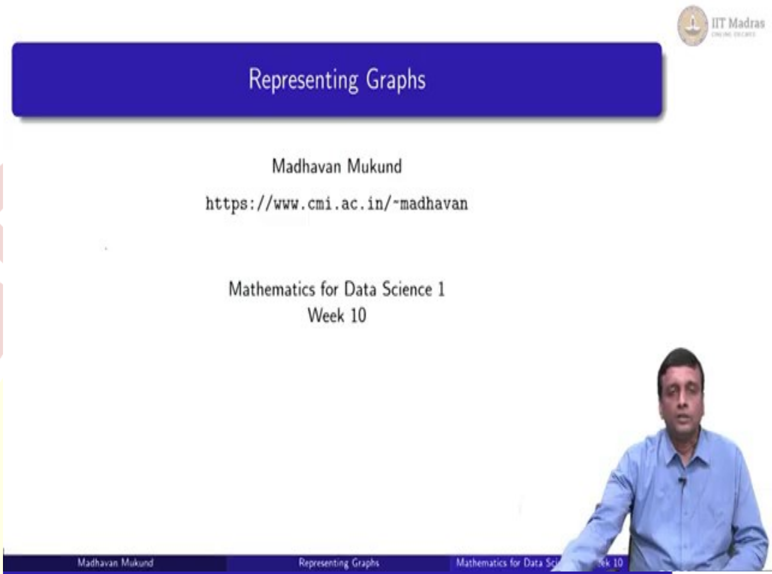
So, although we will not necessarily look at all these problems in detail in this course it is important to understand why there is so much emphasis on graphs and procedures and algorithms involving graphs because the underlying representation of a graph actually is a very rich representation and by solving problems in this abstract world of graphs you can actually solve a number of concrete problems in one shot without having to solve them individually.

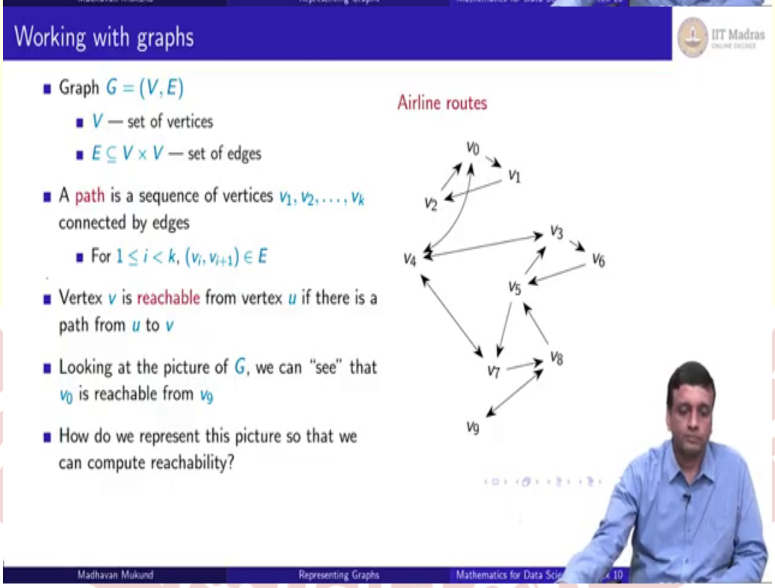
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Madhavan Mukund
Chennai Mathematical Institute
Lecture: 61
Representation of graphs

(Refer Slide Time: 00:14)



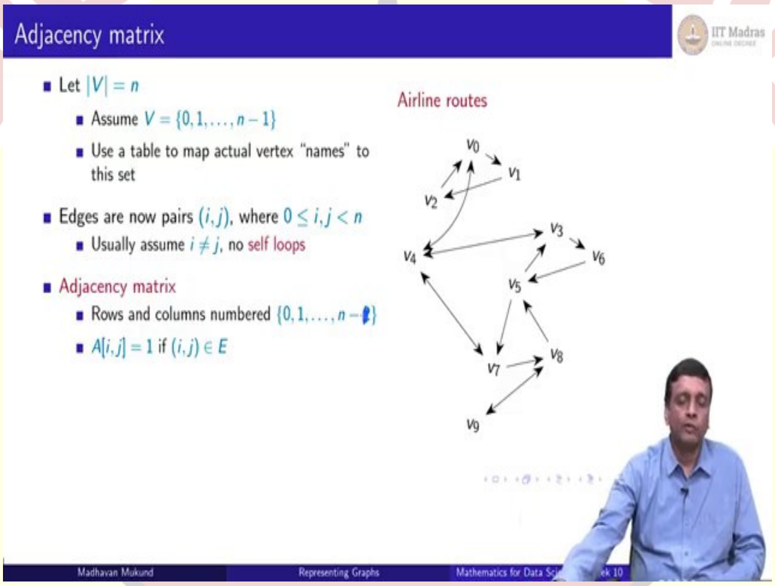


So, we have been talking about graphs and problems on graphs, we started with reachability. And then we talked about very complicated problems like graph colouring, and vertex cover, and so on. So, now let us get back to reachability, and connectivity and ask a more fundamental question, which is how do we actually work with these graphs in a mathematical setting. So, remember that a graph consists of 2 sets, or either a set and a relation, a set of vertices, and an edge relation. And let us focus on reachability back for now. So, a path in a graph is a sequence of connected edges. And we say that V is reachable from u, there is a path from u to v.

So, as humans, if we see a graph like this, then what we will do is take this picture and stare at it, and extract this graph in some sense visually. So, we will take a graph like this. And so, you can maybe by trial and error, but by exploration, we will do this. So, we can see that there is a path from V_9 to V_0.

The problem is that we want to operate on this mathematically, we do not want to have this picture because there is no way to formally represent this picture and how you operate on this picture, in terms of a procedure that we can execute. So, how do we represent this picture in a way that we can compute reachability, for example.

(Refer Slide Time: 01:30)



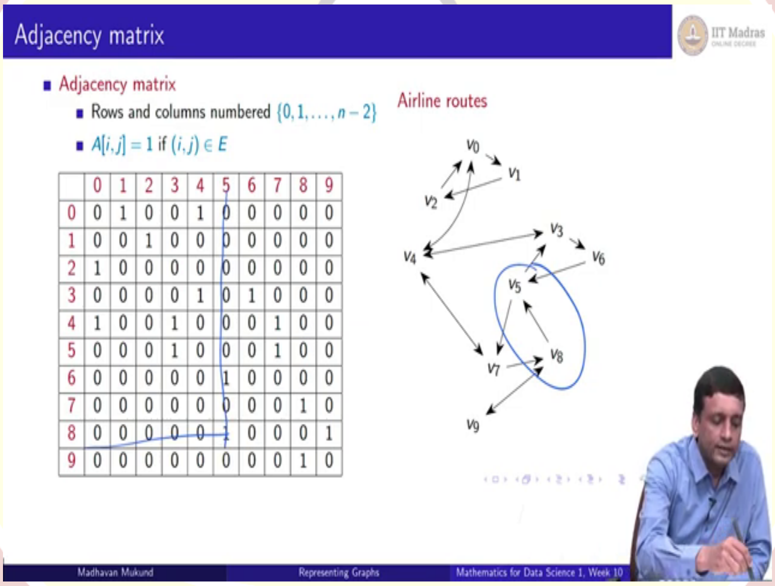
So, first, we need to represent this graph, in a way which is more amenable to computation than a picture like the one on the right. So, one way is to use what is called an Adjacency matrix. So, let us assume that the set of vertices consists of n vertices. So, this is the normal convention and graph theory that we always use small n to represent the number of vertices, and small m very often is used to represent the number of edges.

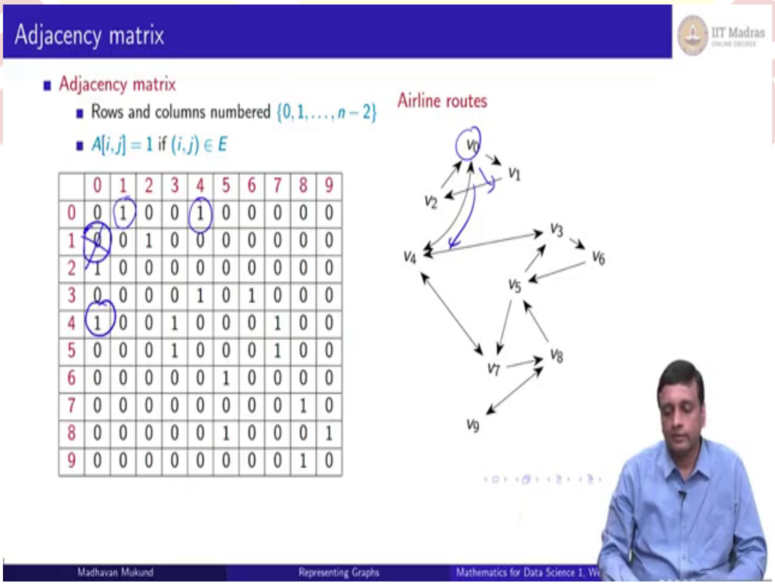
So, in particular, if we say that the vertices are n in number, then we will usually just for simplicity, number them, and call them 0 to $n - 1$, you can also call them 1 to n. But in computing, it is very common to start numbering at 0, So, we will call it 0 to n minus 1. And of course, when we actually have real vertices, like, names of cities, like in this case, Delhi and Madurai and all that, then we will actually use some kind of a table to map the actual vertex names to the set.

So, for instance, we might have a table which says a Delhi is V 0 and Madurai is V 9 and so on. So, Delhi is 0, Madurai is 9, and So, on. So, in this representation, where vertices are now numbers between 0 and $n-1$, and edge is a pair of numbers. So, an edge is a pair i,j, where i and j lie in the range 0 to n minus 1. And we usually assume that we do not have edges like this. We do not have the so-called self-loop, So, we do not have an edge from i to i.

So, we usually assume that $i \neq j$. So, when we have an edge i,j, they both are in the range 0 to n minus 1, because that is our set of vertices, but i is ≠j. So, given this, we now have what is called an adjacency matrix. An adjacency matrix, we have rows and columns numbered from 0 to n minus 1. And then you put an entry in this matrix, 1, if there is an edge from i to j, otherwise, it is 0. So, the default is 0, and you put an 1 wherever there is a matrix.

(Refer Slide Time: 03:31)





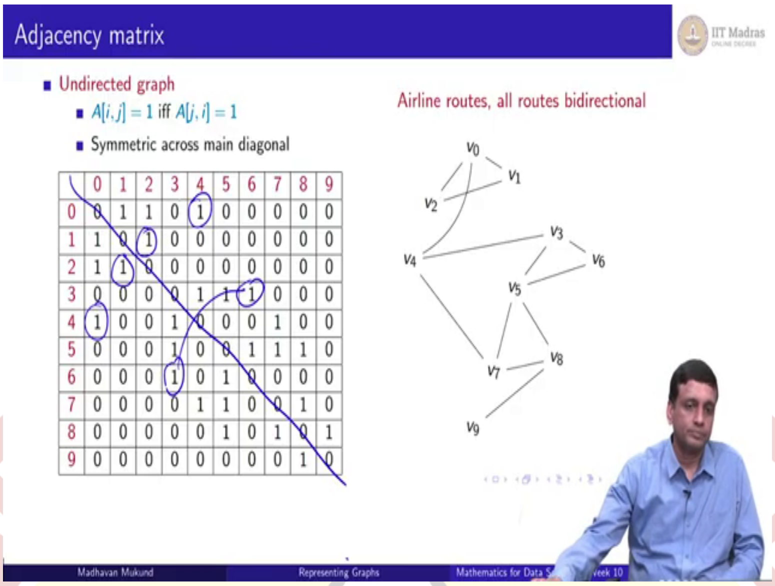
So, if we look at this graph on the right. Here is the corresponding adjacency matrix. So, first of all, because we have 10 vertices numbered 0 to 9, we have 10 rows and 10 columns, and the headers in red and the column headers and the row names on the left are in red, indicating that this is row I in column j. So, now we take an edge, say, for example, we take an edge from 8 to 5.

So, 8 to 5 says that the row 8, and the row, column 5 must be a 1. So, if I look at row 8, and I look at column 5, then I get a 1 at this position. So, that is how I fill entries in this matrix. So, you can check that this matrix represents all the edges in the graph. So, if you take any edge in this graph.

So, if you go here, for instance, and you look at this graph, and you say, 0 is connected to 1 and 4, will indeed 0 is connected to 1 and to 4. And because the edge, this bi-directional edge between 0 and 4 is actually 2 edges as we said, So, this is not an undirected graph. It is a directed graph. So, in this directed graph, we are just using a shortcut to represent it 2 edges by putting an arrow at both ends, but actually, there is a separate edge from 4 to 0.

So, there is an edge from 0 to 4, and there is an edge from 4 to 0, whereas, there is an edge from 0 to 1, but there is no corresponding edge here from 1 to 0 because there is no backward edge on that group. So, this is the adjacency matrix for this directed graph. Now, we can take the same airline route matrix.

(Refer Slide Time: 05:07)

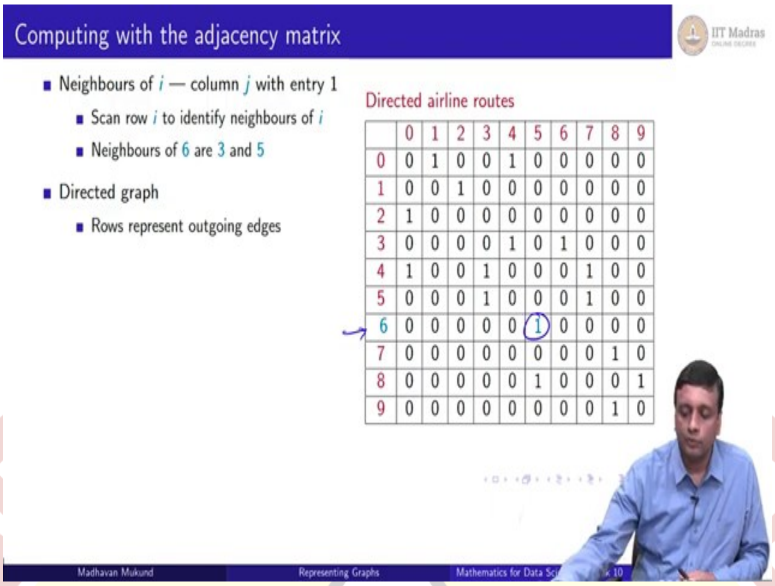


And supposing we assume that all the routes are actually bi-directional. That is, whenever the airline flies from one city to another, it also flies back. So, then we do not need to record these edges anymore. And then it is better, as we said, to take such a graph where the edges are all symmetric, and explicitly call it an undirected graph, rather than recording arrows in both directions.

So, we will draw it in this fashion where we just draw an edge as a line connecting these 2 vertices with no arrows. And now if we look at this graph, every time there is an edge from i to j, there must necessarily be an edge from j to i because it is asymmetric edge relation. And if you look at the matrix, then if you go across down this main diagonal, and then if you look at any entry like this entry here, then if we look at the symmetric entry on the other side must be the same.

If there is an edge from 6 to 3, there must be an edge from 3 to 6. Similarly, there is an edge from 2 to 1, there must be an edge from 1 to 2. So, this is now our thing, this is an edge from 0 to 4, there must be an edge 4 to 0, there must be an edge 0 to 4. So, an adjacency matrix is very simple. So, we just create a row and a column for every node in your graph. And then at the intersection of the corresponding row and a column, if there is no edge, you put a 0, if there is an edge, you put a 1, that is all there is to it.

(Refer Slide Time: 06:23)

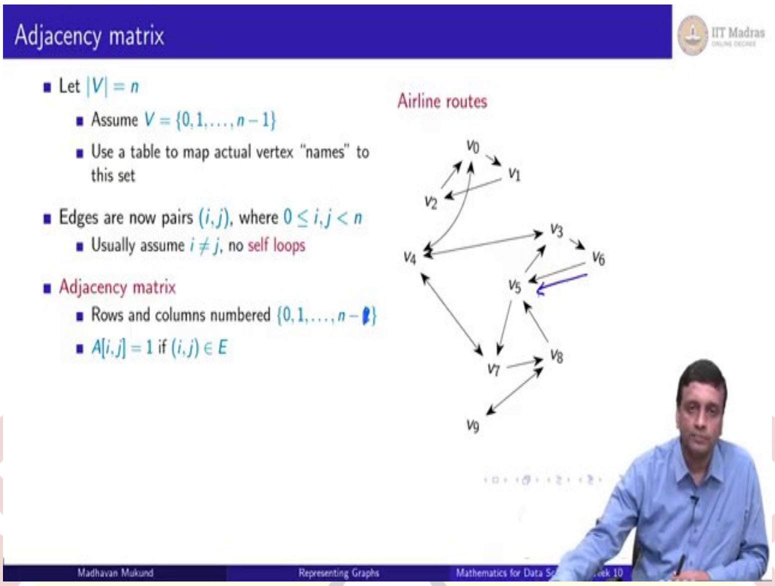


So, now we want to compute with this matrix. So, the whole purpose of doing this is now that this matrix on the right, for instance, suppose to represent the same picture as this undirected graph, So, we have thrown away the picture. And now we are looking at only the matrix. So, in this, if we look at the undirected graph, for instance, then if I want to know the neighbours of I, that is, which are the vertices which is I is connected by an edge, then I go to the row, for example, I am looking for neighbours of 6, I go to the row, where 6 is the starting point.

And then I find that there is a 1 entry in column 3 and column 5. So, this says that the neighbours of 6 are 3 and 5, which if you go up is indeed the case, the neighbours of 6 are 3 and 5. So, without looking at the picture, I can just, in some sense, mechanically analyze this table, or this matrix, and get the same information that I would get by looking at the picture. And this is what we need because there is no way that we can actually design a computational procedure, which looks at the picture and then makes decisions based on the picture.

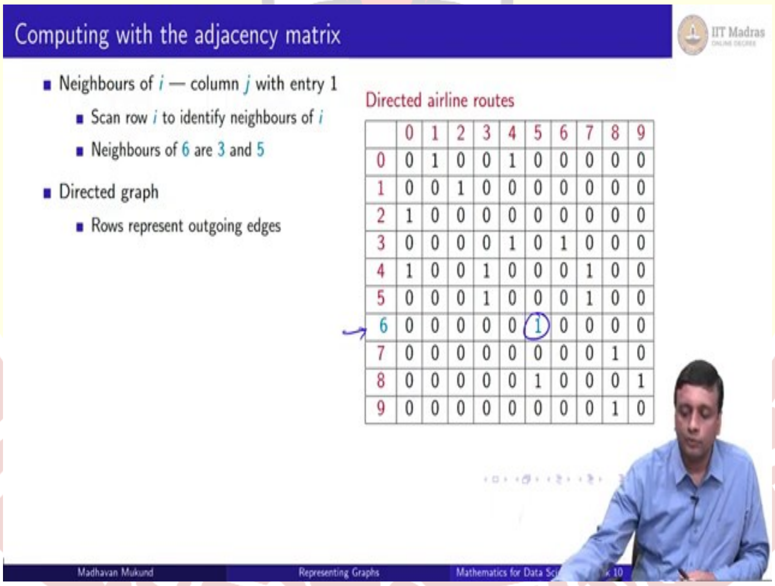
So, if you have a directed graph is slightly more complicated. Because in a directed graph, remember that we have outgoing edges and incoming edges. So, the notion of a neighbour is slightly more complex. So, we have rows which represent outgoing edges. So, if I take the row for 6, So, the row for 6 has an entry pointing to 5.

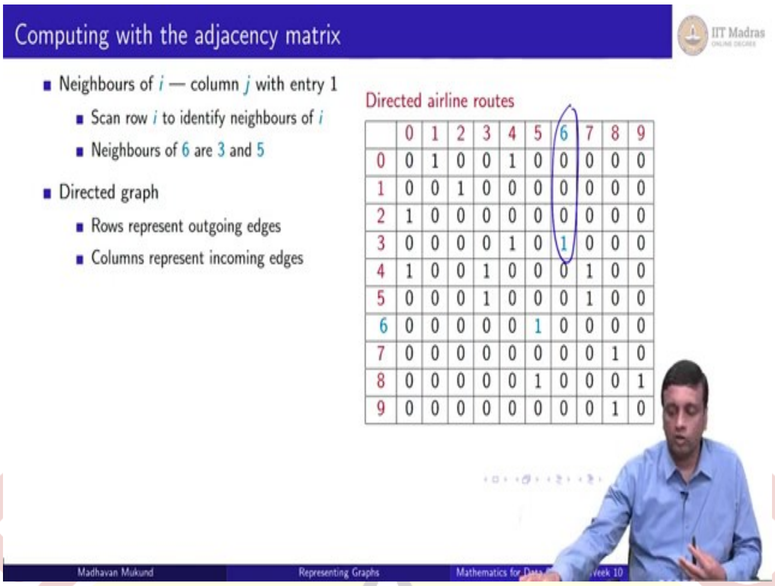
(Refer Slide Time: 07:58)

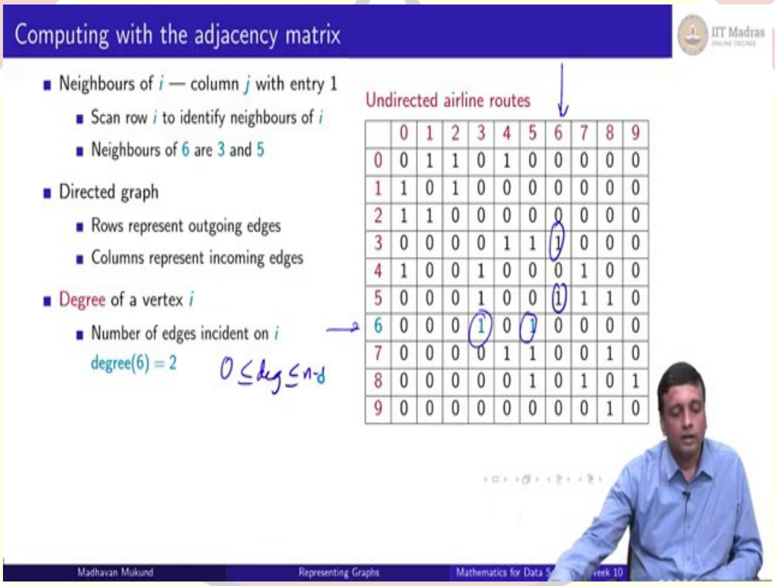


So, now we are looking at this picture, So, we have an entry point into 5 because there is an edge from 6 to 5, but the edge from 3 is coming in. It is not an outgoing edge.

(Refer Slide Time: 08:08)







So, if we look at this graph, for the directed case, then we have to look at the column for 6, we have to see where which things end in 6, So, the column for 6 has an entry in row 3. So, there is an edge from 3 to 6. So, the rows represent outgoing edges. And the columns represent incoming edges. Now, in an undirected graph, these are symmetric, if I have an outgoing edge to j, then i must have an incoming edge from j. So, both of these will have the same information.

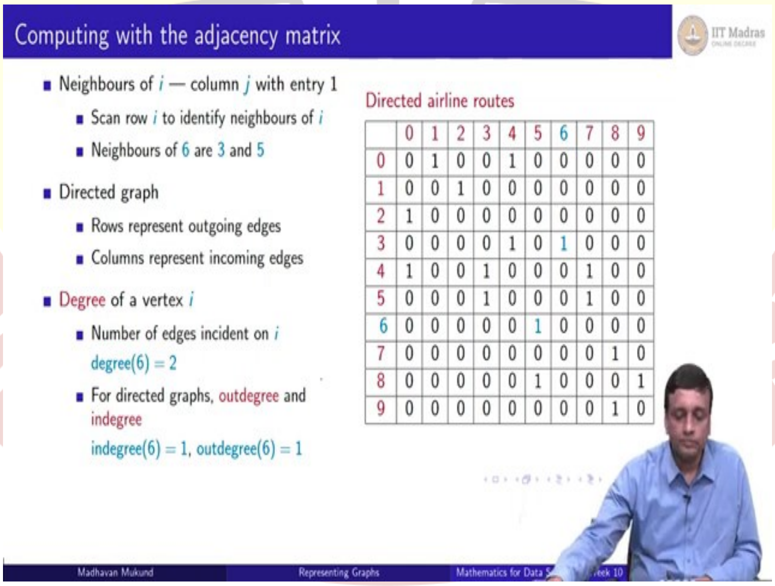
So, actually, I can also, find the neighbours of a vertex in an undirected graph by looking at the column o, it is conventional look at the row I. But the column has the same information. But in a directed graph, this is different. The outgoing edges are in the row, the incoming edges are on the column. So, the number of neighbours has a name and graph there, it is called a degree. So, the degree of a vertex is the number of edges, which are incident on that vertex that is number of edges, which start from that vertex in an undirected sense.

So, here, for instance, we saw that the degree of 6 is 2, because if I look at the row for 6, or if I look at the column for 6. So, if I look at the column for 6, also, I find that there are 2 incoming edges. So, to speak from 3 to 5, 3, and 5, and there are 2 outgoing edges from 6 to 3 and 5, this undirected. So, there is a uniform notion of degree, whether you are counting edges is coming in or going out, it does not matter. So, the degree of a vertex is the number of edges, and notice that each edge must go to a different vertex.

So, if you count the vertex you are starting at there are n minus 1 other vertices. So, the degree could be 0, in which case this vertex is not connected to anybody. I am a friendless soul, for example. Or I am in a city which is not served by this airline. So, I do not have any edges. So, I could have degree 0, or I could at most a degree n minus 1, everybody in the class is my friend.

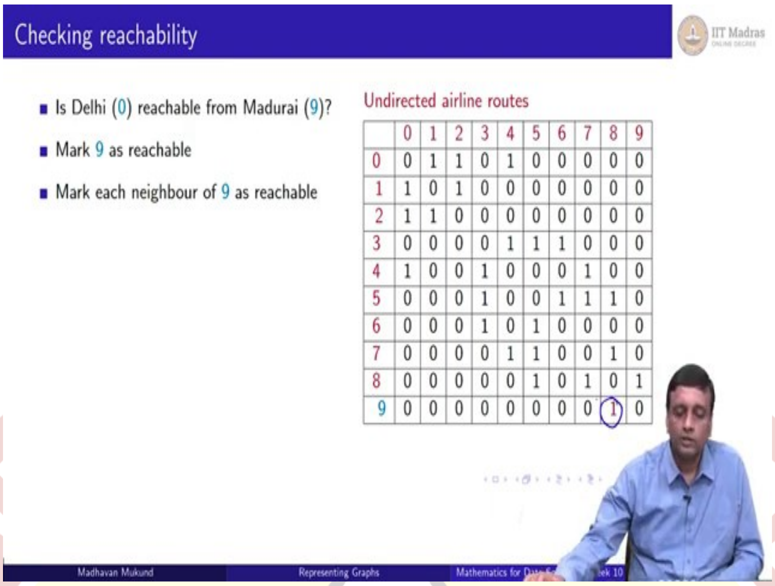
So, I am 1 person, and all the n minus 1 other people are my friend. So, I have degree n minus 1, or I have a direct flight to every other city on the network. So, this is the case. So, the degree is between 0 and, so, 0 less than equal to degree is less than equal to $n - 1$. So, this is something that you should remember.

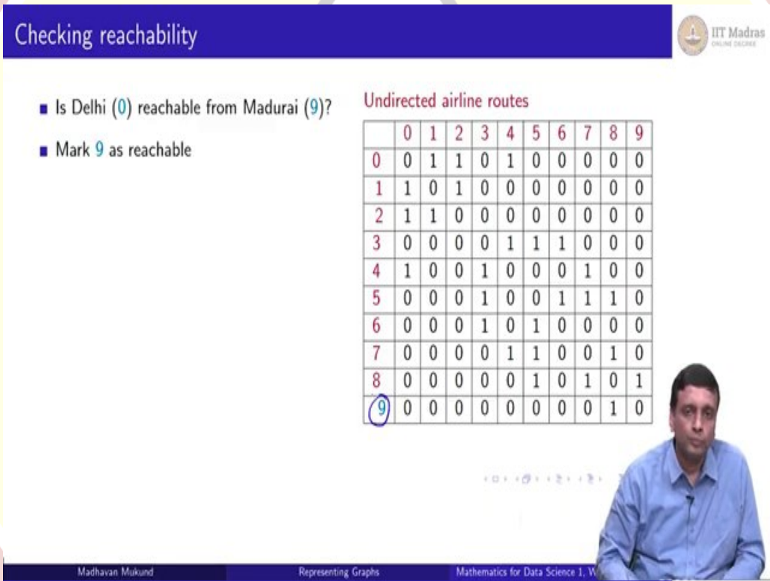
(Refer Slide Time: 10:13)

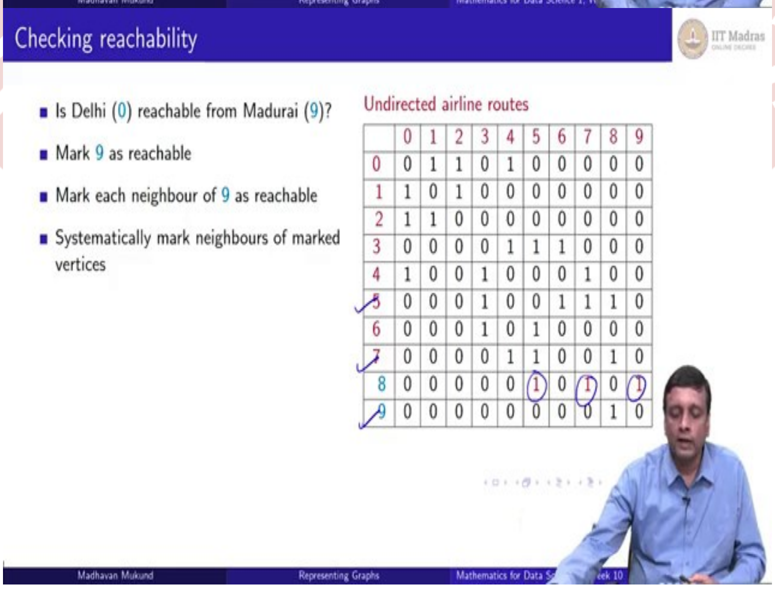


Now, if I have a directed graph, this notion of degree now gets split because I have incoming edges and outgoing edges. So, typically we will talk about the in-degree and out-degree. So, the degree of 6 in the undirected setting was 2, because there were 2 edges, but actually, 1 was pointing out to 5, and 1 was pointing in from 3. So, we can say that the end degree is 1 and the out-degree is 1.

(Refer Slide Time: 10:33)







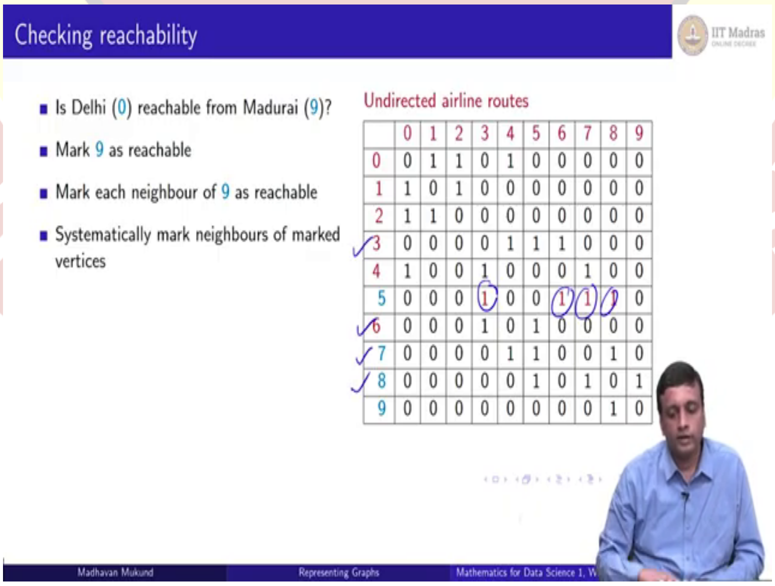
So, our goal, as we said, was to compute. So, we want to do something with this graph. So, in this particular case, how can I use this Adjacency matrix to check whether Delhi which is the vertex 0 is reachable from Madurai, which is a vertex 9. So, we do the natural thing, which is we start at 9, and then we see where all we can go.

So, we first mark that from 9, we can go to 9. So, by marking what I mean is, I will now take the row entry, and I will change the colour from red to blue. So, 9 is now reachable. So, now I can look at the neighbours of 9, in this particular case, there is only 1 neighbour 8. And if 9 is reachable, and I can fly from 9 to 8, then 8 is also reachable. So, I would mark 8 as reachable.

Now, what do I do, I have to start from 8 and see where all I can go. So, I Just have to systematically repeat this procedure. So, I have to systematically mark all the neighbours of marked vertices. So, 8 was marked. So, now 8 has 3 neighbours 5, 7, and 9. Now, notice that I do not have to refer to the picture, I am only referring to a table, I just have to look at the row for 8. And if I look at the row for 8, I know what the neighbours are at 8.

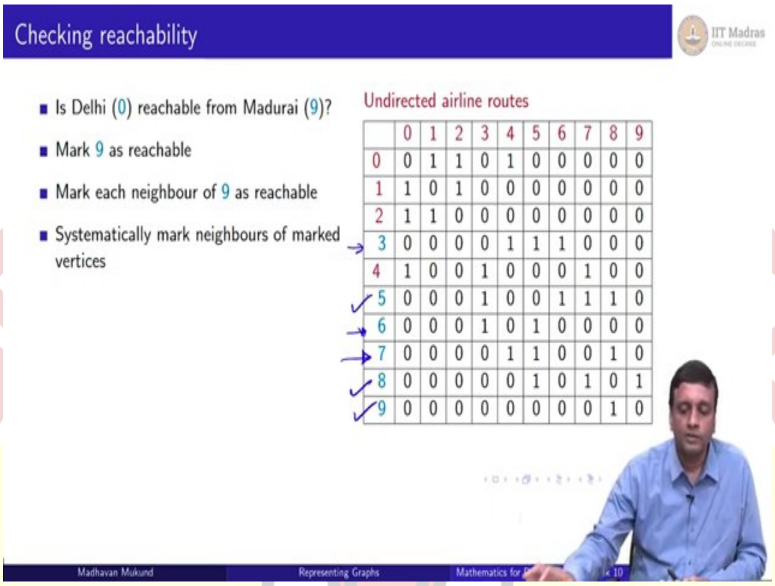
So, I do not have to go back to the picture. And imagine anything, this is just a mechanical analysis of this table. So, I look at this, and this tells me that I must now mark 5, I must mark 7. And I must mark 9, but 9 is already marked, so, it would not have any effect. So, the next step is to mark these new guys as also being visited or reachable.

(Refer Slide Time: 12:14)



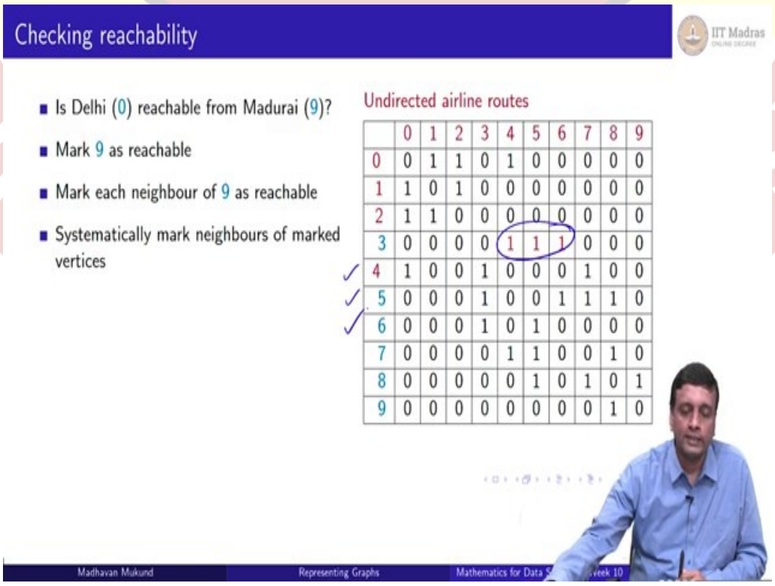
So, I mark 5, 7 and 9, as also reachable. So, now I have from 9, I can reach 5, 7, and 8, other than 9 itself. So, now I have not been very careful to keep track of it. But I know that I have explored the neighbours of 8. But I have now discovered 2 new neighbours, which I could reach 5 and 7. So, I pick 1 of them say pick 5. So, I look at the neighbours of 5. So, 5 has his neighbours 3, 6, 7 and 8. So, I have to know Mark 3, 6, 7 and 8 for which 7 and 8 are already known. So, I would mark 3 and 6.

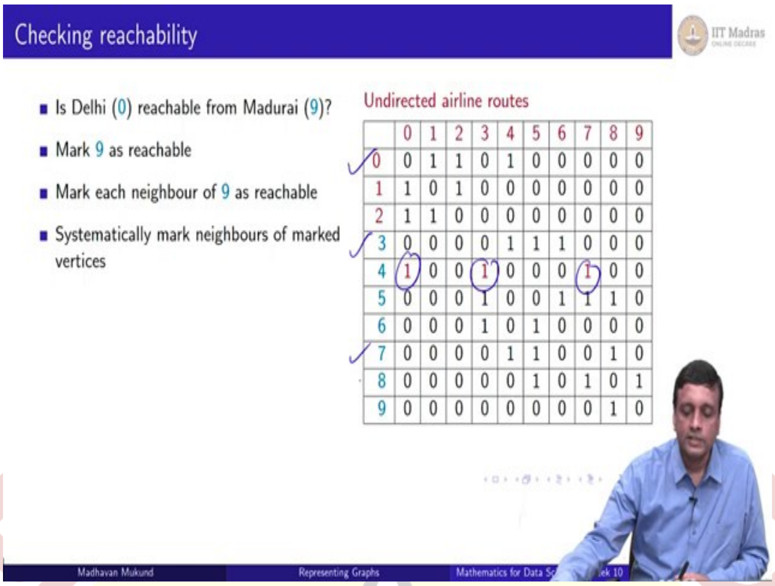
(Refer Slide Time: 12:47)



Now, once again, I have now as unexplored things I had marked 7, but I have not looked at the numbers of 7, I have marked now 3 and not looked at it and. So, 5, 8, and 9 have been explored that is I marked them, and then I also mark their neighbours. So, let me again go to the smallest unmark things with 3.

(Refer Slide Time: 13:11)

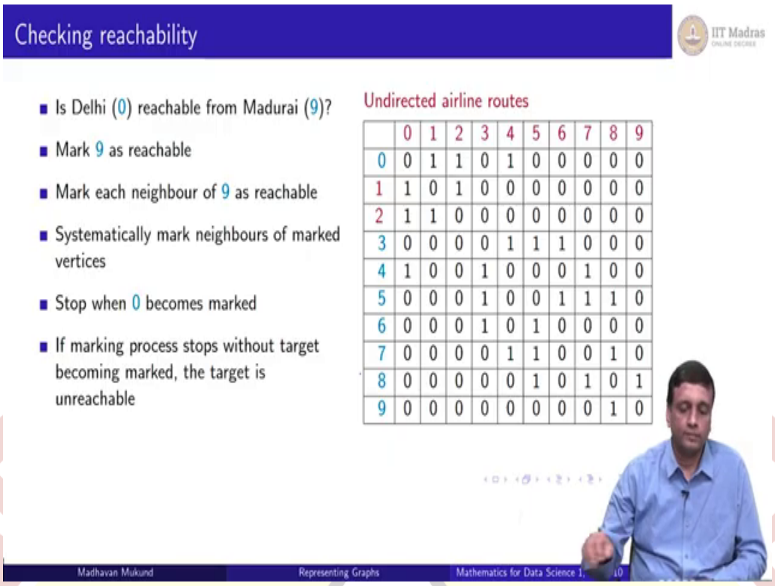




So, I look at 3 and I look at its neighbours. So, the neighbours of 3 are 4, 5, and 6. So, this means I was marked 4, 5, and 6, of which 5 and 6 were already marked before. So, nothing happens, but 4 gets marked. And now the smallest unexplored vertex is 4. So, I look at the outgoing neighbours of 4, and I get 0, 3, and 7, and therefore I must Mark 0, 3, and 7. And, So, once I have marked 0, 3, and 7, I can stop because this was my target, my target was to find out whether 0 is reachable from 9.

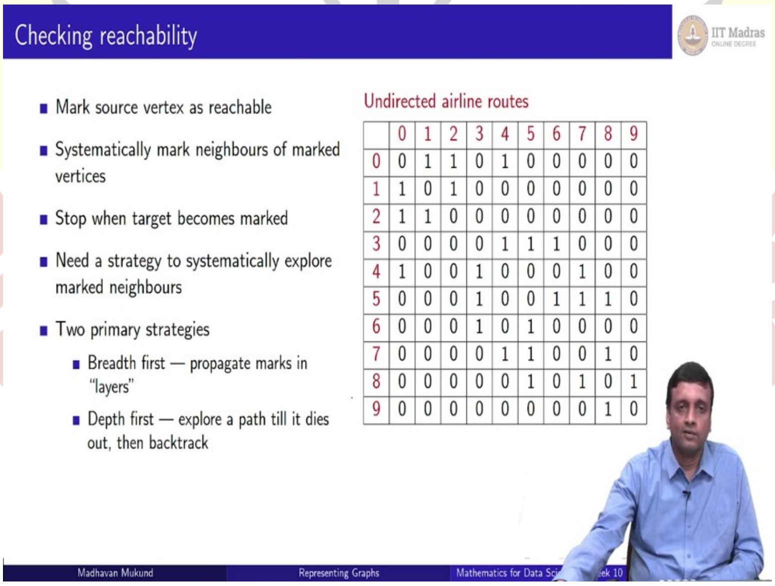
What I have seen is by systematically marking everything which is reachable in 1 statement, 2 steps and so, on, I have found out a way of reaching 0 from 9. So, I may or may not be able to reach 1 and 2, I do not need to do it for this particular calculation, if I wanted to find out what all is reachable from 9, I would continue and see if 1 and 2 get marked. But in this particular case, my only goal was to find out whether 0 is reachable from 9, I can stop.

(Refer Slide Time: 14:08)



So, the other situation is that it is perhaps not possible to reach 0 from 9 and what will happen there is that after I mark everything that can be marked, I will find that 0 is still not marked. So, if at the end of this process, where I have marked all the vertices and all the neighbours of all the vertices and I cannot mark anything more, and I find that some vertex remains unmarked, then that vertex is not reachable from where I started.

(Refer Slide Time: 14:34)



So, abstractly, what we said is we mark the starting vertex of the source vertex is reachable, and we systematically mark neighbours of mark vertices and we stop when the target becomes marked. So, this is what we computed. And we did this using the matrix that was the important thing that we did not go back to the picture and try to follow edges on the graph as a picture, but rather we scan the rows and we did some recolouring of the row headers. And in this process, we were able to explicitly compute whether we could reach 0 from 9 or not.

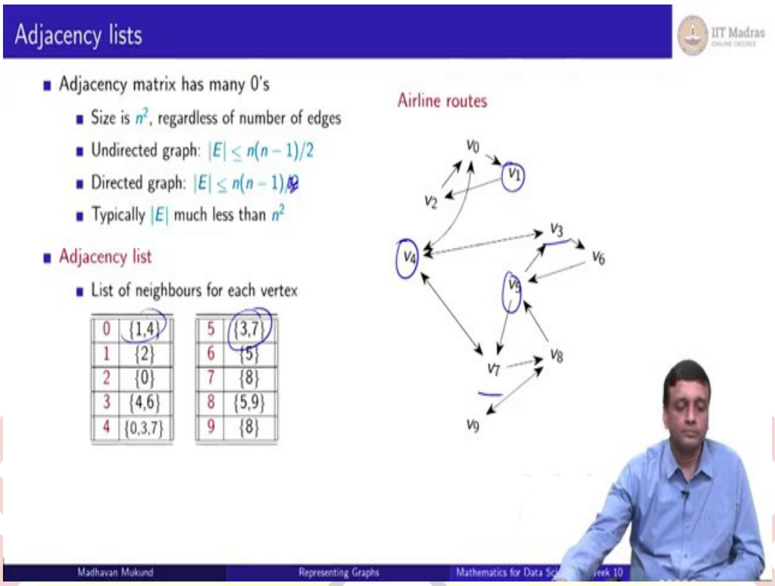
So, we had a kind of ad hoc strategy, which said that we will mark some things, and then we will pick up the smallest thing we have not explored and all that, but we did not systematically tell, how to do that. So, we actually need to elaborate that strategy more carefully, in order to get an actual computational procedure out of this.

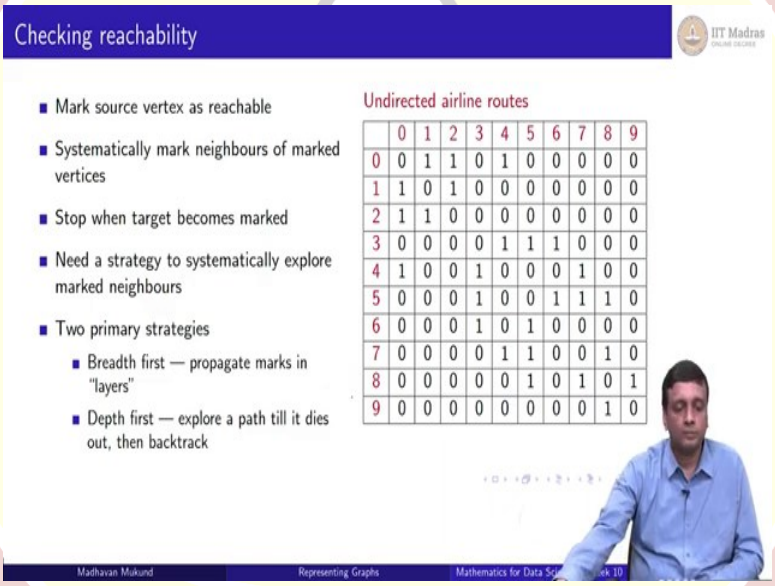
So, how do you systematically explore the mark neighbours? So, it turns out that there are 2 broad strategies for this. So, one is called breadth-first, which is what we were doing in a sense, but not really, which is that you propagate the marks and layers. So, we look at things that are reachable in 1 step. And then from 1 step, we look at things reachable in 2 steps, and so, on.

And the other strategy is called depth-first, which is I find one place, I can go to, there may be 3 places I can go to, but instead of going to the second place, I go to the place I could go to first and from there, I start exploring further where I can go. So, then I keep going down that path. And then eventually, I hit a dead-end, saying that there is no more places I can reach. And then I go back, and I say, Okay, now let me pick up the second place I could have started with and see where all I can go from there.

So, this is called depth-first search. So, you go as far as you can go in 1 direction, and your backup. And then you take the second direction back up until you run out of directions. And you keep doing this at every point. And this is called depth-first search. So, we will study breadth-first and depth-first search in more detail as we go along.

(Refer Slide Time: 16:28)





But before we do that, let us get back to this notion of how to represent a graph. So, the strategy that we have seen so, far is to use this adjacency matrix. So, the problem with the adjacency matrix is that as we have seen before, if you look at the adjacency matrix, here, you will see that there is a large number of 0s and 0s convey no information to us, it is only the once which are useful to us. So, the number of once is relatively small compared to the number of 0s. So, 0s are non-edges, and once are edges. And we are only interested only in the edges.

So, the size of this adjacency matrix, if I have n vertices is going to be n squared, regardless of how many edges I actually have. And now this is not a real problem in the long run, or in the extreme case, because you could have about n squared edges. So, if you have an undirected graph, then every pair of vertices will actually be an edge. So, you will have in terms of undirected edges, you will have n into n minus 1 by 2 because every edge is counted twice. But if you look at the matrix, it will actually have n into n minus 1 entry, because i j will be in the entry, and j i will also be in their entry. So, i j and j i represent the same edge. So, the number of actual edges is half that, but the number of entries is going to be n into $n - 1$.

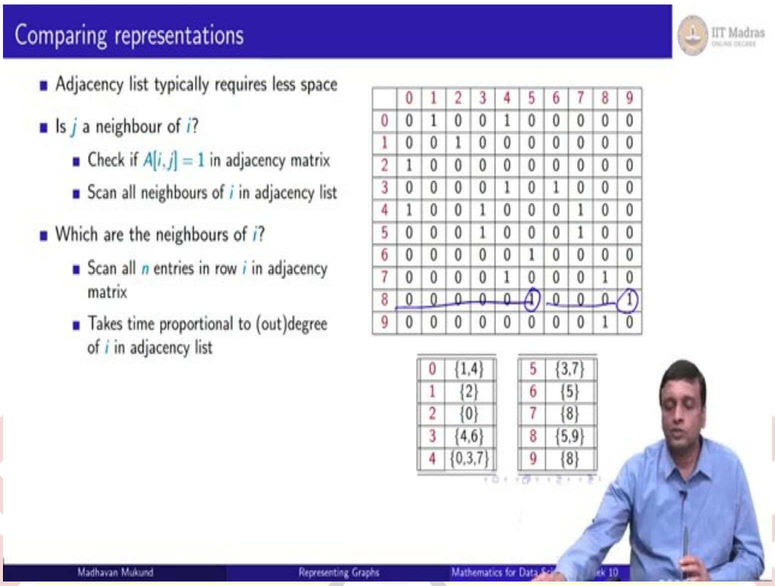
And of course, it was a directed graph. Also, you have n into n minus 1 should not be by 2. So, in both cases, you could have about n squared edges, but in typically you will not have n squared. So, typically a graph will have much fewer than n^2 entries. And if you have much fewer than n^2 entries, then it is not clear that having this large matrix is the best way to represent a graph.

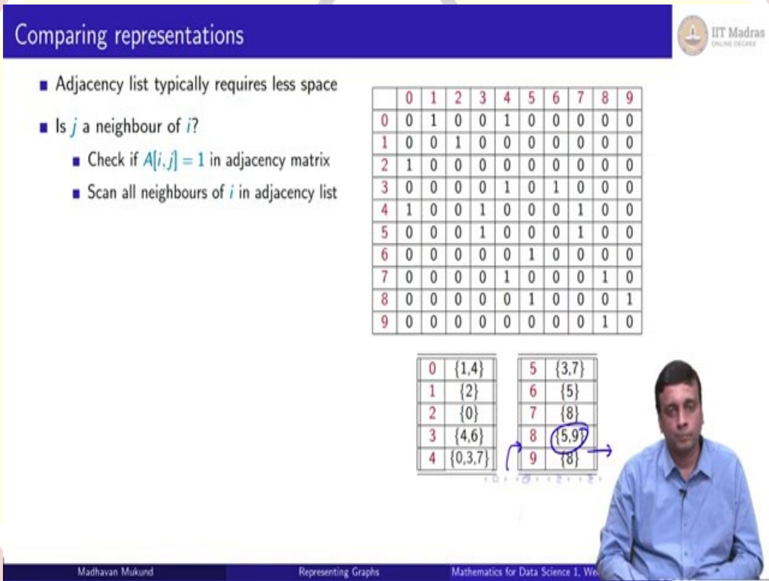
So, the other option is to just directly represent the neighbours of each matrix of each vertex in a list. So, this is what is called an adjacency list. So, in an adjacency list, what you do is you write down for each vertex, which are the neighbours of that vertex. So, it is most sensible in a directed graph.

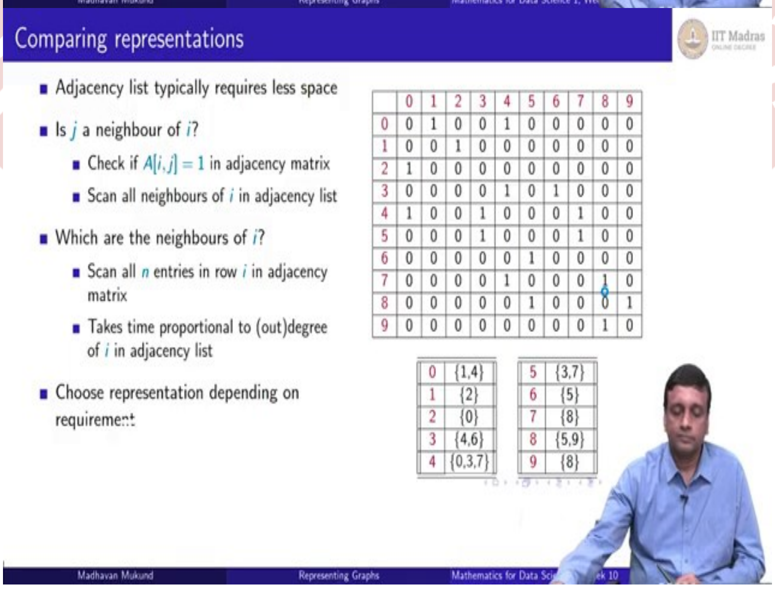
So, let us look at this directed graph here. So, it says that 0 is connected to 1 and 4. So, again, 0, you put this list 1, 4, similarly, 5 is connected to 3, and 7. So, against 5, you put 3 and 7. So, for each vertex in your graph, you just record what would have been in the adjacency matrix, what will be in the 1s in the row for that vertex.

So, if you look at row 5 in the vertex in the adjacency matrix, for this directed graph, it will have 2 once at 3 and 7, So, instead of writing all the 0s on the other 8 positions, we Just write the 1 position as 3 and 7. So, this is an adjacency list. So, this is an alternative way of presenting a graph, and we can work with this representation as well.

(Refer Slide Time: 19:06)







So, on the right-hand side, you see now, the 2 representations for that particular graph we have been looking at the top is the adjacency matrix for that directed airline graph. And the bottom is the adjacency list for the directed airline graph. And it is obvious from this picture, that the adjacency list is typically much more compact in terms of the amount of space that it occupies.

But of course, you have to do different things when you compute with these 2 things. So, for instance, if I want to check whether a vertex j is a neighbour of vertex I, is there an edge from I to j, in the adjacency matrix, I just have to look at the cell i,j, and check whether it is 1 or not.

So, assuming that I can look up every entry in the matrix in constant time, in the same amount of time, then checking with as an edge between i and j is taking the same amount of time for every i and j. On the other hand, if I want to check in the adjacency lists matrix representation. Whether, say, for example, this is an edge from 8 to 9, then I have to go to 8. And then I have to walk down this list. In this case, I have to first check that the first entry is 5, and then I have to look at 9 and so, on. So, I have to go through the entire list for a given vertex to determine whether or not there is a neighbour. So, it is a little bit more expensive.

On the other hand, if I want to know all the neighbours of i, then the adjacency list directly gives it to us. So, if there are 5 neighbours, there are 5 entries anyway, I look at 5 entries, I will find them if there are 2 end neighbours, I will get 2 entries. On the other hand, in an adjacency matrix, regardless of how many real neighbours there are, you have to scan the entire row, because you do not know whether there is a 1 coming up or not.

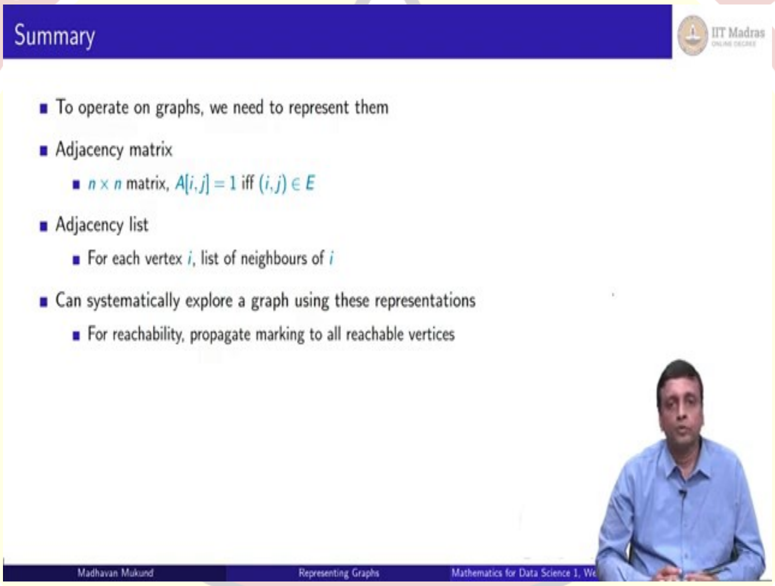
So, you cannot stop and say okay, after 7, I do not believe there are any more neighbours. So, you have to go. For example, if you started 8, I cannot go up to this point and say, Okay, I found 1 neighbour, and there are no more, you have to keep going because you might find a neighbour at last position.

So, regardless of what is the actual degree or out-degree, in this case, because it is a directed graph, regardless of the actual degree of the vertex, you have to spend order n time in order to collect all the neighbours of a given node, in an adjacency list, the time you take at each node is actually proportional to the degree. Now, in practice, many graphs will have a small degree, a given node will not be connected to very many other nodes.

So, therefore, if you have a procedure, which is proportional to degree rather than the number of nodes, it often works much faster. That is why it is useful to have this representation. So, there are trade-offs. So, it is not always the case that 1 is better than the other. So, typically, an adjacency list will save space. But there are some situations where it will incur some additional computation and vice versa. So, if you can make do with an adjacency matrix, and it is very simple to work with it, but then you have to do the scanning of rows and columns.

On the other hand, with a decency list, some things are not so convenient. For example, imagine how you would find out where there is an incoming edge in a directed graph, because an incoming edge will be recorded in the list for the other one, so, you have to go there and look at that list. So, there are other things that you have to do indirectly, with an adjacency list representation. But it is important to recognize that there are these 2 different ways of representing graphs and both of them are used in algorithms.

(Refer Slide Time: 22:15)



So, to summarize, what we have seen is that, we cannot just think of a graph as a picture to describe a procedure on it, we need a representation. And we need a way of writing it down in a way that we can manipulate mathematically. And we came up with 2 different representations.

So, 1 is the adjacency matrix, which is a matrix of n cross n for n vertices, which says that A i j is 1 if there is an edge from i to j, otherwise, it is 0. And the other 1 which is more compact, in general, if we have not very many edges is what is called adjacency list, where for each vertex, we list out the neighbours of that vertex in the list.

And with either 1 of them, we did an example using the adjacency matrix, you can do now systematic computation. So, it is a systematic exploration of whether or not a vertex V is reachable from a vertex U. So, we will look at more such things, in particular, we will look at these 2 strategies which we described mentioned but did not do in detail which is breadth-first search and depth-first search and then we will also, look at other properties of graphs that you can compute using these 2 representations.

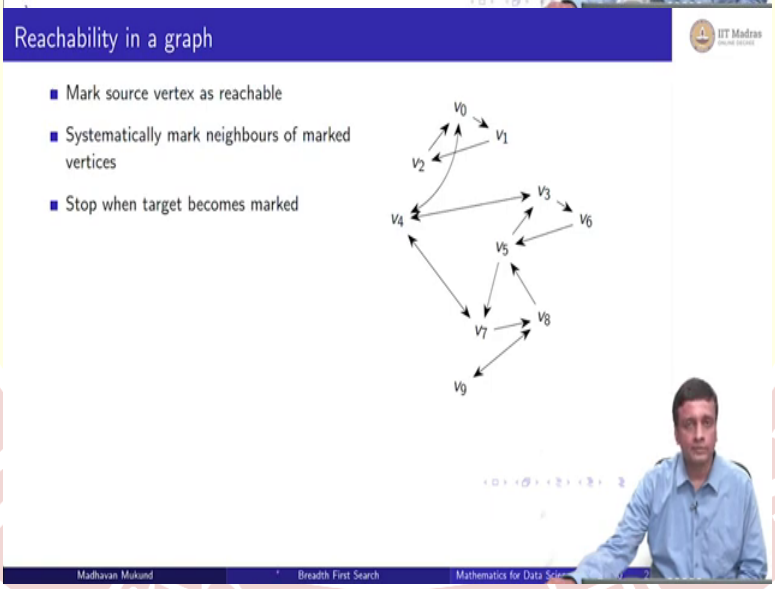
INDIAN INSTITUTE OF TECHNOLOGY MADRAS • सिद्धिर्भवति कर्मजा •

IIT Madras

ONLINE DEGREE

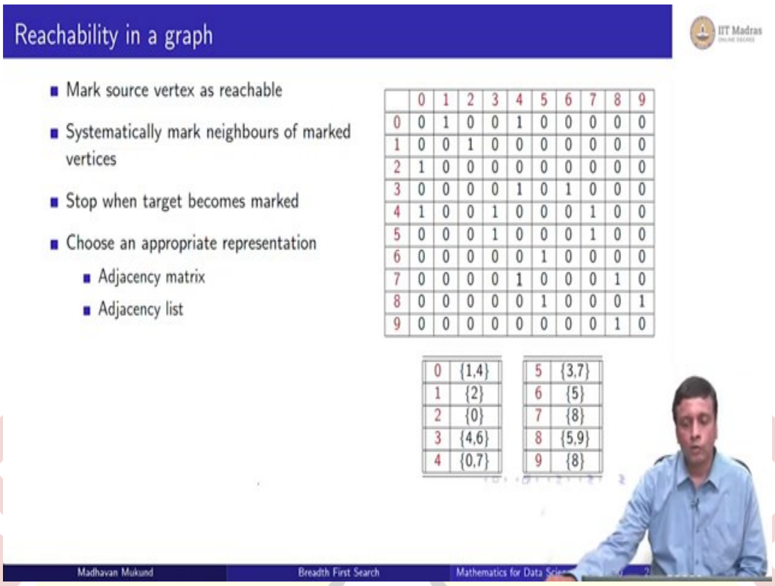
Mathematics for Data Science 1
Professor Madhavan Mukund
Chennai Mathematical Institute
Lecture: 62
Breadth-first search

(Refer Slide Time: 00:14)



So, we have been looking at this question of reachability in a graph. So, we said that to find whether a vertex is reachable for target vertex is reachable from a source vertex, we systematically explore the graph, we mark the source vertex, and then we go to its neighbours mark its neighbours and keep doing this systematically until the target becomes marked.

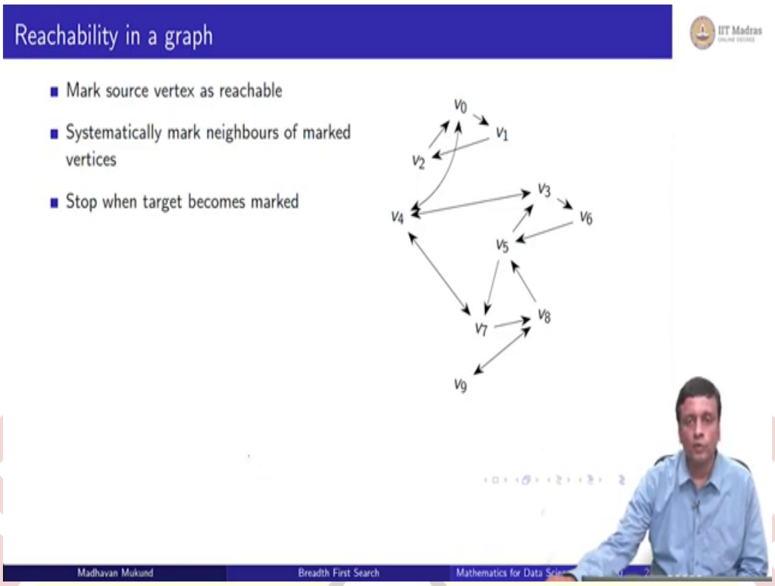
(Refer Slide Time: 00:36)



So, what we saw last time was that, in order to do this as a procedure, we have to have a way of representing the graph, which is not a picture. So, we came up with these two different representations. The adjacency matrix has a row and a column for every vertex. So, remember that we are assuming that our vertices are always numbered 0 to n minus 1 for convenience.

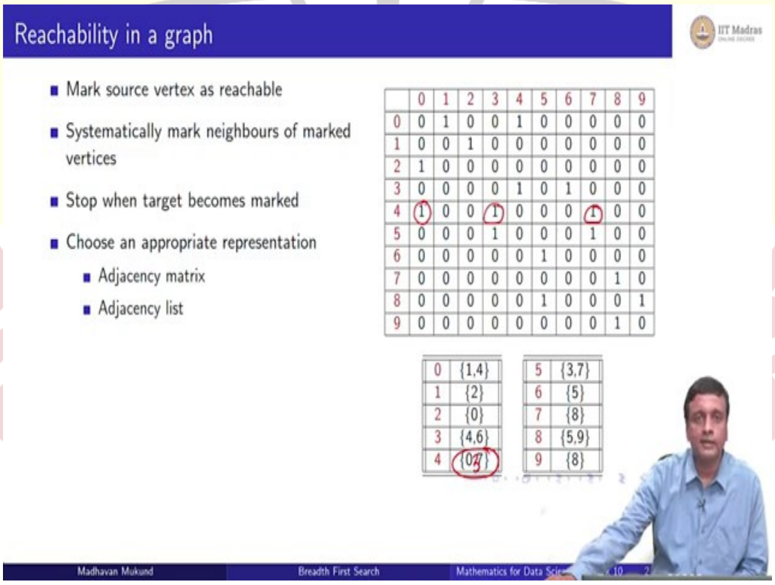
So, we have an entry at row i column j, it indicates whether or not there is an edge from i to j. So, if it is a 1, it means there is an edge if there is no 1, if it is a 0, it means there is no edge. And then we observe that this could be a wasteful representation. If there are not that many edges, a number of entries in this matrix are 0. So, instead, we could Just record an adjacency list for each vertex, just record the neighbours.

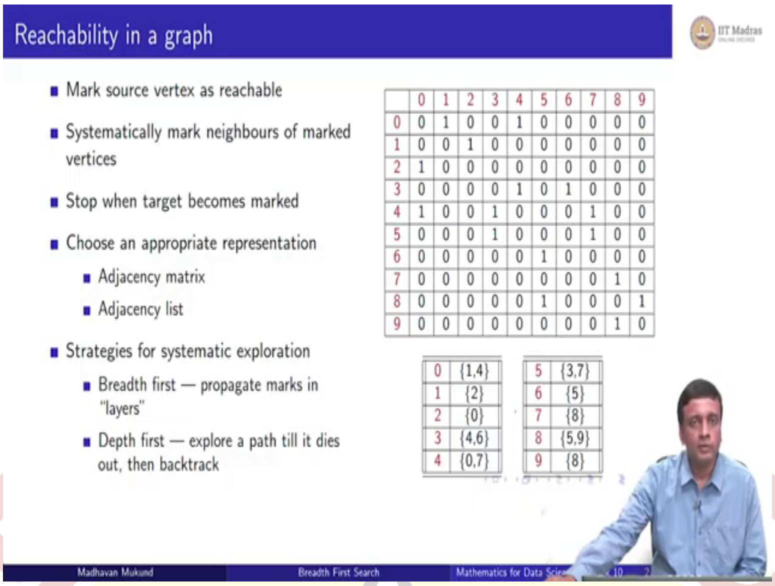
(Refer Slide Time: 01:23)



So, if we go back to this graph that we had just now, So, if we look at V_4, for example, So, V_4 has an outgoing edge to V_0, and outgoing is to V_3, it also has an incoming edge from V_0 and V_3, and then it has an outgoing edge and incoming it from V_7, So, the neighbours of 4 are 0, 3, and 7.

(Refer Slide Time: 01:40)

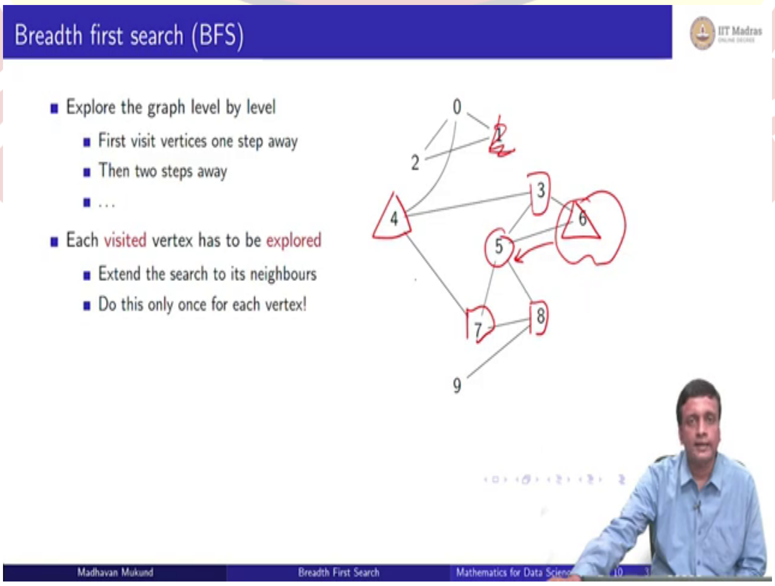




So, if you look at this adjacency matrix, it says that in the neighbours of 4, if you look at the row 4, it has once at positions 0, 3, and 7. And here, it does not have. So, there is 1 entry missing here. So, it should be 0, 3, and 7. So, this is how we would represent the matrix as either the graph is either an adjacency matrix or adjacency list. So, what we are going to look at now is the second part of the story, which is having represented the graph in this way where we can manipulate it, how do we actually systematically explore the graph.

So, we did a kind of ad hoc exploration last time using the adjacency matrix. But this time, we want to do it systematically. And there are 2 systematic ways to do this 1 is called breadth-first and 1 is called depth-first. So, first, we will look at breadth-first.

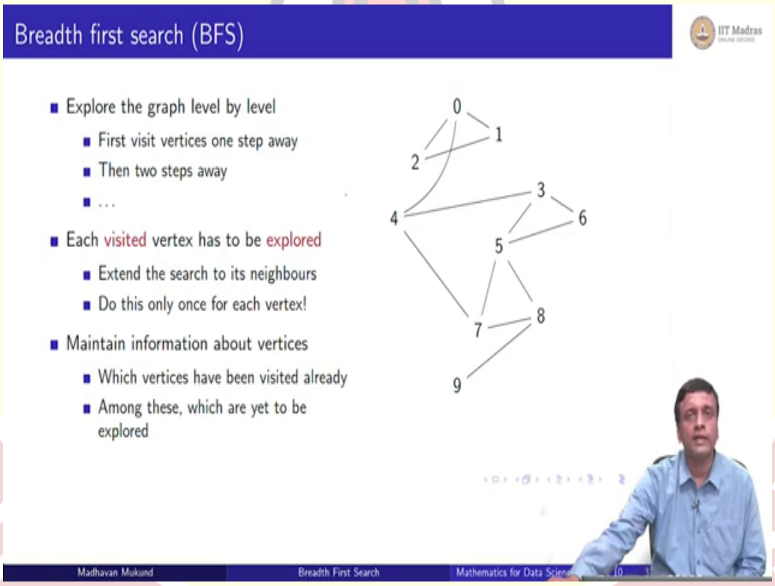
(Refer Slide Time: 02:33)



So, when we explore a graph, breadth-first, we do it level by level. So, we start with the vertex, then we identify all the neighbours have this vertex, that is all the edges, which connect this vertex to its neighbours, we look at those endpoints and say that these are now all reachable from this vertex, then we go to those neighbours, and look at what is reachable from there. So, these are things which are 2 levels away. So, 1 level away at the neighbours of the source vertex, 2 levels away are the neighbours of the neighbours, and so, on.

Now, what happens, of course, is that we might end up with a situation where we start with 5, and then we identify its neighbours as say, 3, 8, and 7. And then we go to 3, and then we identify its neighbours as 4, 6, and 1, or not 4 and 6. And now i look at the neighbours of 6. And it says, oh 5 is a neighbour of 6. But we started with 5. So, we need to record that a vertex has been visited. And we need to make sure that we do not visit a vertex twice. Otherwise, we will go round and round this kind of a triangle or a cycle a number of times.

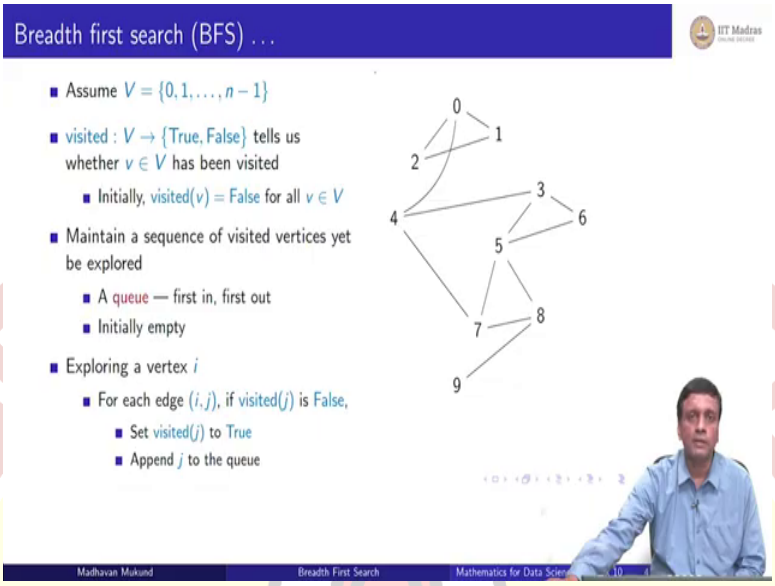
(Refer Slide Time: 03:39)



So, we need to do this visiting and exploring. So, exploring means we go from there to 1 more level of neighbours. So, we need to do this visiting and exploring exactly 1 per vertex. So, we need to maintain some information. So, we need to maintain information about which vertices have been marked as visited, that is, which note vertices have been marked as neighbours of something we have already seen.

And in the process of exploration, these have been marked, but we still have not looked at their neighbours. So, there is a subsequent process after visiting a vertex of exploring its neighbours. So, whether or not such a visited vertex remains to be explored. So, we need to keep 2 things. Has it been visited, and has it been explored? So, of course, it will be explored only after it is visited. But if it has been visited and not explored, that means this is some pending work that we have to do.

(Refer Slide Time: 04:26)



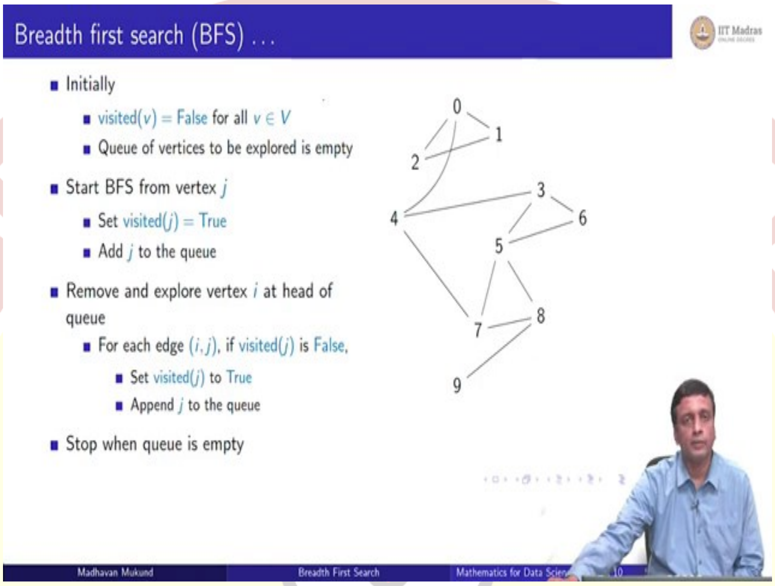
So, as we said, we will always assume that we have n vertices, we call them 0 to n minus 1. So, here is an undirected graph with 9 vertices, 10 vertices, numbered 0 to 9. So, what we will do is record the visited information as a function. If you are thinking about it in programming terms in terms of the computational thinking course, you can think of it as a dictionary whose keys are 0 to 9. It does not really matter, but it is actually mathematically a function for each vertex, it tells us true the vertex has been visited or false the vertex has not been visited so, far.

So, initially, we assumed that no vertex has been visited because we have not explored the graph at all. So, initially visited of V is false for every vertex. And now, we also separately have to look at the vertices which have been visited that have been marked true by visited, but which have not been explored. So, exploration means looking at the neighbours of that vertex. So, we have not yet looked at the neighbours of that vertex. So, we have to keep a record of this. So, this is a sequence of vertices which have been so, far visited but not yet explored.

And we will keep this in a special kind of sequence called a queue. So, the queue has exactly the same meaning that you associate with an English queue. So, if you are standing in a line to buy a ticket, a queue forms you join at the end of the line, the person at the front of the line buys a ticket and moves away, the next person moves up to the front of the line. And gradually as you move forward, when you reach the head of the queue, you get your turn. So, the same way we will maintain these vertices in a queue. As they get visited, they will be added to the queue, when their turn comes, they will be explored.

So, exploring a vertex technically means the following, we want to look at all the edges which are outgoing from that vertex. So, we look at every edge i j, which is there in the graph. And if we have not yet visited j, then we mark j as being visited, and we add j to this queue of unexplored vertices. So, j has been marked now is visited. So, it is due to be explored when it is turned comes. So, we put it at the end of the queue.

(Refer Slide Time: 06:26)

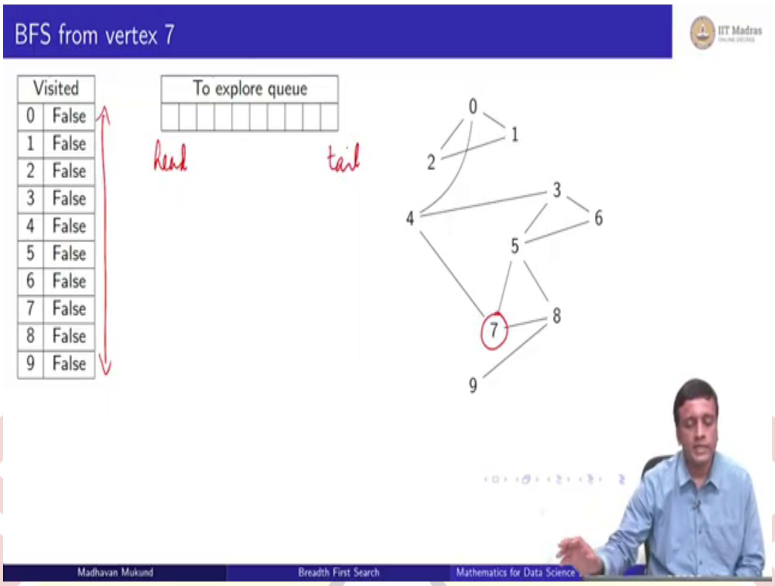


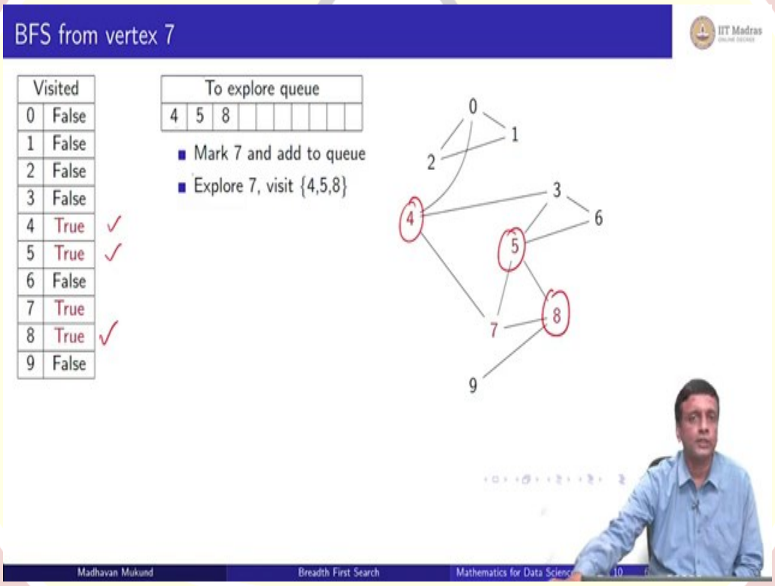
So, suppose, we start our BFS from a vertex j. So, what we will do is initially we will set visited of this vertex to be true, because we start there, So, we have visited that vertex, and now we have to explore it. So, what we do is we just put it into the queue, because our procedure is going to be to systematically explore all the vertices which are in the queue. So, we set the visited of j to true and we add j to this queue. And now, what we do repeatedly is we keep removing the vertex at the head of the queue. So, this queue, as i said is a line of vertices waiting to be explored.

So, we pick up the next 1, which is waiting, which is at the head of the queue the front of the queue, and explore it. And exploring it, as we said, is to check whether its neighbours which are visited, its neighbours have been visited or not. So, if a neighbour has not been visited, we mark it as visited and put it in the queue.

So, how do we stop? Well, if there is nothing in the queue to visit anymore, to explore anymore, this means that we have visited vertices and all the vertices we have visited, we have explored and there is nothing left to do. So, when the queue becomes empty, this breadth-first search as it is called BFS terminates.

(Refer Slide Time: 07:47)



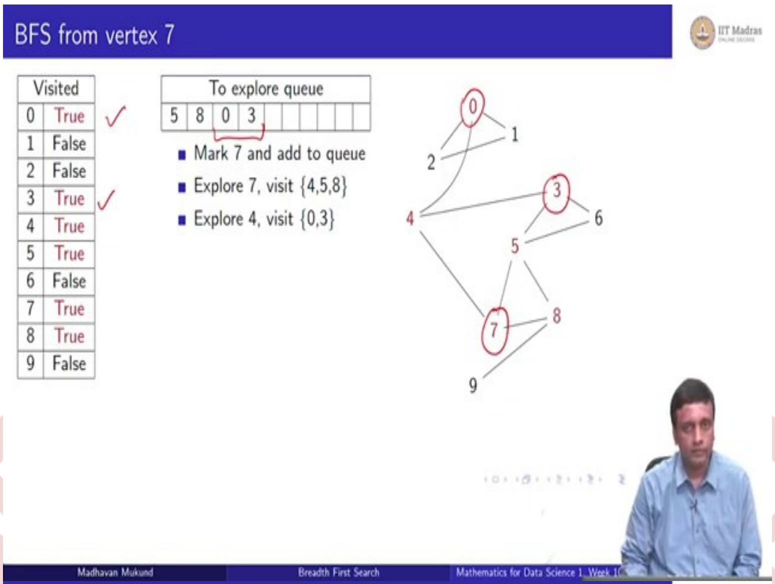


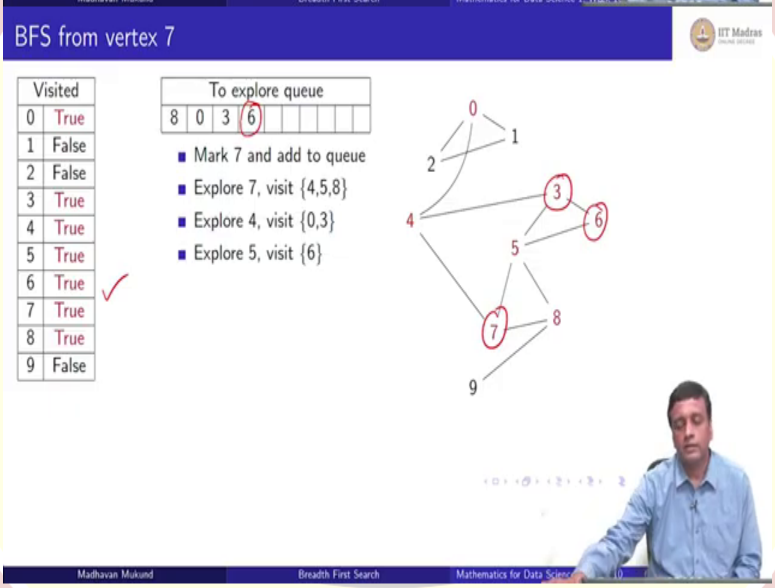
So, let us try an exploration of this graph. So, let us assume that we start at this vertex 7. So, we are going to start with 7. So, as we said, initially, we have set visited to false for all the vertices. And initially, we have this queue. So, the queue, I am going to assume has the left side as the front and the right side as the end. So, the head and the tail of the queue, we joined from the current tail of the queue, and we leave from the head of the queue. So, initially, the queue is empty, because we have not started anything yet. And initially, everything is unmarked. So, everything has visited set to false. So, we are starting from 7.

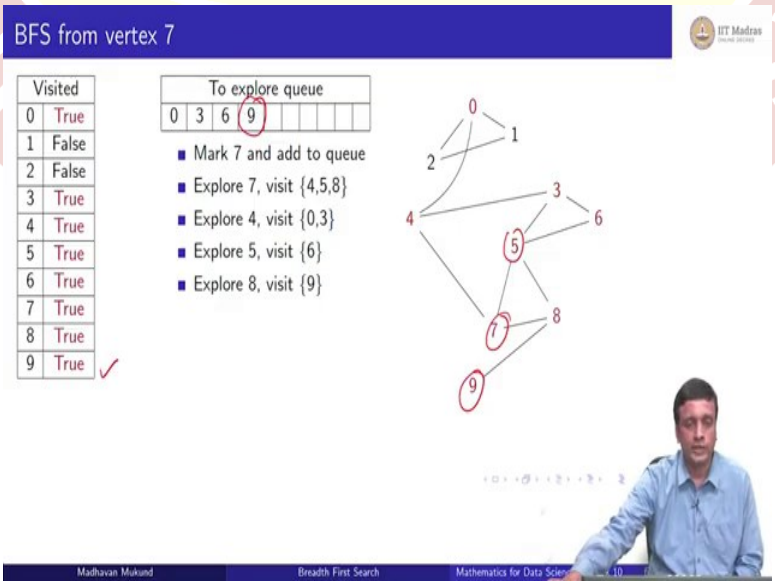
So, the first thing we do is that we mark 7, as visited. So, we mark 7 as visited. And Just to illustrate, we have also marked it on the graph. And now we have also put 7 in the queue, saying that 7 needs to be explored. So, we mark 7 as visited and added to the queue. Now, the real breadth-first search starts, which is you pick up the first element in the queue, and explore it. So, we pick up the 7, explore it. So, what are the neighbours of 7, the neighbours of 7 are 4, 5 and 8. So, in this process, 4, 5, and 8 get marked. And now they also get added to the queue.

So, I put them in some order, I have just put them in this particular case in ascending order, it does not matter, i could put it as 8,5,4 and 5,4,8. But is just convenient to put it in some fixed orders, I always put them from smallest to largest. So, 4, 5, and 8 were neighbours of 7, they had not been visited. So, now I have marked them as visited and added them to the queue. And notice the 7 has gone from the queue because 7 has been explored. So, 7 is no longer in the queue. Now, the first vertex in the queue to be explored is 4. So, the next step is to pick up the first vertex in the queue and explore it. So, I explore 4.

(Refer Slide Time: 09:35)





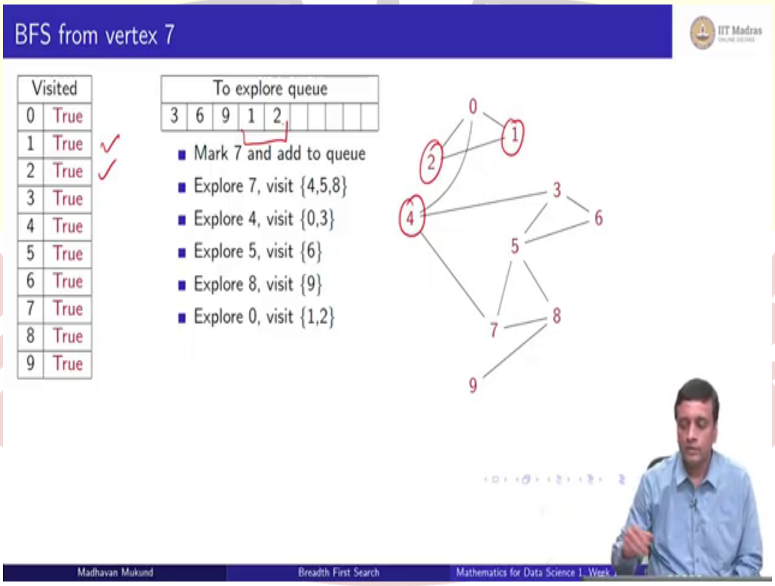


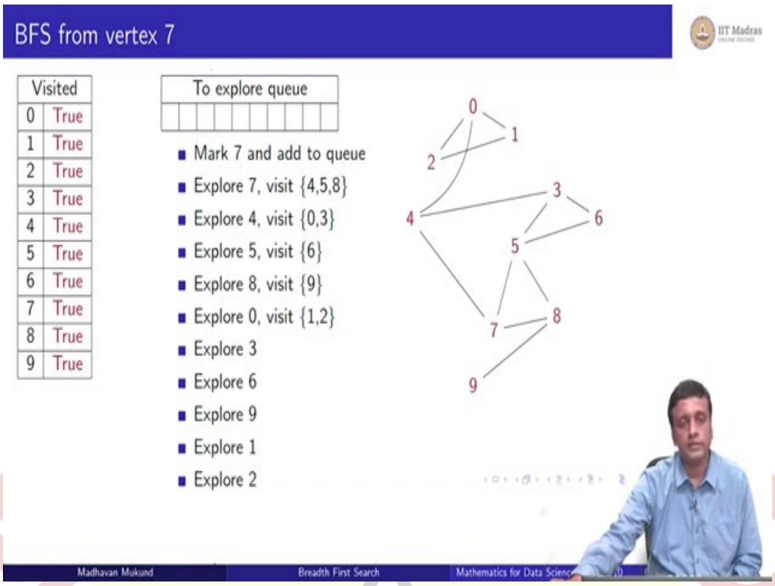
So, if i look at 4, it has neighbours, 0 and 3, and 7. But 7 is already been visited. So, i do not have to do anything about 7, i pick up the 2 which have not been visited, which is 0 and 3, and set their visited status to true and i add them in the queue in some order. In this particular case, as I said, I will put 0 before 3 Just because it is a smaller number. So, now i have finished with 4. So, 4 has left the queue. But from the previous rounds 5 and 8 are still pending.

So, 7 added 4, 5, and 8, i finished 4, 5, and 8 are still pending. So, 0 and 3 will have to wait their turn, they will have to wait until we are finished. So, in some sense, 4, 5, and 8 were 1 level away. So, until I finish all the vertices, which are 1 level away, I am not going to explore the vertices, which are added at level 2, namely, 0 and 3. So, what I do is I know to pick up 5 and explore it. And in the process, I look at the neighbours of 5, so, there are 3, 7, and 6. But 3 and 7 have already been visited so far, so, we do not have to look at them again. But 6 is new.

So, I marked 6, and I put it in the queue. Similarly, now i will pick up 8. And from 8 again, I have vertices, which are 5, 7, and 9. But since 5 and 7 were already visited, what is added now is 9 and 9 gets put into the queue. So, now I finally finished the level 1 vertices in the queue and have come to the 0 which is added at level 2.

(Refer Slide Time: 11:04)





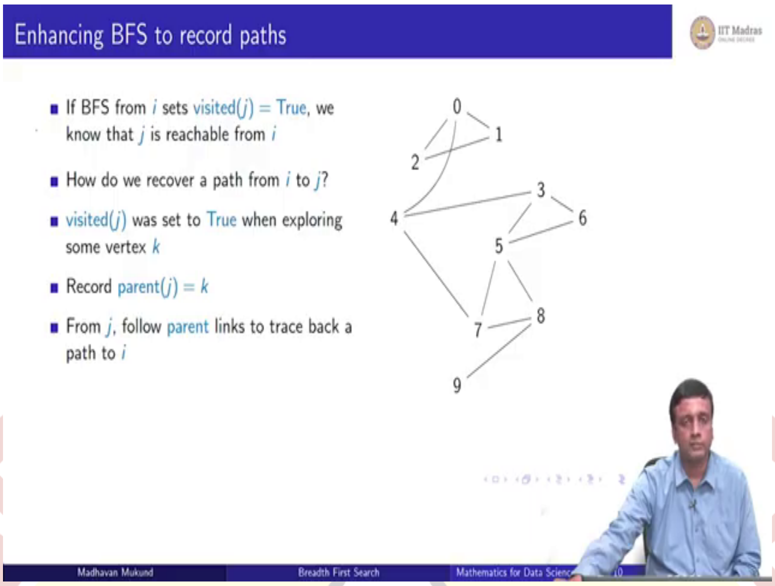
So, I explored the 0. So, 0 has 3 neighbours, 1, 2, and 4, but 4 is already visited earlier. In fact, 4 is where we came to 0 from in some sense. So, we now mark, 1 and 2. So, now you can see actually that everything has been marked, and 1 and 2 get into the queue. So, at this point, in principle, everything has been marked and you could stop.

But this is not how BFS stops, BFS stops by checking that every visited vertex has been explored. So, what we will do now is we will go to 3 and explore it, but we find that all the neighbours of 3 are already visited. So, exploration does not add anything to the queue, it only removes the 3 from the queue.

Next, we explore 6 similarly, all the neighbours of 6 have already been explored. So, exploration of 6 only remove 6 from the queue does not add anything to the queue does not mark anything new. Similarly, we have to explore 9 again, nothing new, explore 1, again, nothing new, explore 2, again, nothing new.

And now the queue is empty, I have run out of work to do. So, I have processed every vertex that I visited during my breadth-first search. And as a result, I have marked all the visit vertices, which I could visit starting from 7. In this particular case, you can see obviously, from this graph, that everything can be reached by some parts of the other from 7. So, everything is marked as true. So, this is breadth-first search.

(Refer Slide Time: 12:27)

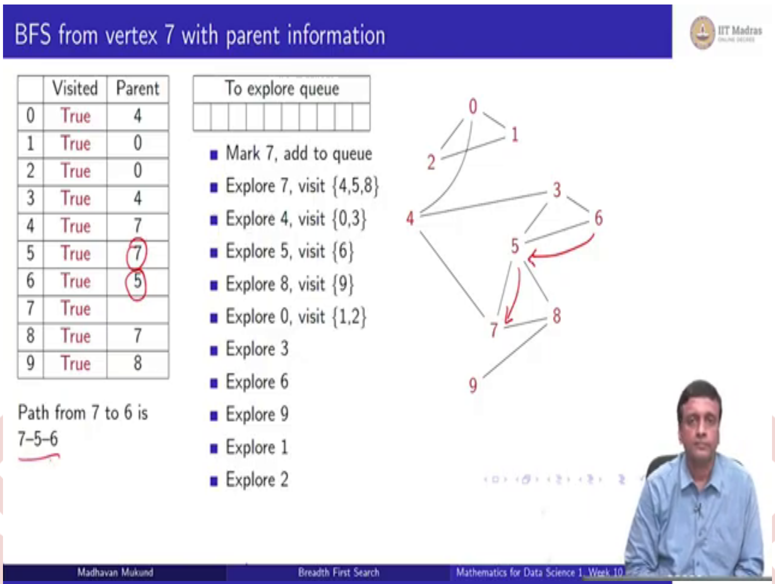


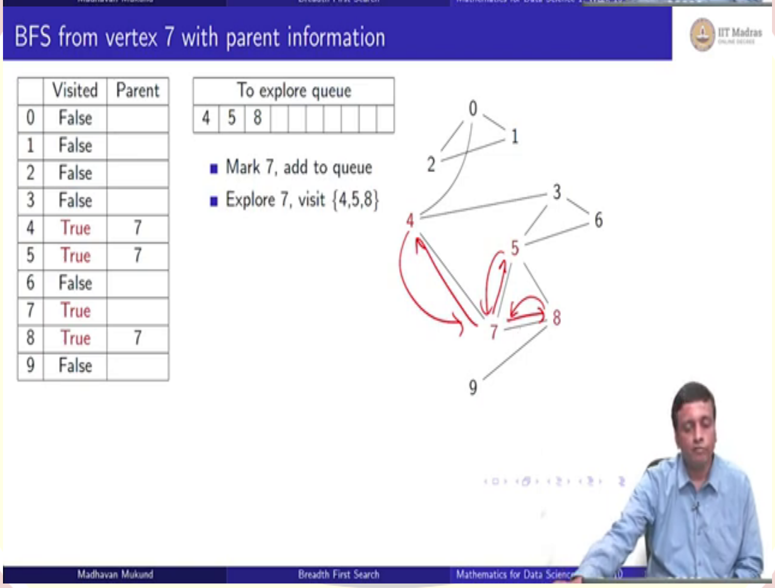
So, now we know for instance, that we can reach 1 from 7 because everything was marked as true. So, visitor, 1 is true. So, clearly, there was a way to go from 7 to 1. But this information that we have recorded in this visited vertex does not tell us anything about how to do this. So, if we have visited have j equal to true, after we have explored breadth-first search for I, we know that j is reachable from i. But we do not have any information about the path. So, this is now something that we can fix. So, how did we get to i, j, we reached j, because we explored j from some k.

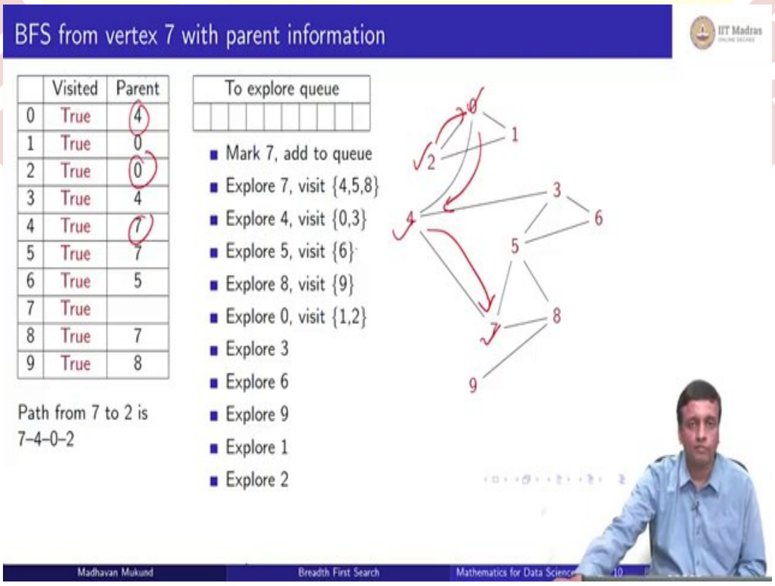
So, if we keep track of how we reached each vertex, we can work backward and extract the path. So, visited j was set to true when we explored some vertex k. So, we can say that the parent of j, the reason that j got added to the visited set was k. So, now if we go follow back from j to k, through this parent link, k itself would have been added because of some other vertex. So, we can look at parent of k.

So, maybe parent of k, some vertex L. So, we go to L, and we look at parent of L, everything was eventually traced back to the starting vertex. So, that is what marked the first set of vertices visited. So, in this way, we will walk backward and find the reverse path, which we can then, of course, enumerate in the forward direction and be done with it.

(Refer Slide Time: 13:51)







So, this can be done along with breadth-first search with no extra cost, except that we record more information. So, as before we start by setting visited to false, we are again starting with 7. So, that we Just have the same familiar computation to do. But now we have this extra column called parent, which records the parent that is the vertex from which this merge vertex was marked as visited. So, as before, we start by marking 7 and adding it to the queue.

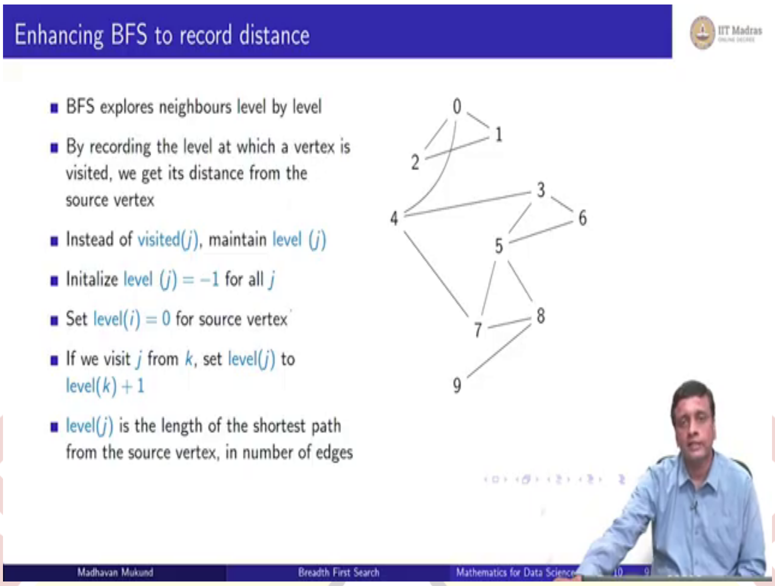
And then when we process 7, we add 4, 5, and 8 as marked, but we also set the parent of 4, 5, and 8 to be 7 to indicate that I came from here. So, let me draw an arrow like this. So, it says that I came from here, that is what this parent is saying that I came this way. So, whichever way you want to think of the arrow is going, it is pointing to the vertex which market. Now, when I process the 4, similarly, it is going to mark 0, and 3, and for 0 and 3.

Now the parent becomes 4 because that is how they got mark 4 and 0 and 3 did not get marked by 7 they got mark by 4. Same way, when I explore 5, the parent of 6 becomes 5. Same way, when I explore 8, the parent of 9 becomes 8, because I got to 8 and 9 from 8. And finally, when I come to 0, then the parents of 1 and 2 becomes 0 because 0 marks them. Notice that 7 does not have a parent. And that is because we started the search at 7. So, 7 got visited, because we initiated the breadth-first search at 7 not because of some other vertex marked it.

So, except for the Source vertex, all the other visited vertices will have a parent node marked in there. And now we can recover this information. So, we come back to the end. And finally, after we have explored everything we have emptied, the queue and breadth-first search is over. Now, we can ask for instance, what is the path from 7 to 6. So, the path from 7 to 6, that breadth-first search discovered, well, I go to 6, and 6 says the parent of 6 is 5. So, 6 says I came from here. And 5 now says I came from 7.

So, by following these parent licks, and traceback this path from 6 to 7, and I read this path in reverse and I get 7, 5, 6. So, here is a longer path, what is the path from 7 to 2. So, 2 says I came from 0, 0 says I came from 4, 4 says i came from 7. So, therefore reading it backward, it is now 7, 4, 0 and then 2. So, in this way, by keeping this parent information as we are exploring the graph, we also record the path for every reachable vertex or a path for every reachable vertex from the source vertex.

(Refer Slide Time: 16:45)



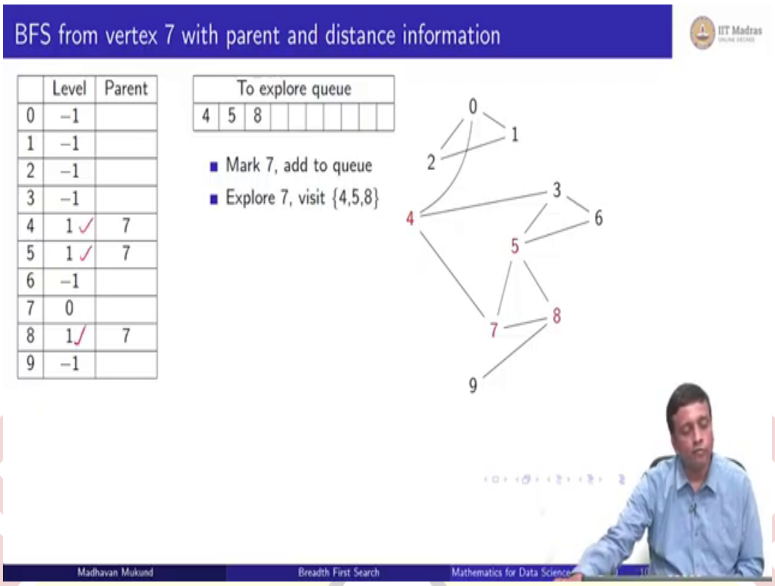
So, how about the distance. So, we have explained that BFS explores neighbours level by level. So, in some sense, the level of a vertex indicates the earliest that I can get to that vertex from the source vertex. So, all the vertices which are level 1 could be reached directly in 1 edge from the source vertex. Anything which is at level 2 is reachable from level 1 but was not reachable directly. So, I need two edges to reach it, and so on. So, the level gives us some notion of distance from the source vertex. And can we do that? So, it turns out that we can modify our breadth-first search.

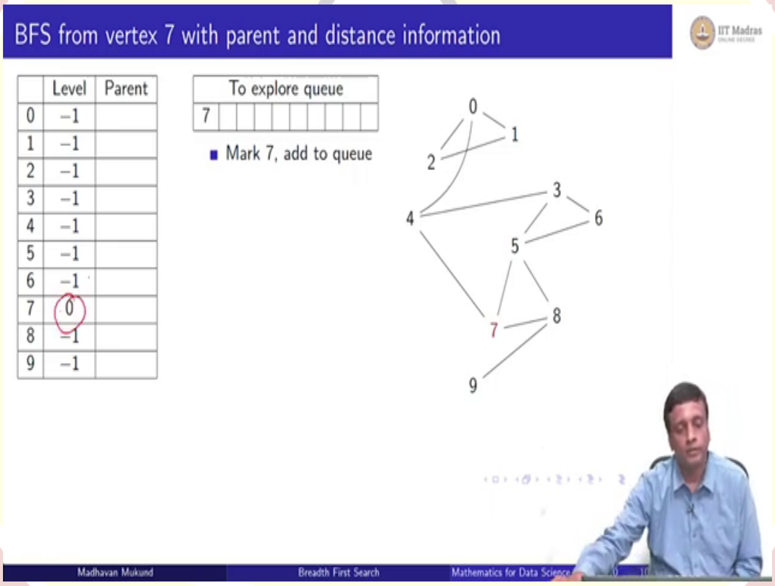
So, instead of just keeping this true false information for visited, we can replace it by this level information. So, every vertex which is reachable will have a level, which is 0 for the source vertex, and it has 1 or more for every other vertex that is reachable. So, since the minimum level is 0, for any reachable vertex, I can initialize the level to be minus 1.

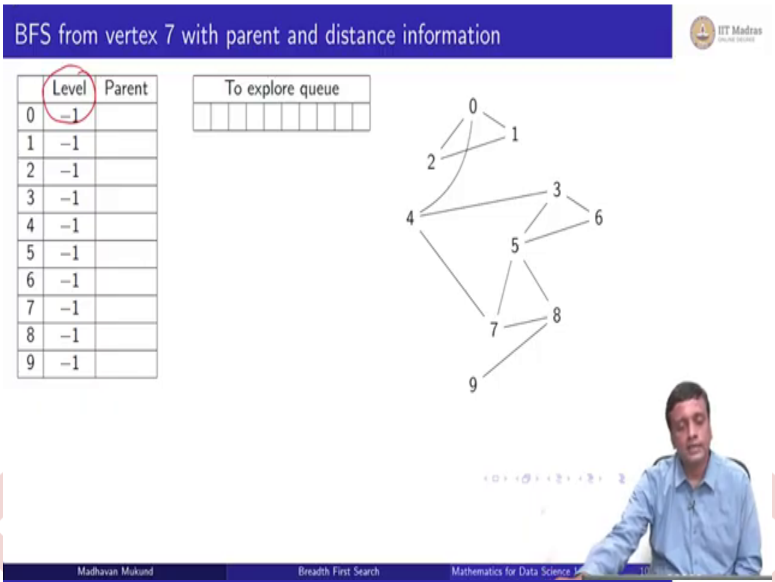
So, any vertex whose level at the end remains minus 1 is as good as something whose visited value was false. That is, i never reached it. Otherwise, what I do is I set the level to be 0 for the source vertex, assuming that I am starting from i. And whenever we visit a vertex j from vertex k, we already have a level assigned to k.

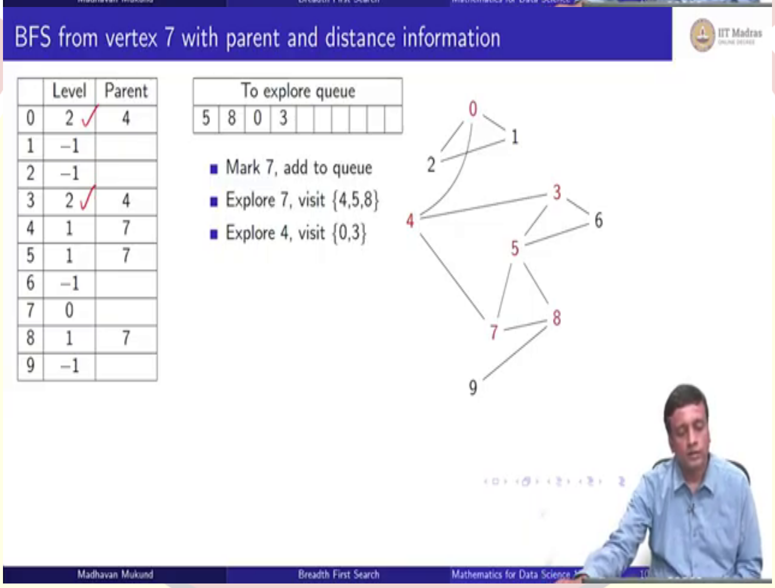
So, we assign the level of j to be the level of k plus 1. So, it should be clear that because we are doing this level by level, the length of the shortest path, in terms of number of edges that we have to travel is given by the level. So, if I look at level of j, I know exactly how many edges I need to take to get from i to j, there might be longer ways, but there is no shorter way.

(Refer Slide Time: 18:35)



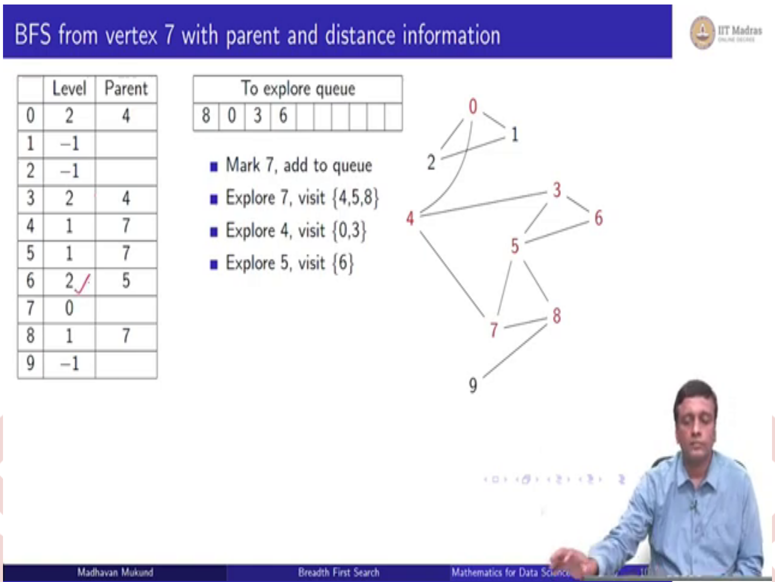


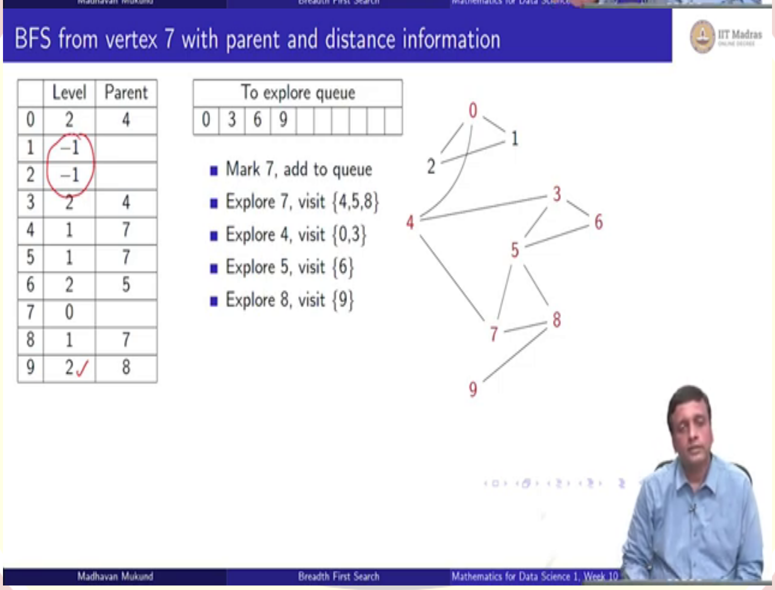




So, here is that breadth-first search again. So, I have now replaced this column visited by this column level. And as we said before, we put all the levels to be initially minus 1. So, now we start with our favourite starting point 7. So, we start by setting this level to be 0. So, instead of setting it to be true, as visited, we set its level to be 0. Now, when we explore 7, its neighbours get to level 1. So, 4, 5, and 8 are all at level 1. Now, when we explore the neighbour of 4, I get to 0 and 3. So, these now have level 0032 because the level of 4 was already 1, so, it is 1 plus 1 is 2.

(Refer Slide Time: 19:17)

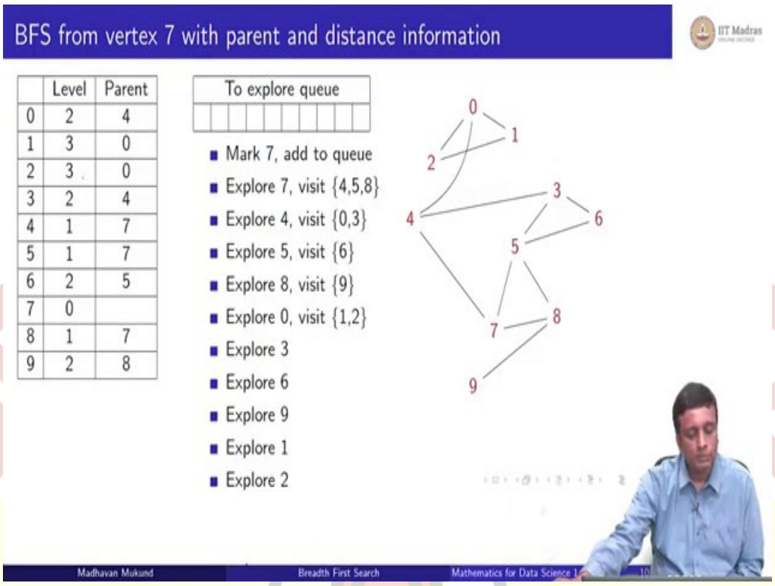




Now if I explore 5, though I am exploring 5, after 4, the level of 5 was the same as the level of 4. So, what I can reach from 5 is also at level 2, it is not that it is level has increased because I am exploring it after 4. So, both 4 and 5 were added at the same time. So, they were both at level 1 I Just happen to have happened to choose to process 4 before 5. So, 6 also becomes level 2.

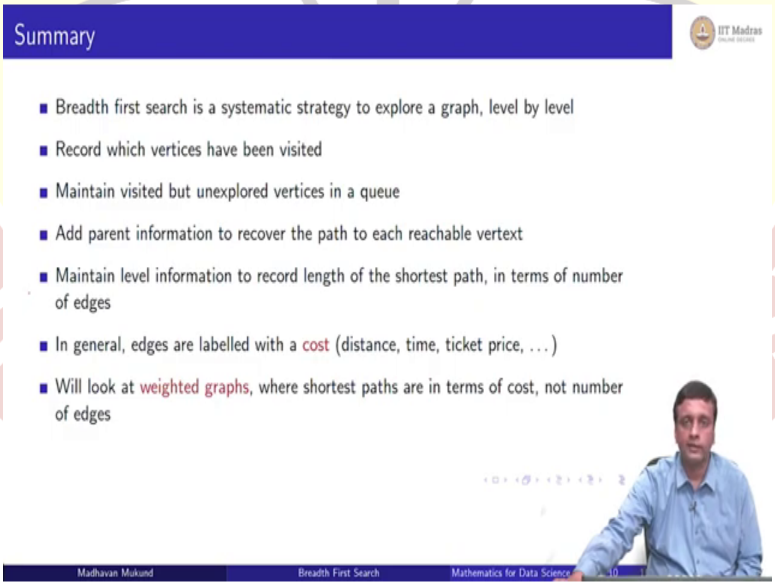
Similarly, when I look at 8, 8 was also originally a level 1 vertex, so 9 which is reachable from it also becomes. So, at this point, if we look, we can see that the 2 vertices which have not yet been visited, are the ones which have level minus 1. So, that is, of course going to be fixed next. When i explore 0, so notice that explorer 0 means that 0 has level two. So, these two become 3, because in order to reach 1 and 2, I have to first go to 0, and 0 was already 2 edges away. So, this gives us the breadth-first search. And as before, so, this is about paths, but as before, we could also record the, we have the parent vertices.

(Refer Slide Time: 20:23)



So, you can get the path. But we can also get the shortest distance. So, you can keep only the level only the parent both whatever you want.

(Refer Slide Time: 20:33)



So, to summarize breadth-first search is a systematic strategy to explore a graph level by level. So, remember that we said that broadly, the way we explore a graph is to systematically propagate these marks from a source vertex to the neighbours, to the neighbours, and so, on. But we need a way to do this. So, that we do not go around and around in cycles. So, we need a way to do it. So, that we terminate sensibly.

So, breadth-first search is one such strategy. So, what we do is we record the vertices which have been visited, and we maintain this queue of visited vertices, which are yet to be explored. So, exploration means explore the neighbours, then we saw that we can add parent information to recover the path. And we can maintain level information to record the distance, which is the shortest path in terms of number of edges. And what we will see is that in general, when we have a graph, we could record more information than just the edge.

For instance, if you had airline time graph, or a railway graph representing the route of railway network or an airline network, then typically with each edge, you would have associated with Some number, which we will abstractly call a cost. Now the cost could be a real cost, it could be the price of a ticket to go from the station to that station, or from this airport at that airport.

But the cost could also be a distance, it could also be how many kilometres this route travels. Or it could be time. For instance, if you have a train, the same distance could be covered in different times depending on the quality of the tracks and the number of stops in between.

So, in general, to go from one point to another point, you have to traverse real-time or pay Some real money, or spend Some distance traveling, say, supposing it is a road network, then it will give you a measure of how many liters of petrol you would expect to consume complete this distance.

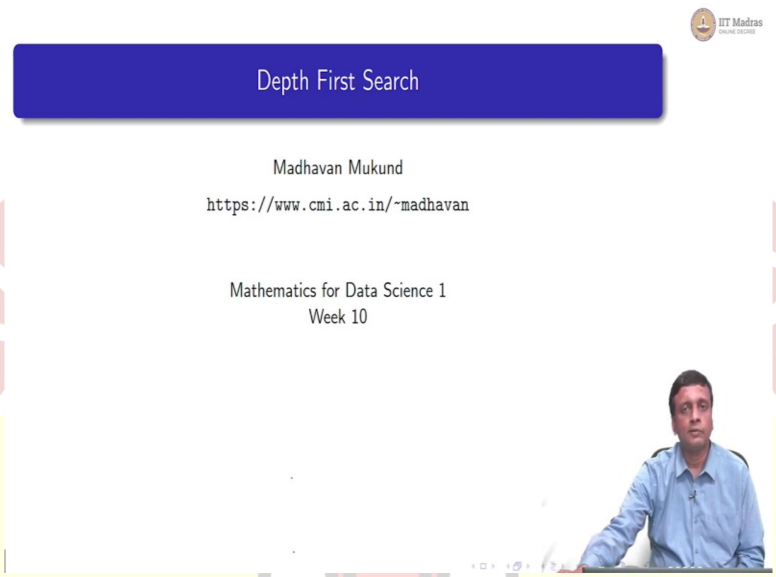
So, in these kinds of graphs, which are called waited graphs, the shortest paths are no longer just in terms of the number of edges, we could have a short edge, a single edge, which has a very high cost. It could be that a direct flight costs much more than a flight which goes by an intermediate airport because the airline is trying to encourage people to take these flights to unpopular airports in order to fill the planes. So, it is not always the shortest number of edges is also the shortest path. So, we will explore this separately when we look at waited graphs.

IIT Madras

ONLINE DEGREE

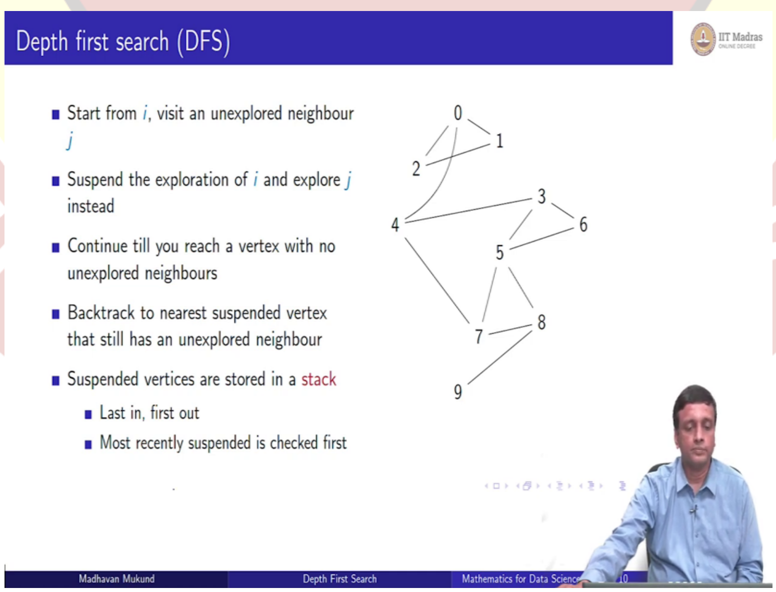
Mathematics for Data Science 1
Professor Madhavan Mukund
Depth first search

(Refer Slide Time: 0:14)



So, we had said that there are two systematic ways to explore a graph and we earlier looked at breadth first search which explore a graph level by level using a queue to maintain the unexplored vertices as we go along. The other strategy that I would use depth first search.

(Refer Slide Time: 0:31)



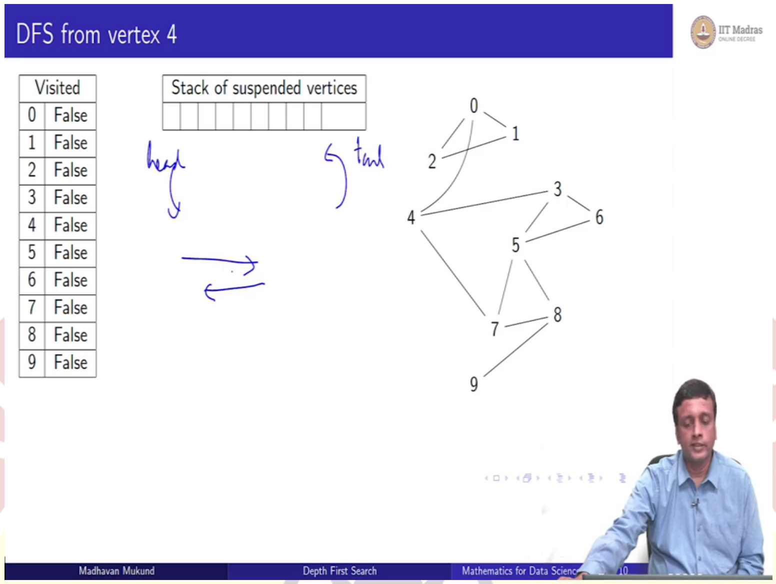
So, in depth first search we start from a vertex i and we pick anyone of its neighbor is not been explored in breadth first search. We pick all its neighbors and put them into a queue, here we pick any unexplored neighbor j and now what we do is we basically suspend the exploration of i and start exploring j instead. So then we look at j, we look at the neighbor of j, we pick again some neighbor of j which has not been explored.

We suspend j and go to the next vertex and we keep doing this until we run out of vertices that we can reach down this path. And now when I reach a vertex through this process which I cannot explore any further, I come back along the path I have taken. And find the first point where there was another choice for an unexplored. So you back trap to the nearest suspended vertex that still has an unexplored neighbor.

And then you explore that neighbor and so on. So here unlike in breadth first search where we had to keep track of these vertices which have been visited level by level and then put them into a queue and then we process this queue in this first in first out manner. So the queue, the vertices which are added level 1 get processed before the ones at level 2 and they get process before the ones at level 3 and so on.

In a depth first search if I have walked down some distance then I will come back to the point where I last stopped and see there was something else I could do from there and then I will walk back and so on. So I need not a first in first out but what is called a last in first out. So the last vertex has suspended is the first that I restore and check again. So this is called a stack, so these are 2 very fundamental ways of organizing sequences. A queue is a first in first out structure and that is used in breadth first search and a stack as we will see is used in this depth first search.

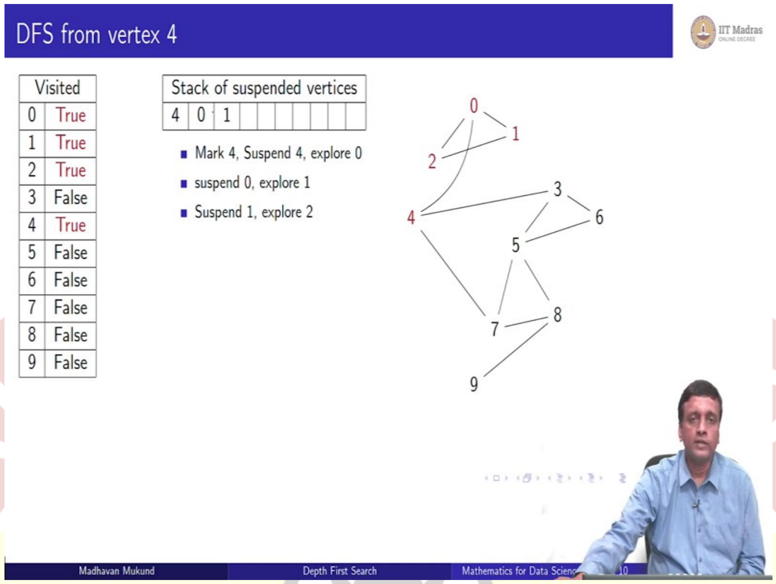
(Refer Slide Time: 2:24)



So, let us try to explore this depth first search, so for this our stack will again like our queue grow from left to right, our stack will again grow from left to right. So when I add this thing to the stack I will put it to the right. And now unlike a queue when I remove things from stack I will remove it from the right not from the left. Remember in a queue we had a head here and we had tail here.

So, we put things here and we brought them out here instead here I am going to grow the stack this way and also remove it from this end. So, that is going to be our strategies so for a change instead of 7 let suppose we start our depth first search from vertex 4.

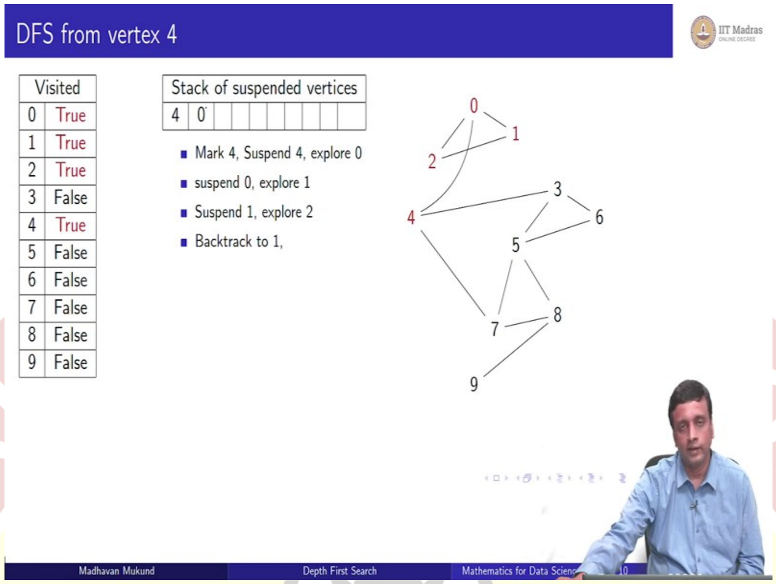
(Refer Slide Time: 3:04)



So, we first start as before by marking 4 as visited. And now we have to pick a neighbour of 4 and suspend 4 and start exploring that neighbour instead. So the neighbour of 4 has 0, 3 and 7 so we can pick anyone of them so let us pick the smallest one. So I suspend 4 and I start exploring 0 instead. So now I look at the neighbours of 0 and explore them if they have not been visited.

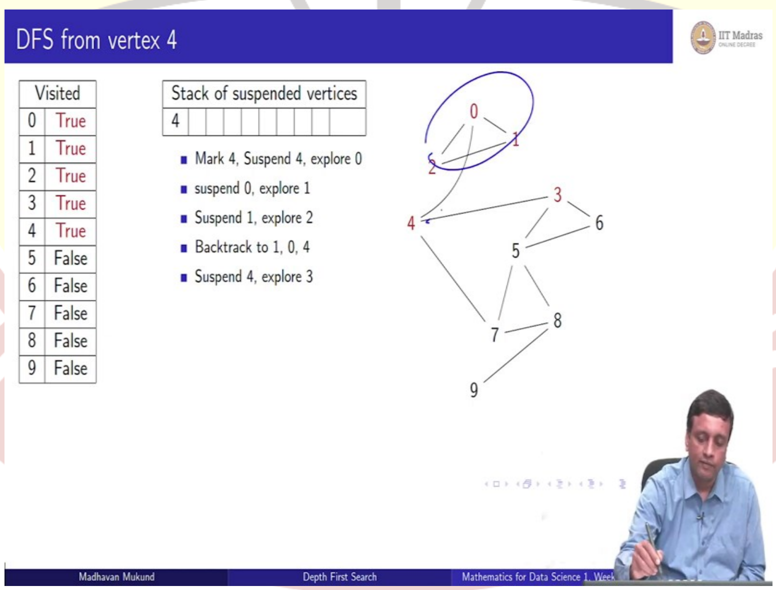
So notice that now visited this 2 for 4 and 4 and the stack has 4 in it but now I am going to suspend 0 and pick up one of its unexplored neighbour say one. So now I have marked 1 and 4. And so the stack is growing from left to right remember. So 0 has come on top of 4 in the stack. So now I have to explore 1, so explore 1 and I find it has only 1 neighbour which is unexplored namely 2. So I suspend 1 and I start exploring 2. Now when I look at 2, 2 has no neighbuors left to explore because it has only 2 neighbours 0 and 1 both have which I have already been visited.

(Refer Slide Time: 4:13)



So, I start moving back. So I back track from 2 to the most recently visited one which is 1 and see whether 1 has anything more to be done. But 1 has nothing more to be done because 1 had only 2 neighbours I have visited both of them.

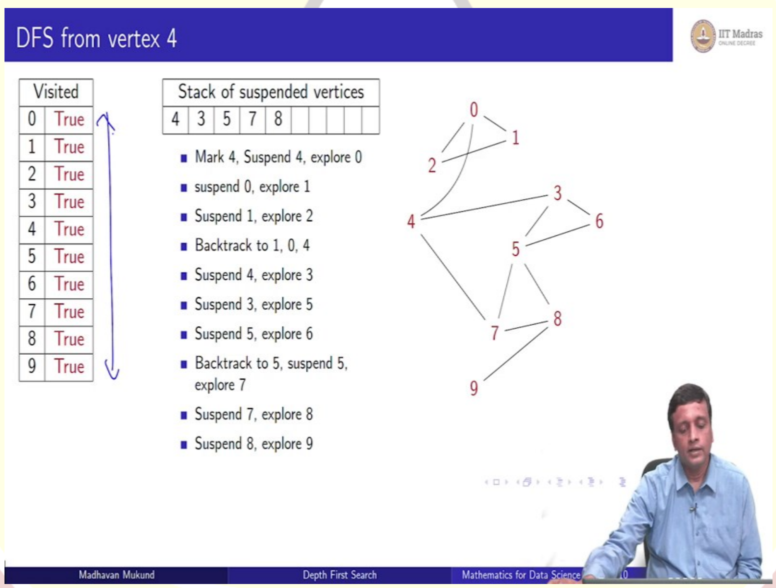
(Refer Slide Time: 4:26)



So, I back track to the previous one which is 0. So I ask whether 0 has anything more to be done. So notice that from 0 I have visited 1 and from 1 I visited 2. So though when I started with 0 I had not seen 2 yet by that time I have come back to 0 through back tracking 2 has already been visited. So at the current state of 0 it has no unvisited neighbours. So 4 was visited before it came to 0.

1 was visited from 0 and 2 was indirectly visited from 1 and therefore is no longer available. So I have to back track from 0 as well. So I back track from 0 back to 4 and now I ask whether 4 has anything interesting to say. And 4 says yes I have an another vertex to visit which is 3. So I explore this 3 and I suspend the 4. So now I have finish this whole section and I have come here.

(Refer Slide Time: 5:21)

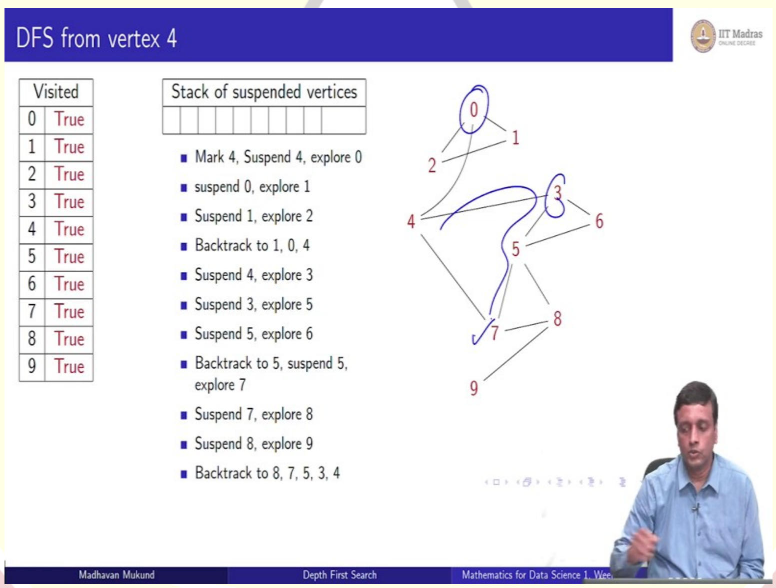


So now from 3 I have two neighbours 5 and 6. And I saw a suspend 3 and maybe I explore 5. And then from 5 I have I mean 2 unexplored neighbour 6 and 7. So maybe I suspend 5 and I explore 6. So now when I come to this point I have 4 which triggered 3 which triggered 5 which triggered 6. So I have 4, 3, and 5 on the stack and 6 has no new neighbours to explore because both 3 and 5 are already visited.

So, I will back track from 6 to the previous one which is 5 and ask whether 5 has anything more to do. Well, 5 does it has 7 and 8. So I will perhaps pick 7, so I will again suspend 5 this is the new suspension of 5 first time I suspended it because I have explored 6 when I came back to 5 and I have suspended it again to explore 7. 7 has a neighbour 8. So I will suspend 7 and explore 8. 8 has a neighbour 9.

So, I will suspend 8 and explore 9 and at this point if you look at the visited matrix everything this list, everything here has been marked true. But I still have this long queue a long stack of things which I have suspended. So I have to make sure nothings is missed out. So I will now look at 9 and 7, 9 has nothing left explore because it only had 1 neighbour 8 from which it came.

(Refer Slide Time: 6:47)

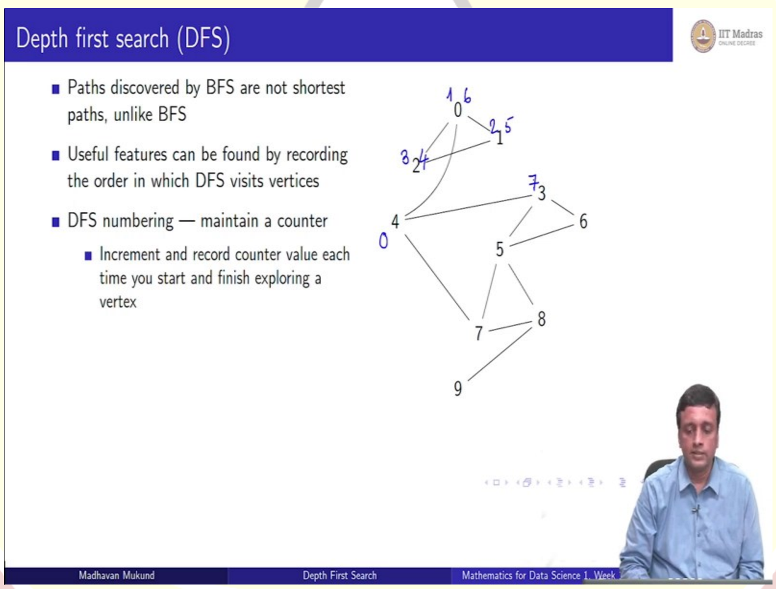


So, that is already marked. So I will go back to 8 but 8 has nothing more to do because 5 and 7 were also marked but 8 was triggered by 7. So I will go back to 7, 7 was triggered by 5, so I will go back to 5, 5 was triggered by 3 so I will go back to 3, 3 was triggered by 4 so I have come back to the empty stack. So I have nothing on the suspended list I have also now like before from 4 I explored 0 and 1 but indirectly through this 6 I have already explored 7.

So, 4 also has nothing to be done. So I say 4 is terminated and I quite. So this was how depth first search work. So in a way you can imagine depth first search the way you kind of browse on the internet. You start reading something interesting then you click on a link because you sees interesting you follow that and before you know where you are you had started reading something and you are somewhere far away.

So, then you have to go back follow back this links and go back to where you started and continue. So that is more or less what depth first search is doing. So you find the first interesting vertex and you go there. You keep distance suspension go there. Then find the first interesting vertex go there and so on.

(Refer Slide Time: 7:43)



So, depth first search finds these long paths like we said the path it found from 4 to 7 was it said that 4 triggered 3 in case we could have kept this parent information which we did not but for everything which is marked as visited like in the breadth first search. You could also mark in depth first search why it was marked visited. So you could keep this parent information. So we said that 4 marked 3, 3 marked 5, 5 marked 6 then came back and then 5 marked 7.

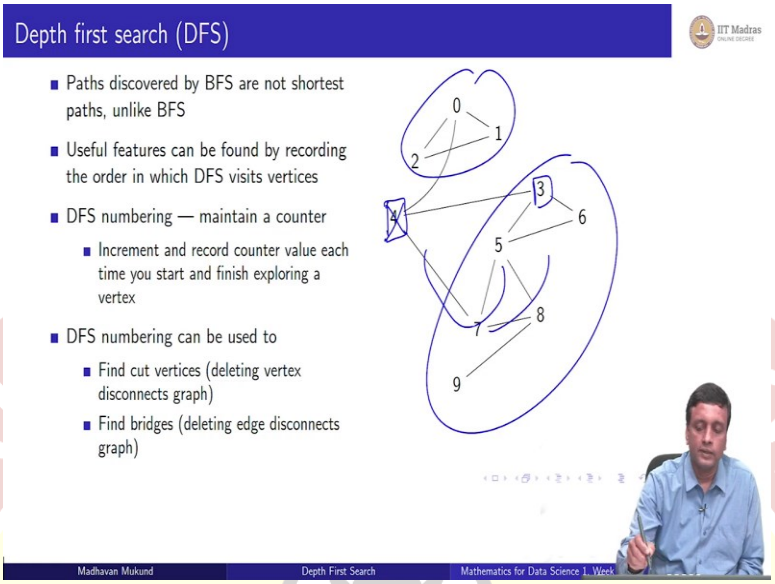
So, we found this long path 4 to 3 to 5 to 7 if we are kept the parent information you would have said parent of 7 is 5, parent of 5 is 3 and parent of 3 is 4. But this is obviously not the shortest path. So it does not do what breadth first search does not terms of shortest paths. So it seems a bit strange that we use this kind of indirect way of exploring when we have something which seems to be a better one namely breadth first search.

It turns out that actually depth first search is very useful for other things. So one thing that we can do in depth first search is keep track of how we visit these vertices. So what I can do, we will do it formally next time but informally I can say that we keep a counter. So we say that when the counter is 0, we entered 4. When the counter was 0 we entered 4 and from 4 we entered the vertex 0. So at this point the counter became 1.

From 0 we entered 2, entered 1 so at this point a counter became 2 from 1 we entered this. So this point the counter become 3. Now, we finished because the nothing to do at 2. So we finished 2 at the counter was 4, when we finished 1 the counter is 5, when we finished 0 the counter is 6. And now we have come back to 4 so now from 4 when we explore 7 the 3 the counter is 7 and so on. So in this way we can keep a counter incremented every time we enter a vertex for the first time and record it against that vertex as the incoming number of that vertex.

And then every time we finish processing a vertex that is all its edges have been visited then all its neighbours have been visited then we mark it saying that this is when we finished it. So we started 1 vertex 1 at counter 2, we finished vertex 1 at counter 5, we started vertex 0 at counter 1 we finished vertex 0 at counter 6. So notice that the counter value tells you something. It tells you that 1 was visited after 0 because 1 started at 2, 0 started at 1 and 0 finished after 1 because 1 finished at 5 and 0 finished at 6.

(Refer Slide Time: 10:19)

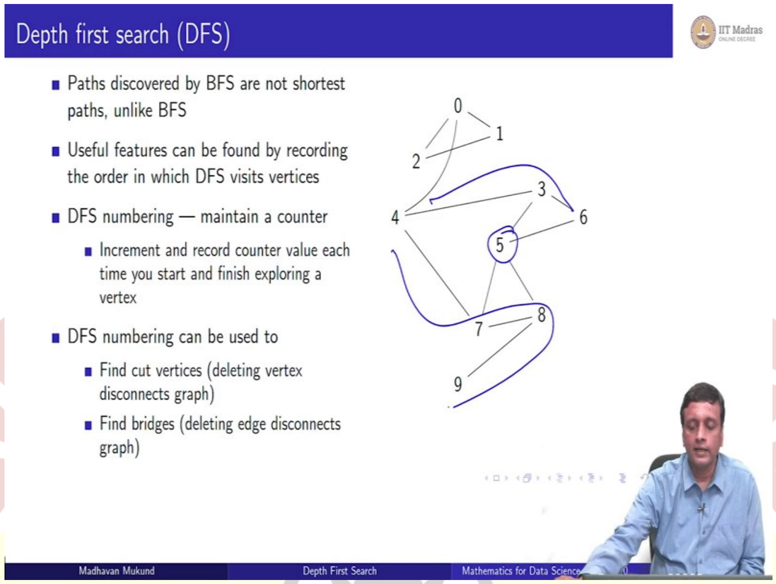


So, this DFS numbering is very useful and we can use it to find many interesting things. So there are some things that we talk about last time as problems on graph. We talked about coloring, we talked about matching and so on. But sometimes you also want to finds special vertices. So for example, supposing there is a critical vertex in your graph. So imagine an electrical network.

Supposing there is one power station which is very critical that anything happen to this power station then the entire electrical network will get divided into 2 disconnected portions. So this is called a cut vertex or sometimes called an articulation points. So in this network for example if this is a airline network. And we are imaging that some airport is unavailable because of bad weather.

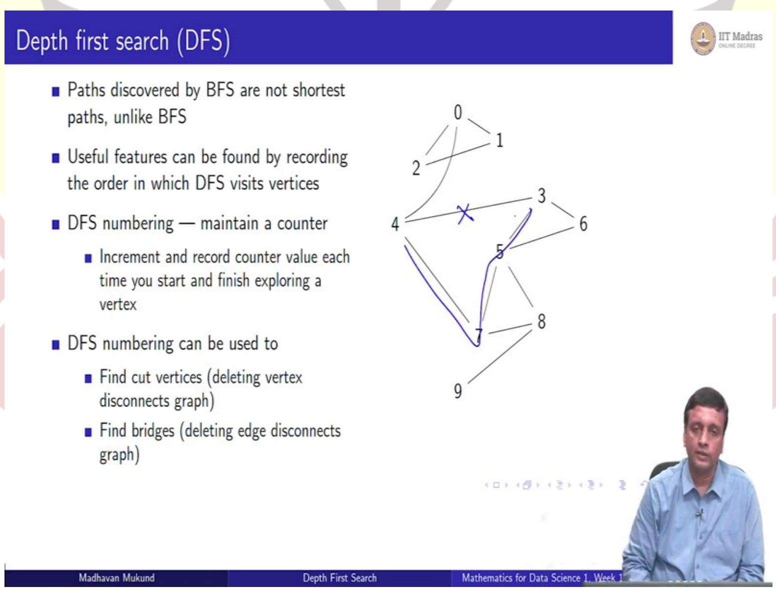
So, supposing there is a cyclone, so supposing vertex 4 the airport at city number 4 is knocked out of service because of a cyclone then you can see that there is no way to get from these vertices to these vertices. So we can still travel between 3 to 9. We can still travel between 0 1 2 but we cannot travel from 0, 1, 2 section to this. So now the graph has become disconnected. So would we discover that 4 is such a vertex. For instance, 3 is not a such a vertex if I remove 3 the graph still remains connected by this outer path.

(Refer Slide time: 11:37)



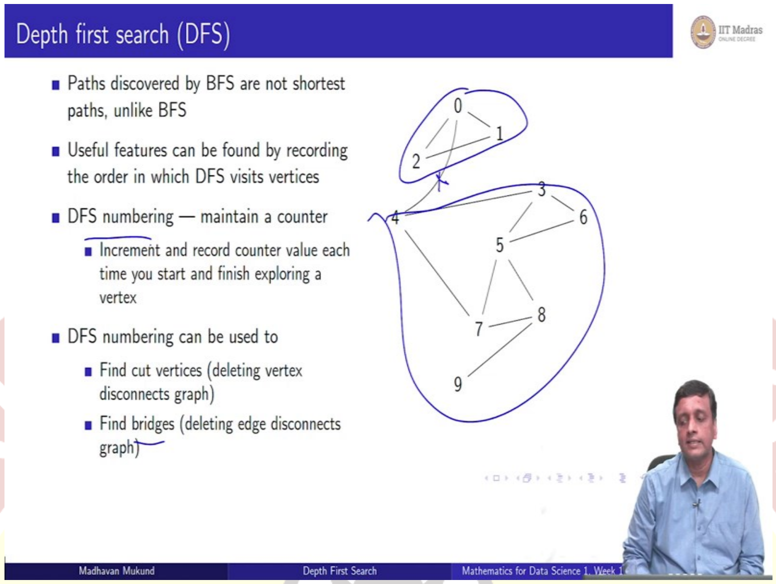
So, this 3 is not a, so if I remove 3 there is no (())(11:39). If I remove 5 also there is no problem, I can still go from 4 to 6 this way and I can come to these, this way. So which are the articulation points of cut vertices, you can discover that using DFS numbering. A related thing is this is for vertices, related question is for edges. So are there edges which are critical for me.

(Refer Slide Time: 11:59)



If there is a root which I cannot follow will that disconnect things for me? So now you can say that this route is okay because it is not a critical route because of if I cannot go from 4 to 3 directly I can still follow an indirect path and go from 4 to 3.

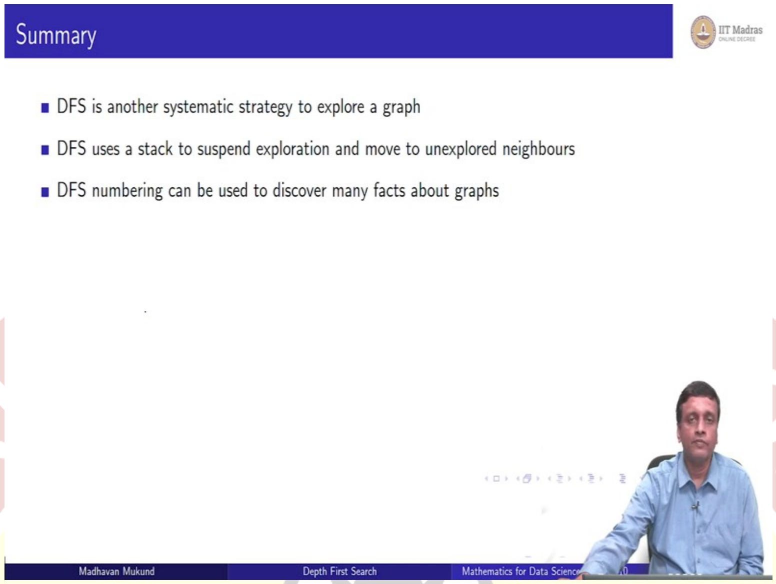
(Refer Slide Time: 12:16)



But again if I knock off this one, if I knock off the route from 4 to 0 then it turns out that 0 1 2 is disconnect from the rest of the network so this is called a bridge. So these kinds of properties of graphs, these cut vertices, bridges and many other interesting things can be discovered as a byproduct of depth first search.

So, you do a depth first search then you do this DFS numbering which we will describe in a later class. And using this DFS numbering you can actually find out interesting properties of the graph. So that is why depth first search though it does not find shortest paths it finds out interesting structure within a graph.

(Refer Slide Time: 12:51)



So, to summarize DFS is a different strategy from breadth first search, is another systematic strategy to explore a graph. So what we do is we start for the vertex, suspend it go to an unvisited neighbour, suspend it go to an unvisited neighbour and so on. And in order to keep track of these suspended vertices and how to resume them in a systematic way we use a stack. So we use this last in first out data structure in order to keep track of how to resume vertices when we come back from a terminated computation.

And we saw that although not in detail just informally, we saw that if we keep track of the sequence in which we visit vertices, when we finish, when we entered, when we finish we get this DFS numbering and with this DFS numbering we can actually uncover some structural properties of graphs which are quite interesting.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

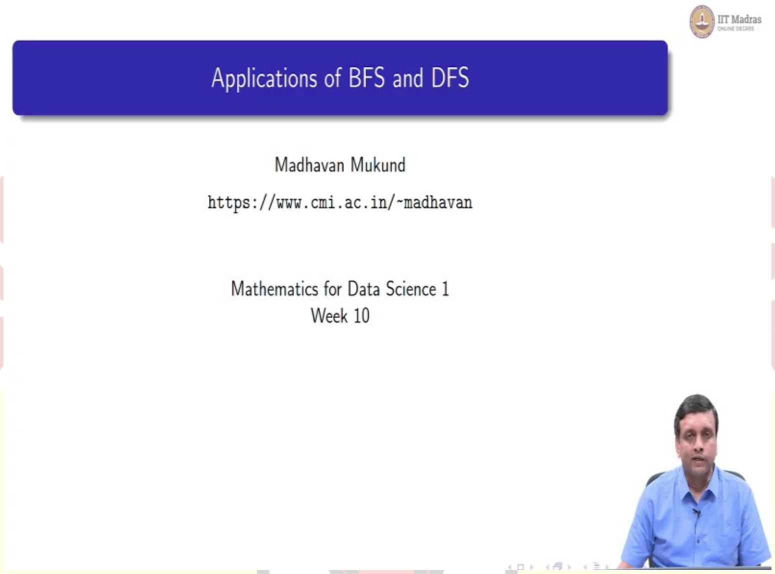
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

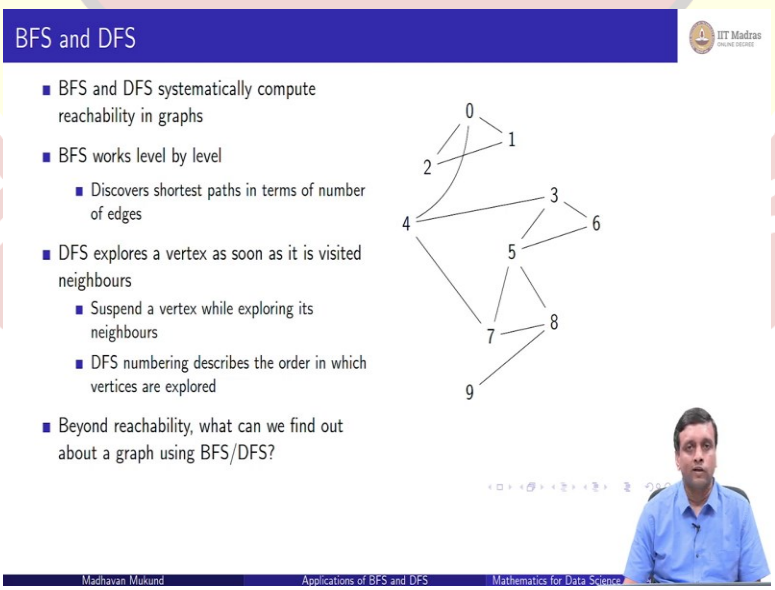
Mathematics for Data Science 1
Professor Madhavan Mukund
Applications of BFS and DFS-1

(Refer Slide Time: 0:15)



So, we have looked at breadth first search and depth first search as two ways to systematically explore a graph.

(Refer Slide Time: 0:21)



So, what we are going to look at now is how to go beyond just reachability, so what we have done so far is starting with a vertex how to find out what all we can reach in that graph, and we said that BFS and DFS are two systematic strategies to do this. So when we do BFS, we do it level by level, so we start with its vertex, we go to its nearest neighbors, then from those nearest neighbors we go to the next set of neighbors and so on.

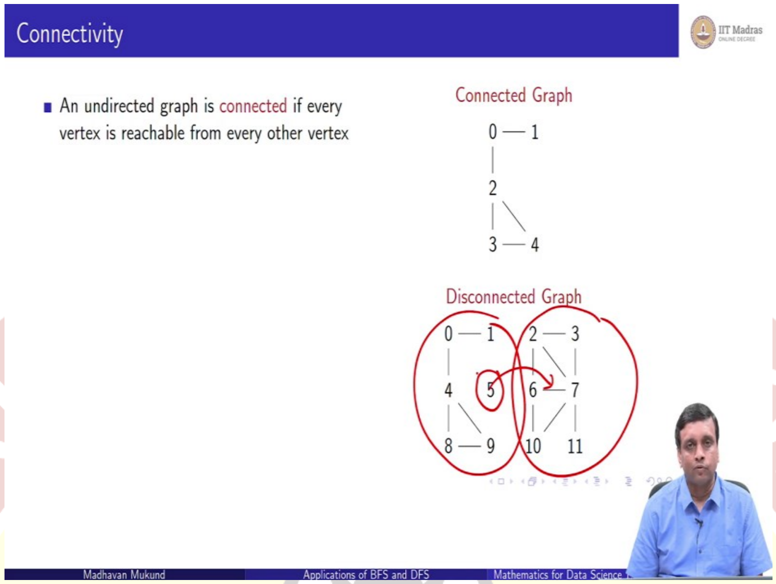
So in this process, one of the things we said is that BFS will discover the shortest path, the shortest level for every reachable vertex because it is processing the graph layer by layer in some sense, everything which is reachable in one round, in two edges and three edges and so on. And of course the whole point of calling it breadth-first search is that we need a systematic way to do this, so we had this queue which kept track of how to make sure that we explore all the vertices in this level by level order.

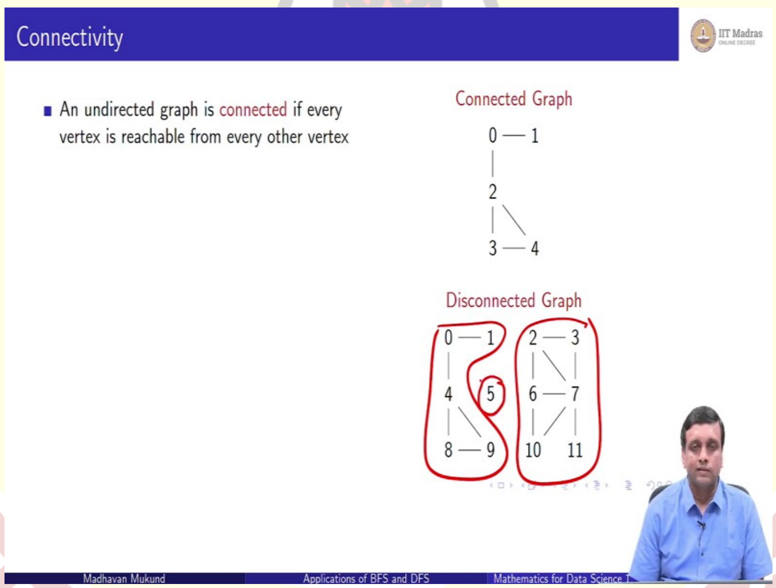
So everything at level one is put into the queue and it will get processed before everything at level two and this will guarantee that everything is reached in the shortest number of levels. Now DFS was a very different strategy, it was in some sense an aggressive strategy, the moment it moved to a neighbor, it would suspend the current vertex and then it will start exploring the neighbor. I mean in a way it is a bit like how when we start looking up information on the internet, right?

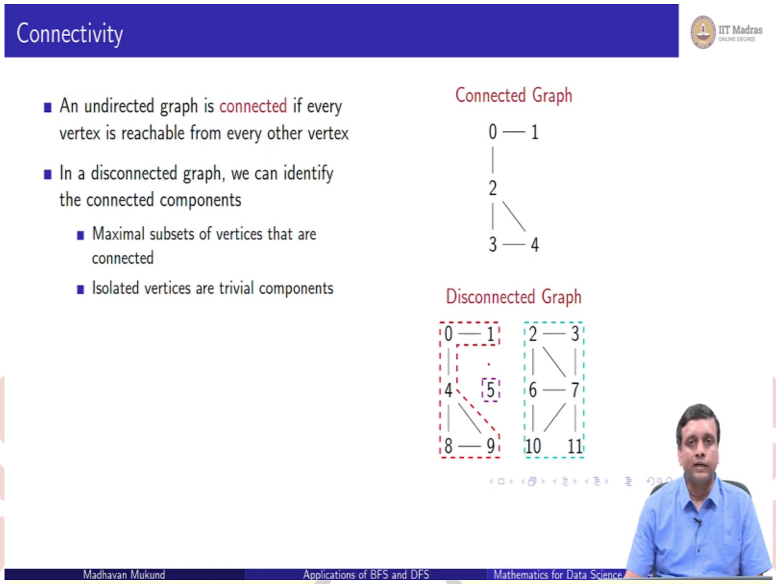
So, we start looking for something and then before we finish reading the article we find an interesting link and we go to that link and we start reading that link, that has another link and we go to that link and so on. So eventually we have to remember to come back to where we were reading in the first place.

So DFS is like that, you start with a vertex, look at any vertex that is neighboring it which you can explore, go down that path and only when you run out of things to see you come back to the original vertex. So you keep this stack of suspended vertices in DFS. So the question that we are going to address in this lecture is what more we can do then just reachability with BFS and DFS. So, is BFS and DFS only for reachability or are there more interesting things that we can do?

(Refer Slide Time: 2:16)







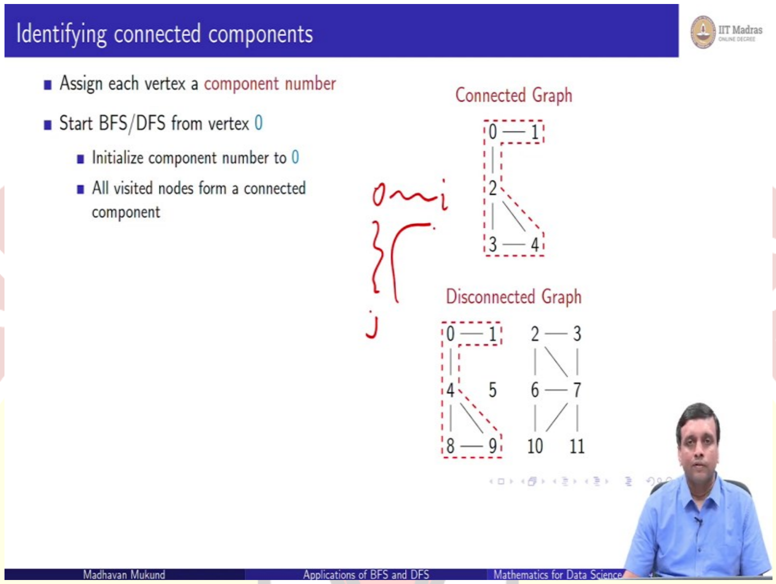
So, first aspect of a graph that we will explore is that of connectivity. So we say that a graph, an undirected graph is connected if every vertex is reachable from every other vertex. So you can on the right two graphs, the first graph is clearly connected from any vertex of 0 1 2 3 4, you can go to every other vertex. But if you look at the bottom you have clearly two disconnected components.

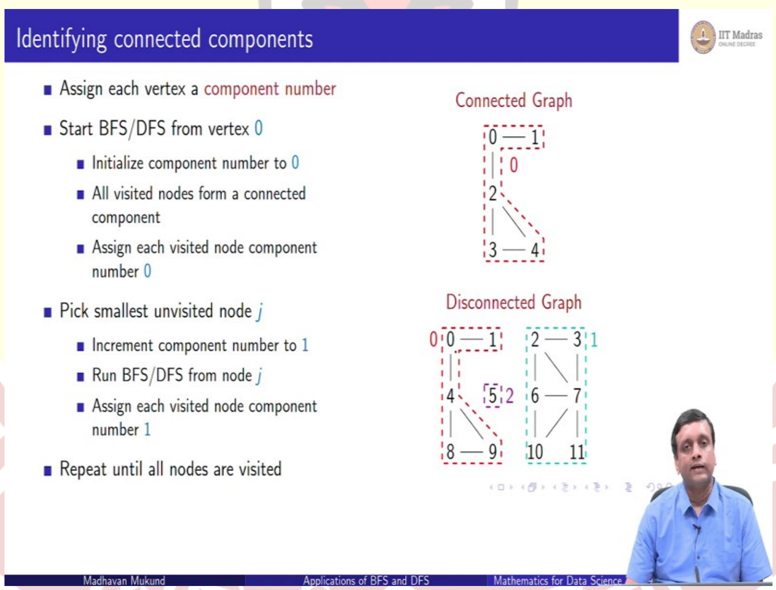
So, you have this left hand side component and you have this right hand side component and there is no way you can go from here to there and in fact 5 on its own is also isolated from everything else. So that is not a component on its own, so really technically you have these components which are, so you have this component, everything here is reachable from itself. Here this component, everything inside this component is reachable from within that component and finally we have this component which consists of just one vertex

So, in a disconnected graph we can identify these components. So what we want to do is put these red border around this component and this blue border around the second component and then find also in particular these isolated vertices which are what you might think of is trivial components. So we said that technically there are no self-edges in a graph, we said that we do not assume there are edges from i to i.

So, when we say a single vertex is a component we are just saying that it cannot reach anything and so there is nothing else that it can be connected to which can come back to it. So these are the trivial components. So our goal is to see how BFS or DFS can help us identify these components.

(Refer Slide Time: 3:54)





So, when we are doing one of these, either BFS or DFS, what we do is, just like we kept track of extra information like in BFS we kept track of the level number and the parent information and even in DFS we said we could keep track of parent information, so we will have one more component number that we keep track of. So we have a number which we are going to use to label these components, so we are going to call it component 0, component 1, component 2 and so on.

So, we have a component number and we attach component numbers to vertices. So initially we want to do it for the whole graph, so we might as well start at vertex 0, we want to find out how many components are there in this graph. So we start at vertex 0, with either BFS or DFS it does not matter, and then this new quantity that we are going to assign to vertices which is the component number, we initialize this quantity to 0.

So now, when we do a BFS or a DFS from vertex 0, we will reach some vertices and all of these vertices will be connected because they are all reachable from 0 and therefore they are reachable from each other, in fact can reach it from 0, if I can reach say 0 to i, and I can reach 0 to j, then I can go from j to i by going back to 0 and then because these are undirected edges, you can always go backwards.

So, if I can go from 0 to some vertex i and I can go to 0 to some vertex j, then there is a connection from i to j also. So everything that you get in a single component, in a single scan of BFS or DFS is going to form a connected component. So what you do is that, while you are performing the scan you remember this component number is 0 and you just assign component number 0 to all of these. This is an extra piece of information that you keep just like we keep visited v, we keep component of v and we just keep assigning component of v equal to 0 for all these vertices.

Now, if you are in the connected graph like in the first case, all the vertices would have been covered. But if you are in a disconnected graph like in the bottom case, there are some vertices after you have reached everything that you can reach from 0, there are some vertices which are not yet marked as visited. So this means that they are not in component 0, they are in another component, so you have to find that.

So, you have to pick any one of them, you have to pick any one of the vertices which are not visited and start a new breadth-first or depth first search from there. So in particular let us pick the smallest one, so supposing we identify the two in the second graph, in the first graph there is nothing to do because everything is already being visited, but here we identify that vertex 2 is a candidate to start of a new BFS or DFS because it was not visited when I started from 0.

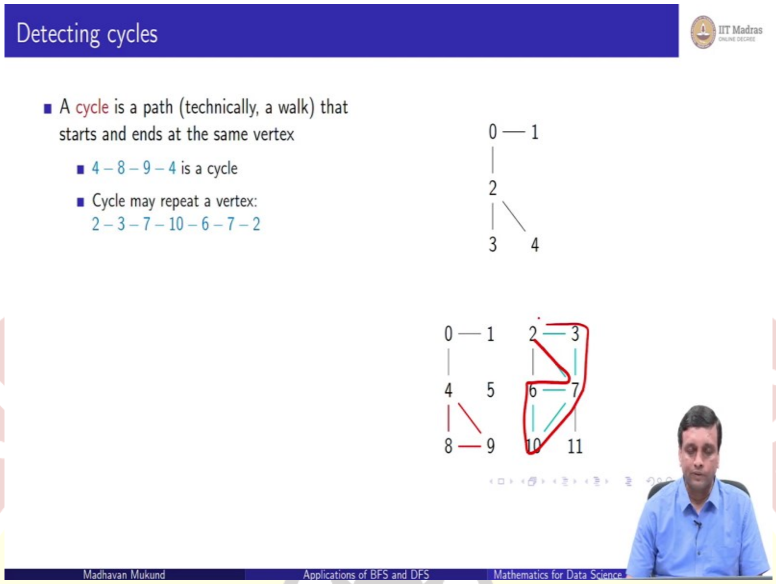
But now this is going to be a different component. So we cannot call the component that we are going to discover starting from 2, the same as the component we already discovered. So we have to increment the component number, so now we are looking at vertices which will be called component 1. So we perform this breadth-first search or depth first search, we will find that all those 6 vertices on the right hand side are reachable from 2. And as we are done with the first component while we are doing this we will attach the current component number which is 1 to all of these.

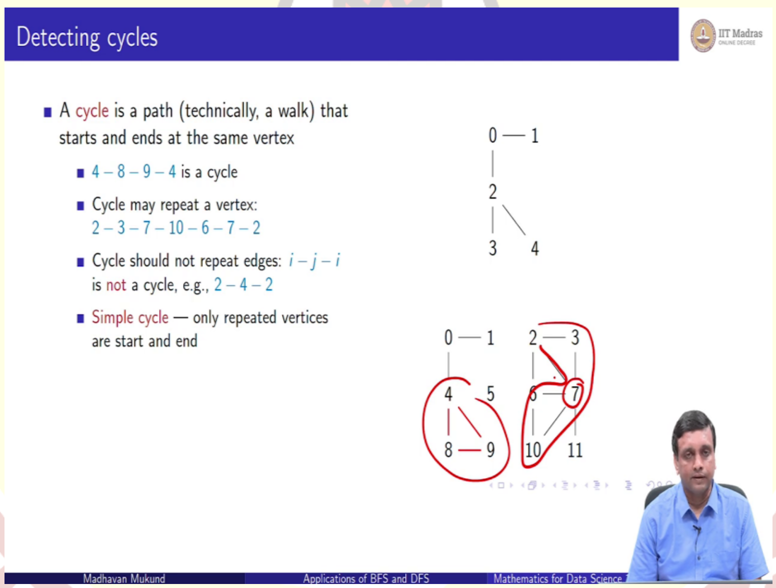
So, now after two DFSs or two BFSs I have identified component 0 and component 1, but at this point I still have an unvisited vertex, so I will repeat this. So in general there may be many components, so each time I finish one component I will look at the remaining unvisited vertices and start yet another DFS or BFS from safer instance the smallest numbered vertex in that set.

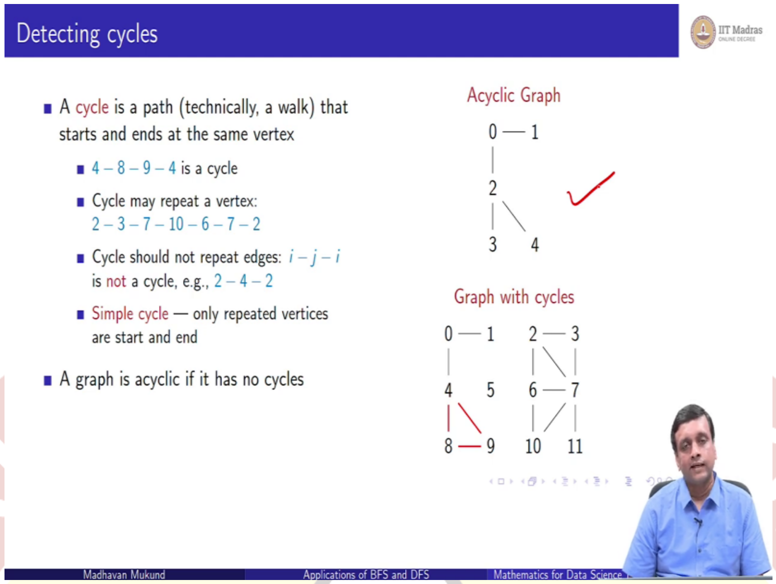
So, we just keep repeating this, in this case we only have to do it one more time because we have only one such vertex, vertex 5 and because we cannot go anywhere from 5 this BFS rapidly stops and we get a component 2. So remember each time we repeat we also increment the component number, so we increment the component number and start a fresh scan. So this is how we can build in a component discovery algorithm to BFS or DFS.

So, while we are doing BFS and DFS we can also discover what are the connected components in the graph and label them. So we now know at the end of this, that for example 9 and 4 in the bottom graph are both in component 0, so they are in the same component, whereas 9 and 10 have different component number 0 and 1, so 9 and 10 are not in the same component. So it is important that we can at the end of this look at two vertices and decide are they connected to each other or not without having to start another breadth-first search from those vertices and then try to again connect them.

(Refer Slide Time: 8:27)







So, related to breadth-first search and depth first search is also the idea of a cycle. So a cycle as you could imagine is something which is circular, so a cycle in a graph is a path which starts at some vertex and then comes back to that same vertex. So remember that we said technically that a path does not repeat vertices, so technically this is a walk because we start at a vertex, traverse some edges and come back to the same vertex.

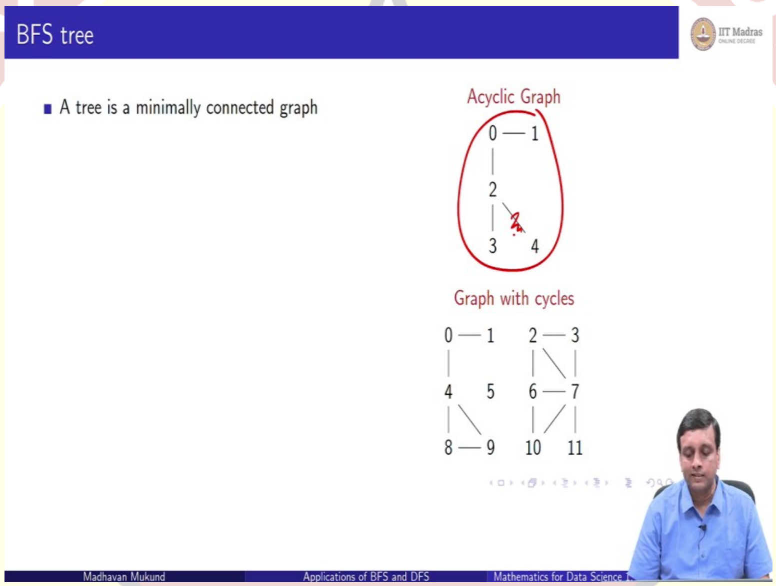
So, if you look at the bottom graph for instance, 4 then 8 then 9 and then back to 4. So this is a cycle. Here is a more complicated cycle, 2 to 3 to 7, so I start at 2, then I go to 3 and then I come to 7, but then I go to 10 and then I come back to 6 and then I come back to 7 and then I come back to 2. So in this case the walk not only repeats the starting and the ending point 2 but it also repeats 7 along the way. So this is also called a cycle.

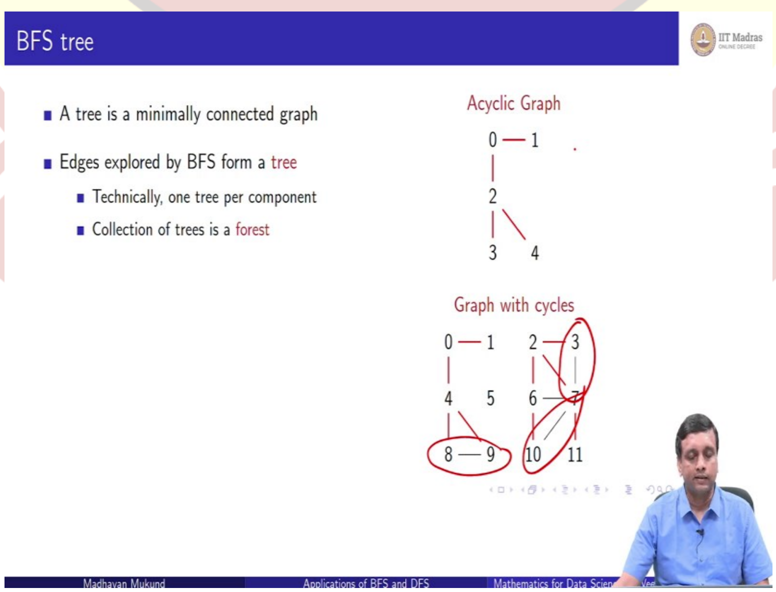
But though you can repeat vertices, you cannot repeat edges, so you cannot claim that this is a cycle, I went to 2 and then I came back, this is not a cycle. So a cycle cannot go back and forth along the same edge, otherwise every edge will become a cycle and we do not really intend that. So finally what we are typically interested in a what are called simple cycle. So simple cycle is like this one.

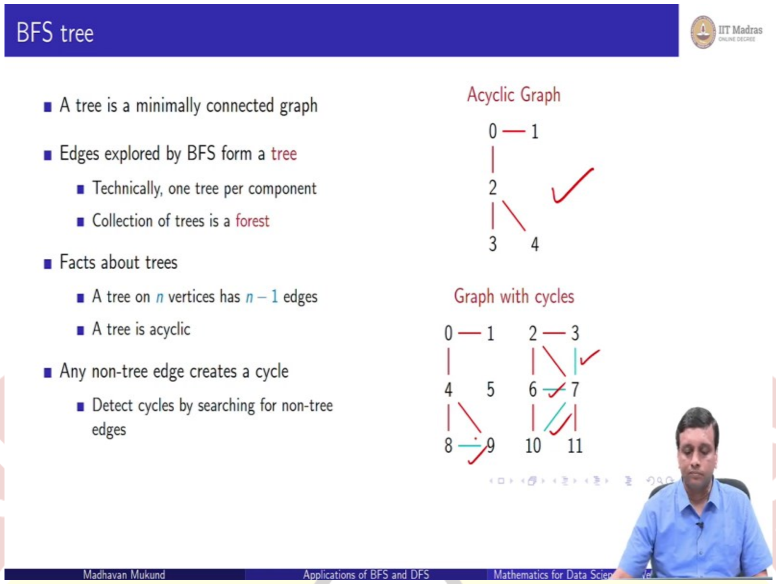
So, this one is a simple cycle because it went from start to end and came back to start rather, so the only vertex that was seen twice was a starting vertex when we closed the cycle. Whereas, this one was not simple because we went down and then we came back up and then we visited this 7 twice, so it is actually two simple cycles which have been joined at 7. So we can do 2 3 7 and then we can do 7 10 and 6 7.

So, if a graph does not have any cycles, then it is called acyclic. So this graph is acyclic because there are no cycles in it. Whereas, the graph on the bottom is not acyclic, so we do not really call it acyclic graph, we only interest, we are interested in whether it is acyclic or it is not acyclic. So these graph has cycles and what we would like to do now is say that if a graph has cycles how do we find these cycles. That is our goal.

(Refer Slide Time: 10:51)







So, we start by looking at what is called as tree. So tree is a minimally connected graph. So here look at this example here, so this acyclic graph that we drew, so it is connected because we saw that it is one connected component. It is also acyclic and in particular now if we drop any edge like for example if we drop the 2 4 edge, then this graph will become disconnected. So if I want to minimally connected, I need to draw at least these many edges in this case there are five vertices, 0 to 4 and I have drawn four edges.

So, it turns out that when we explore for instance the tree using BFS, then the edges that we use to visit new vertices, so I start at a vertex and I visit a neighbor if the neighbor is not visited, so then I will count that edge as being visited by BFS. So the edges that BFS visits actually form a tree. So if you look at this acyclic graph on the top, it actually visits all the edges because the graph itself is a tree and there are not fewer edges which form a tree in that case.

Now, if you have a graph like the bottom one which has multiple components then technically you have to start BFS each time from a new component. So it is not one tree because the tree as a whole would connect the entire graph, but the entire graph is not connected. So each component gets connected by a tree. So here for instance BFS does not visit this edge, so this edge is outside the BFS tree.

Similarly, here, this edge and this edge are outside the BFS tree. Now from English since we say that a forest is a collection of trees, in this kind of a thing also we talk about multiple trees are forming a forest. So technically what BFS does is it discovers a collection of trees or a forest inside the graph. So some useful facts about trees, so the first fact is that if you have n vertices then a tree on n vertices will have n minus 1 edges.

Now, it does not specify which edges, so you can connect them in many different ways, but whichever way you connect them you will have to use n minus 1 vertices. So just as an example supposing we have four vertices, so one way to connect this in a tree is to connect it like this in a single path. So this is one way to do it. And of course we can choose different paths, so we could connect it this way or we could connect it this way and so on.

But a different way to connect it which is not a path is to connect it all to one vertex like this, so this is also a tree. But notice that all these ways of connecting these four vertices, so that each one is connected to the other but in a minimal way, all of them have three edges. So whenever you have n vertices, you will have exactly n minus 1 edges and of course the tree is acyclic. Because if you just think of it as a minimally connected graph, then if it has a cycle for instance if I had another vertex, another edge from here to here, then I have two ways of going from all the vertices on the cycle.

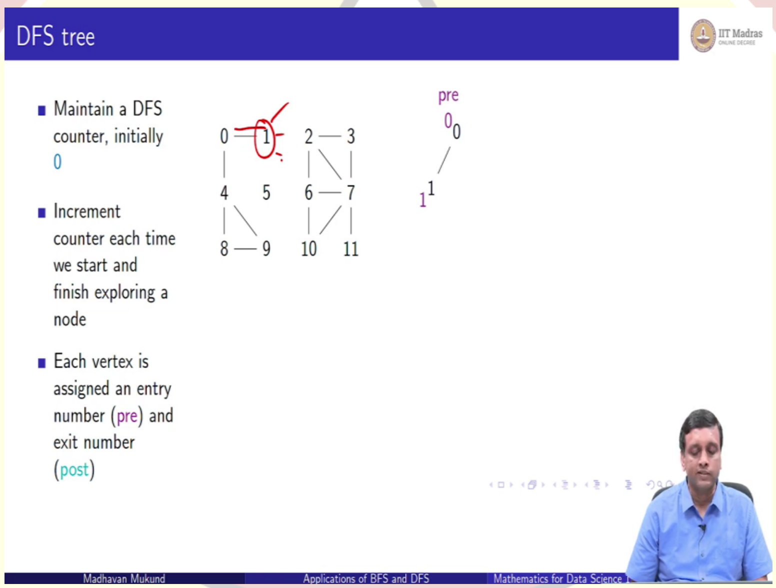
Two ways of going from 2 to 1, two ways of going from 0 to 4 and so on. So if I delete one, for example now if I delete this vertex or this edge, then I still have a connected graph. So if I had a cycle, then it cannot be minimal because some edge along the cycle can be removed and I will still be able to reach everything by going around the cycle the other way. So we are now going to prove this fact, but it is useful to know that these are all equivalent ways of thinking of a tree, so tree is a minimally connected undirected graph, a tree is necessarily acyclic.

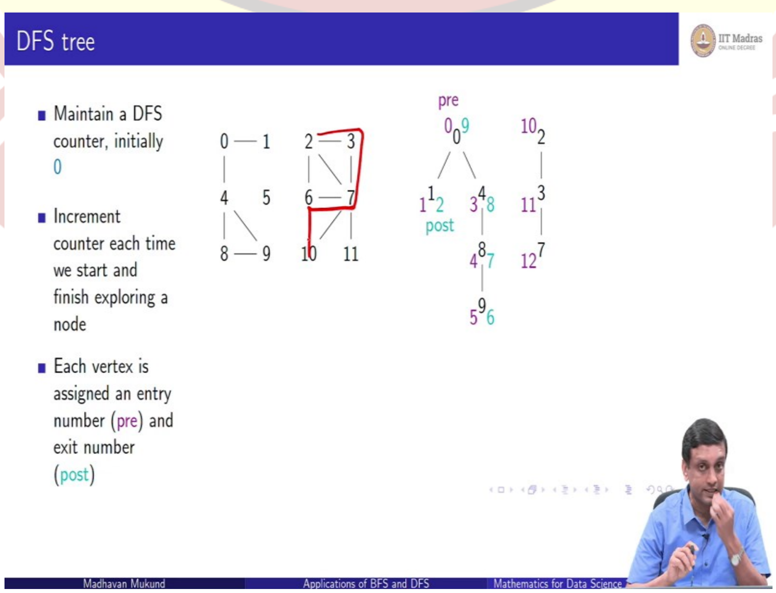
So, an acyclic connected undirected graph is a tree. And on the other hand if it is connected and it has n minus 1 edges, then it is a tree and if it is a tree, it has n minus 1 edges. So just remember all these, because these are all different ways of thinking about a tree. So now coming back to our question, our question was how do we detect whether the graph we have is acyclic or not. So what we saw is that we have acyclic as like the first one, then the BFS tree that we get covers all the edges.

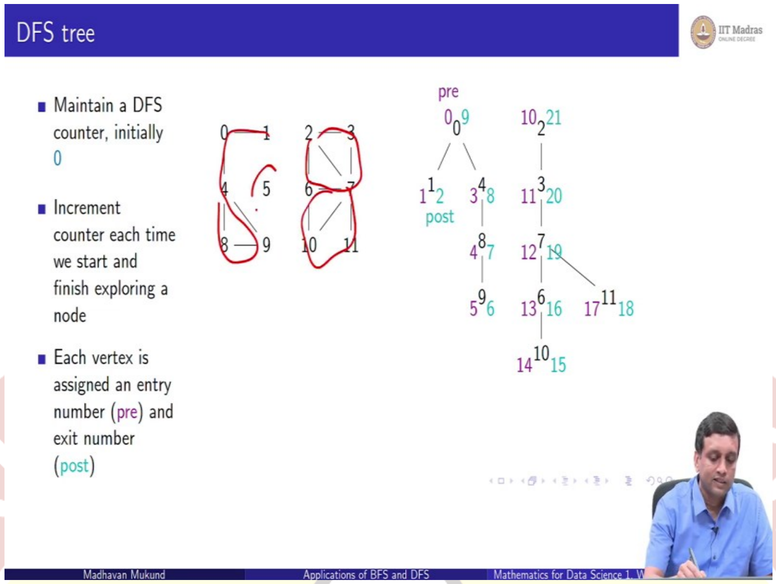
So, if we do not cover all the edges on the graph through our tree, if there are non-tree edges, then those non-tree edges if we add them must form a cycle and that is exactly what we get, so we can detect a cycle by searching for non-tree edges. So as we are going along just like we mark vertices as visited, we can also if we want mark edges as visited or we can keep track of them by looking back afterward we have done it to see which edges we use, say the parent thing also gives us the visited.

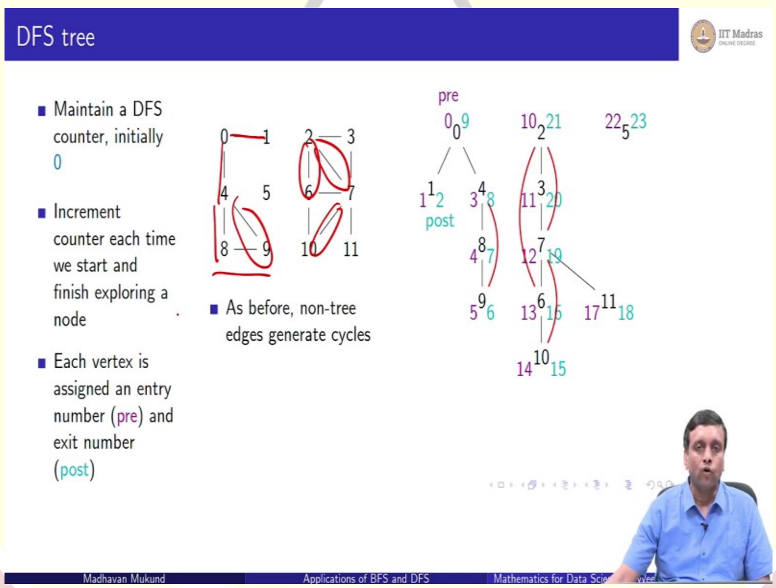
So, if I say parent of i is j, that means I went from j to i to visit i, so therefore ji was a tree edge. So whichever way we can recover the tree edges at the end of our BFS and then any non-tree edge if it exists must form a cycle. So here we have these non-tree edges, so 6 7; 3 7; 7 10 and 8 9; which are not part of the BFS tree that we constructed. So since there is at least one such edge there must be a cycle in particular in this case both these components have a cycle.

(Refer Slide Time: 15:53)









So, we could do this using DFS also in the same way, but we mentioned last time that DFS also comes with a different type of strategy to keep track of it called DFS numbering. So let us build a DFS tree for this same graph and let us describe formally how this numbering works through an example. So we initially maintained a separate counter for numbering when we enter and exit a vertex and we will start with 0.

So, we increment this counter every time we enter a vertex, every time we start exploring it and every time we leave the vertex, every time we finish exploring it. So this will become clearer as we go along, so in this process, this counter value is assigned to the vertex as an incoming number and then when we leave the updated counter value is assigned as an outgoing number. So we have a pre number and a post number for every vertex.

So, let us try and do a DFS for this graph here, starting at vertex 0. So we initialize our counter to 0, so we are starting, so the black indicates the vertex number and this purple number is the pre number, a number of our DFS counter when we entered vertex 0 was 0. Now we explore for instance 0 to 1. We have to explore the unvisited vertices which are neighboring 0, so we can do 1.

So, now we increment our counter and say and we entered vertex 1 with counter 1. So now when we come to vertex 1, we have no further vertices from 1 which are unexplored, the only neighbor of 1 which I can go to is 0 but 0 is already been visited, so now I am going to leave 1. But before I leave 1, I will increment the counter, so I will leave 1, I will assign it the post number of 2, and now in my stack I am coming back to 0. So I am back at 0.

So, 0 is not finished because 0 has another neighbor which is unexplored which is 4. So I will increment the counter and enter vertex 4 with the pre number 3. So notice that this number is increasing 0, then 0 plus 1 is 1, then 1 plus 1 is 2, then I went back but did I, I did not leave 0, I am still processing 0. So I will assign a post number to 0 only when I have finished all the neighbors of 0, in this case I am not finished, I have gone down another path.

So, 0 1 2 and now I assign the pre number 3 to this vertex. So now I am at 4, so 4 has two unexplored neighbors other than 0, because 0 is already explored. So I go to the smaller one say 8, so again I increment my number from 3 to 4 and I enter 8. So I enter 8 with the pre number 4 and now from 8 I can go to 9, because 9 is not yet been visited. So from 8 I enter 9 with pre number 5, so each time I am just incrementing this one counter, the pre and post is the same counter is being incremented whether I go in or go out.

So, at 9 I get stuck, because I have only two neighbors 4 and 8, both of which have been visited. So 9 now terminates, saying I have finished processing 9, so I increment to 6 and I get out of 9 and comeback to 8. At 8 again I have nothing more to do, so I increment to 7, get out of 8 and come back to 4. At 4 I have done 8, but I cannot do 9, because 9 was visited through 8, so 4 is also done, so I will assign my post number to 8 and get out of 4.

And now I come back to 0. Now at 0 I had two neighbors 1 and 4 both are done and I am now finished with 0, so I will exit 0, so this is now my first component that I have discovered, starting from 0 and coming back to 0 and I entered and exited each vertex by updating that counter. Now I go to an unmarked vertex 2 and I continue the same numbering, so I enter 2 with vertex, with pre number 10, from 2 I go to 3 with pre number 11, from 3 I go to 7 with, so I am following this path.

So I have the smallest neighbor, then the smallest neighbor, then the smallest neighbor and then I will get stuck. So 2 will go to 3, 3 will go to 7, 7 will go to 6, 6 will go to 10 and then at 10 I am stuck, because I cannot go back to any vertex which is not visited, so I will exit 10 with counter 15. Then I will come back to 6, again 6 has neighbors 2 7 and 10 all of which have been seen, so I will exit 6.

I will come back to 7, now 7 has another neighbor 11 which I have not yet explored. So when I come back to 7, I am not done, instead I start exploring vertex 11 with counter 17. Now 11 obviously has nowhere to go after that, so I exit 11 with 18. Now I am done with 7, so 7 becomes exited with 19, 3 is exited to 20 and finally I come back to 2 and the other vertices which are neighboring 2 namely 7 and 6, have already been explored through 3, so there is nothing more to be done at 2, so I exit vertex 2 with 21.

So, at this point I have visited all these vertices and I have visited all these vertices. So this vertex 5 remains, so I have to start a third DFS as we did before. So I started with a new counter value 22, because I finished the last one with 21 and then I immediately exit because 5 has nothing to do. So this is, at this point we have not use in these numbers, we will see soon why we are going to use these numbers, but at this point it is just to show that when we are doing BFS we construct in a tree, when we are doing DFS also we construct a collection of trees is one for each component.

And it is now useful to actually describe the order in which this tree was drawn. How did we add edges to the tree and when did we back track up the tree? That is what we are keeping track of for this pre and post number. So ignoring the numbering, now there are some edges which did not come into the tree. So these are these red edges here, so we have an edge for example from 4 to 9, which did not come into the tree because our tree went 0 to 1, 0 to 4 and then 4 to 8 and then 8 to 9.

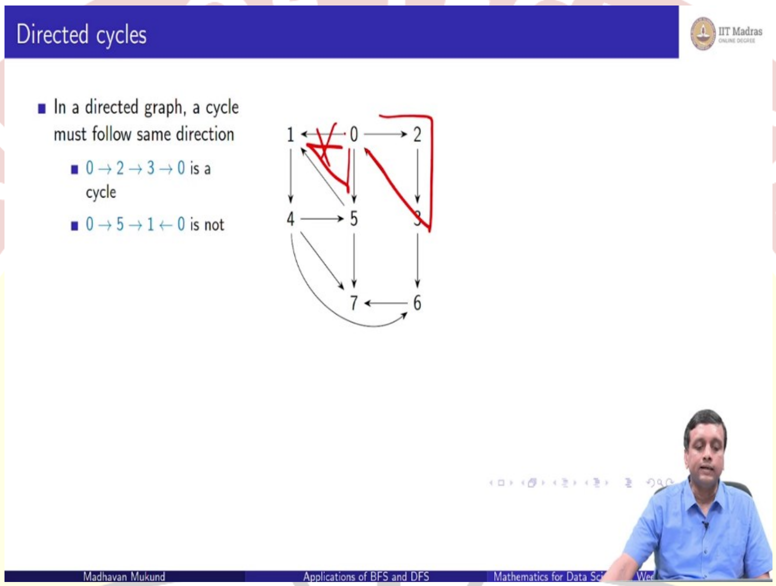
So, since I covered 9 through 8, I never got to explore 9 from 4, so 4 to 9 is a non-tree edge. Similarly, 2 to 7 and 2 to 6 are non-tree edges, because I did not explore them and so is 7 to 10. Because I went from 7 to 6 to 10, so I never explored the edge 7 to 10. So these 4 edges are non-tree edges, and each of them as you can see creates a cycle, so exactly like in the BFS, in an undirected graph, in a DFS also all the non-tree edges create a cycle.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Chennai Mathematical Institute
Lecture No. 65
Applications of BFS and DFS-2

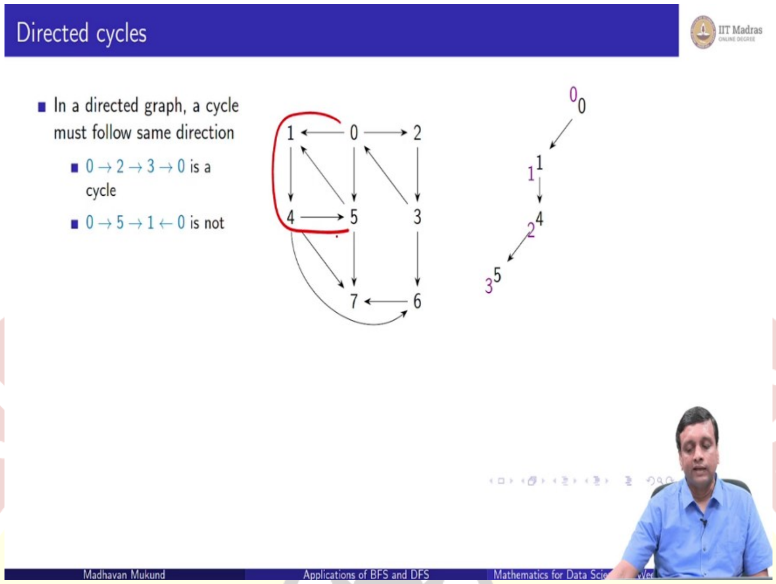
(Refer Slide Time: 0:4)

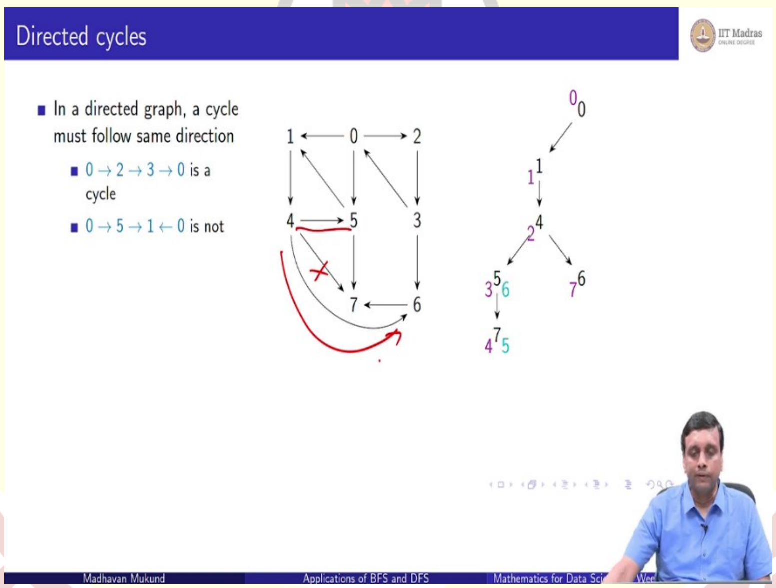


So, now what happens in a directed graph? So, in directed graph a cycle also is directed, that is I must go around from a vertex through a set of neighbors and come back, but following the same direction. So, it is like going around a circle in a one-way street, you have to follow the one-way street, you cannot suddenly go down in one-way street the wrong way. So for instance, I can go from 0 to 2 then I can go from 2 to 3, and then I can come back from 3 to 0. So, this is the cycle in the directed sense because I am going forwards at every step.

On the other hand, if I try to go from 0 to 5, and then from 5 to 1, and then I try to come back from 1 to 0, this is not allowed because 1 to 0 is not in the same sense. So, if I ignore the direction there is a cycle 0 5 1 0, but with directions there is no such cycle, so we are interested in directed cycles.

(Refer Slide Time: 1:08)



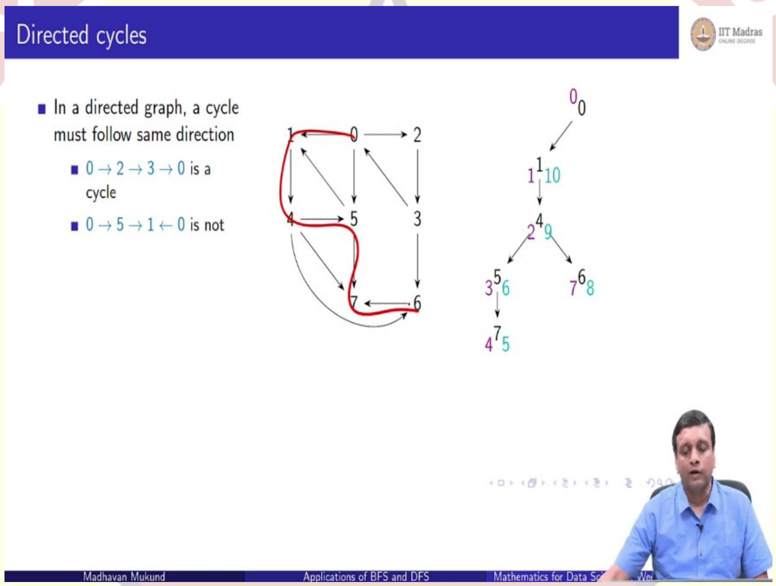


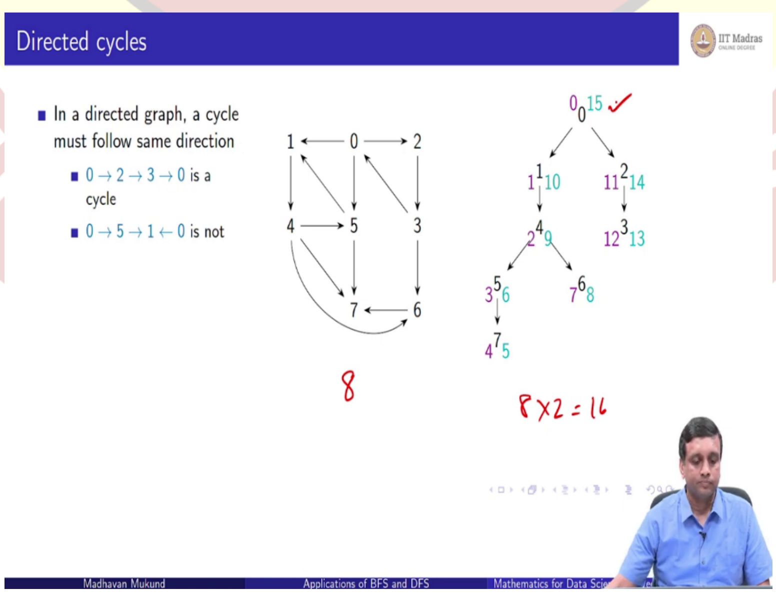
So, again we again do a DFS and do the DFS numbering, it is exactly the same, there is no difference in DFS whether it is directed or undirected, so we follow the same protocol for these pre and post numbers, so we start at 0 in this case. We are starting at 0 and then we are going to systematically explore its neighbors. So, we enter 0 with pre number 0, we enter its first neighbor with pre number 1 and now we explore 1.

So, we enter 4 with pre number 2, from 4 we have many ways to go, so the first way is 5, so we enter 5, so we have now come down this way. So, we have gone from here and we have reached here. So, we enter 5 with pre number 3, from 5 we can go to 7 with pre number 4. And now notice that 7 has no outgoing vertices, edges at all. So, from 7 I cannot go anywhere because it has no outgoing edges, so from 7 I exit with number 5. So, I have a post number 5 and I come back to 5.

Now I ask whether we can do anything from 5, well 5 had only two outgoing edges, back to 1 and forward to 7. But 1 was already covered and 7 has just been covered, so 5 also exits with number 6. Now I come back to 4, so 4 I explored 5, 7 is no longer available because 7 is already done through 5, but this long edge from 4 to 6 is there, so I enter 6 with pre number 7. Again from 6 the only thing I could have done is go to 7 which I have already seen, so I exit 6 with number 8.

(Refer Slide Time: 2:36)



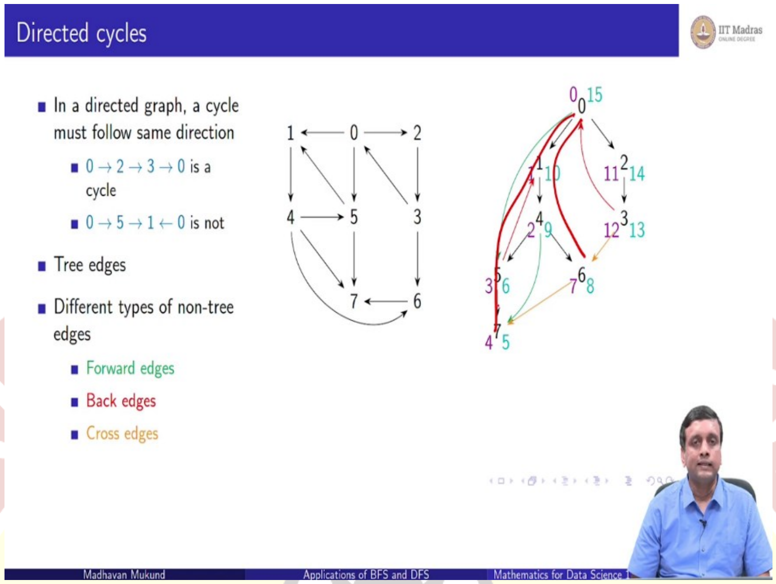


Now I am done with 4, so I exit 4 with number 9. Now I am also done with 1, so I exit 1 with number 10, because 1 had only one outgoing edge to 4. So, I come back to 0, so I in some sense I started here I went here, then in the process I explored everything on this side. So, now I go to the right side, so I explore the other neighbor of 0 which is 2 by entering it with number 11 because number 11 because I exited last one with number 10.

From 2 I can go to 3 with entry number 12, from 3 I cannot go anywhere else, I can only go to 6 or 0 which are both already visited, so I exit 3 with number 13. I go back and exit 2 with number 14, and I go back and I exit 0 with number 15. So, you should just do the sanity check, there are 8 vertices here, so there should be 8 into 2 is equal to 16 numbers.

Every time I enter a vertex and I exit a vertex, so if I start at 0, I must end with 15 and I indeed do end with 15. If you do not end with 15 then there is a problem. So, if you start with n, vertex with n graph, graph with n vertices and you do a DFS, then at the end the last exit number should be 2n minus 1.

(Refer Slide Time: 3:46)



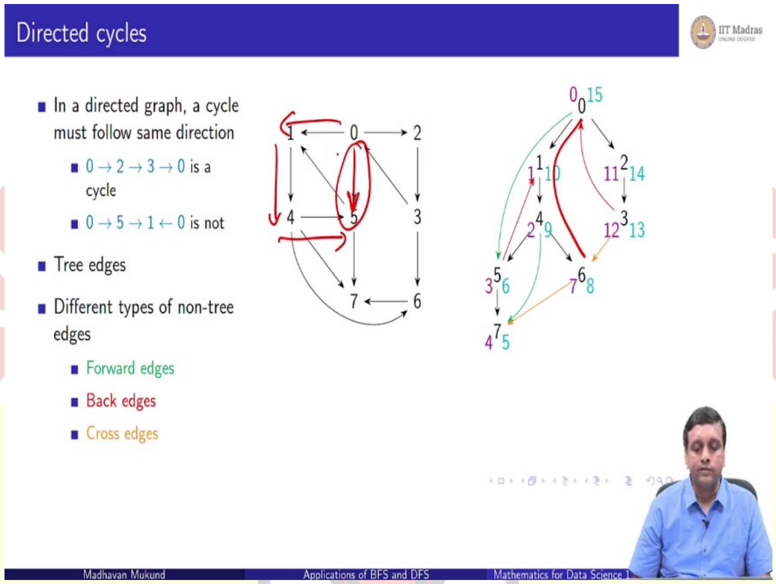
So, here we have tree edges which is what we have drawn when we were drawing these, so all these edges which we followed when we did the DFS are the tree edges. So, this is now a directed tree but otherwise is the same structure as before, so it is tree which connects all the vertices that we visited. But the non-tree edges now come in different flavors.

So, one type of non-tree edge is one which follows the direction of the tree, so it goes from a higher node in the tree to a lower node in the tree. So, the non-tree edge is in the same direction as the path that it is by passing in some sense you are kind of short circuiting, you are going like a flyover on a road, you are going over some intersections and reaching a later point. So, 0 to 3 is a forward edge, so is 4 to 7.

So these are not part of the tree, but they traverse the tree in the same direction as the edges that they are skipping. The converse would be a backward edge, it goes up a path in the tree. So it goes from a lower node in the tree to a higher node but again along a path which already exists. So, I am going from 5 back to 1, and there is a path from 1 to 4 to 5 or I am going from 3 back to 0, and there is a path from 0 to 2 to 3.

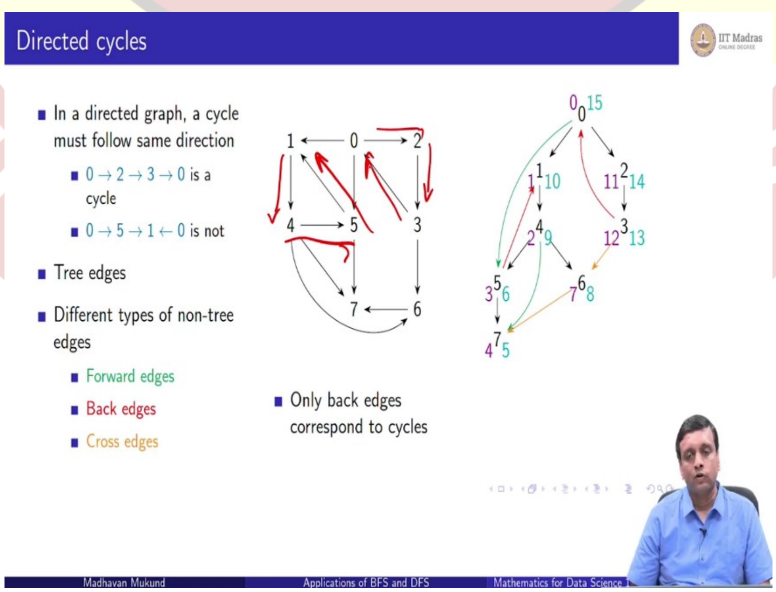
And finally there could be edges which cut across different branches of the tree, for instance, I can go from 6 to 7, so 6 is on this branch and 7 is on this branch, so it is not that I am going up or down from 6 to 7, I am going across. So, these are called cross edges. So, these are the 3 types of edges that you could have in a directed graph which are not in the DFS tree; forward edges, back edges and cross edges.

(Refer Slide Time: 5:28)



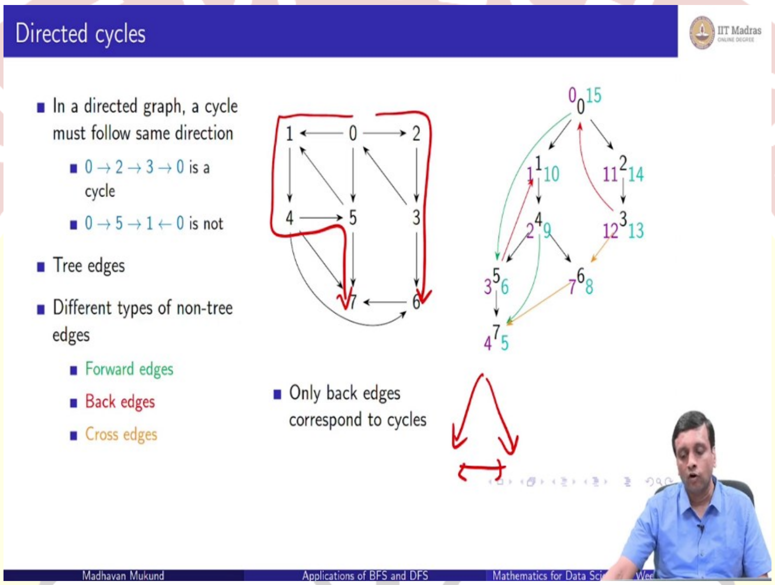
So, now if you look at this carefully, so let us look at this forward edge 0 to 5. So 0 to 5 is a forward edge and the other edges that it was corresponding to are 0 to 1, 1 to 4 and 4 to 5. So, this was the new edge that I added. So, clearly adding this new edge to the existing path did not create a cycle, because it is going in the wrong direction. So, a forward edge does not create a cycle.

(Refer Slide Time: 5:59)



On the other hand, a backward edge does create a cycle. So, if I go from 1 to 4, and then I find that there is a backward edge from 5 to 1, then there is a cycle. Similarly, if I go from 0 to 3 via 2 and then I find that there is a backward edge from 3 to 0, then there is a cycle. So, forward edges do not create cycle, backward edges do create cycles and you can also check that cross edges will not create a cycle.

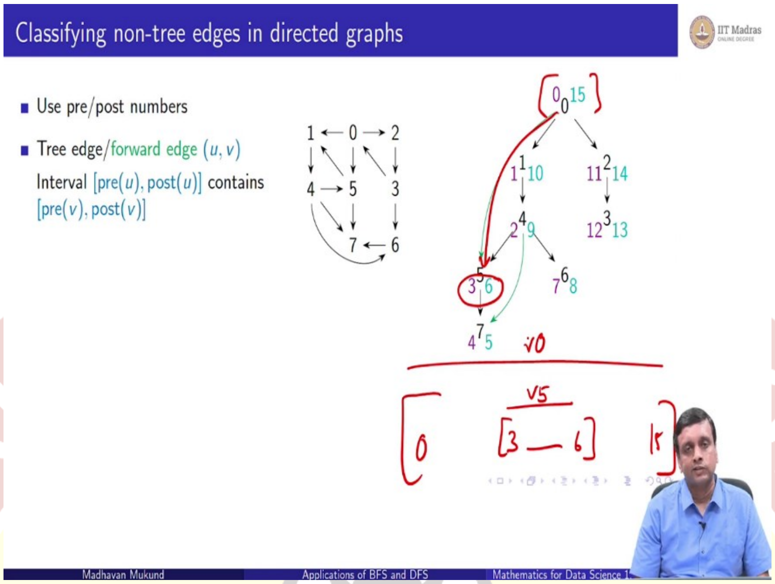
(Refer Slide Time: 6:26)



So, cross edges will actually go down different, different paths and so therefore they do not create a cycle. So, for instance, if I look at this here the cross edge from 6 to 7, so I have a 0 1 4 5 7 path, so I have this path and I have 0 2 6 path. And now these two paths are going in opposite direction, so no matter how I connected this way or that way, so I have a path like this and I have a path like this, so there is no way that I can connect these two paths either left to right or right to left, form a cycle, because the paths themselves are going in the wrong direction.

So, what we want to do now is to identify not just the non-tree edges, so in the undirected case it was very simple, every non-tree edge indicated a cycle. Now in the directed case we are saying it little more subtle than that, it is not enough to just to find a non-tree edge, you must find a non-tree back edge. So, how do we know which of the edges which are not in my DFS tree are forward edges, which are back edges and which are cross edges.

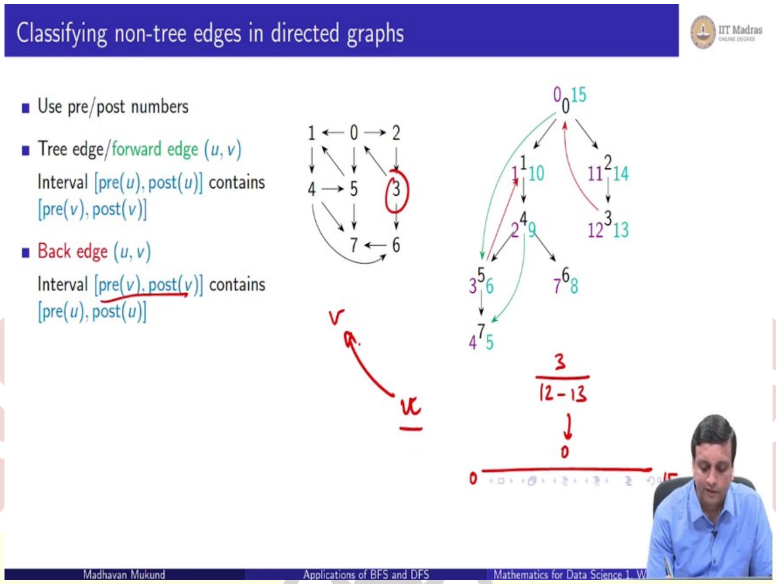
(Refer Slide Time: 7:31)



So, the problem is that of classifying these non-tree edges. And this is the first instance where we will actually use these pre and post numbers. So, if I have a forward edge from u to v, so in this case I have this forward edge here, from 0 to 5, then we will look at the interval, I say that I entered 5 at 3 and I left at 6. So, I have an interval 3 to 6, during this period I was processing 5, I was going to neighbors of 5 and so on. And when I finished processing 5 I was at 6.

And look at 0, 0 has an interval which is from 0 to 15. So, I started processing 0 when the counter was 0 and I finished it when it was 15, so what this says is that the entire processing of vertex 5 happened before I finished vertex 0. So, vertex 5 was processed as a part of vertex 0 processing, so if now the back edge, if the edge goes from the bigger interval to the smaller interval, then it means it is a forward edge, because it is a vertex which was processed earlier to a vertex which was processed later, because the interval is smaller.

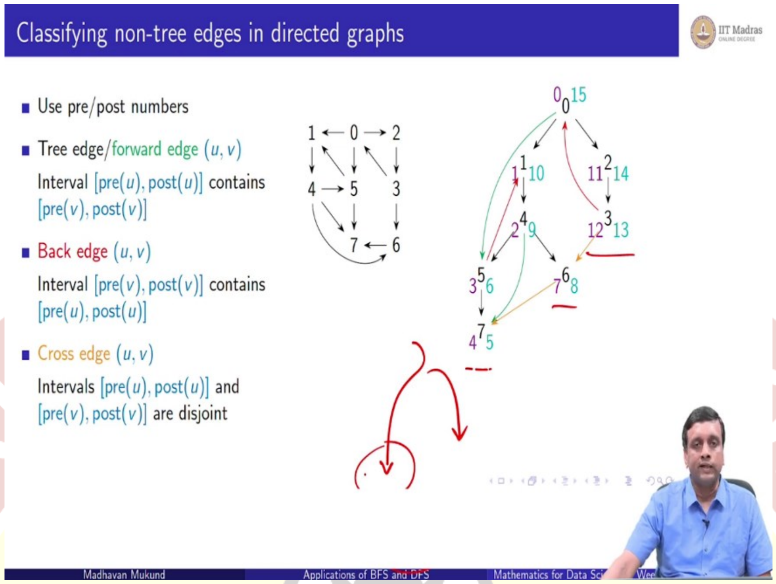
(Refer Slide Time: 8:42)



On the other hand, if I have a back edge, then it is exactly the reverse. So, I am going from say 3 which has an interval 12 to 13, and I am going back to 0 which has its interval 0 to 15. So since I am going backwards, again this indicates that I did the processing of 3 while I was doing the process of 0. So, therefore, this is an edge back from 3 to 0, 3 happened later, so now because the edge is reversed, it is asking that the starting interval is included in the ending interval or it rather the ending interval is included in the starting.

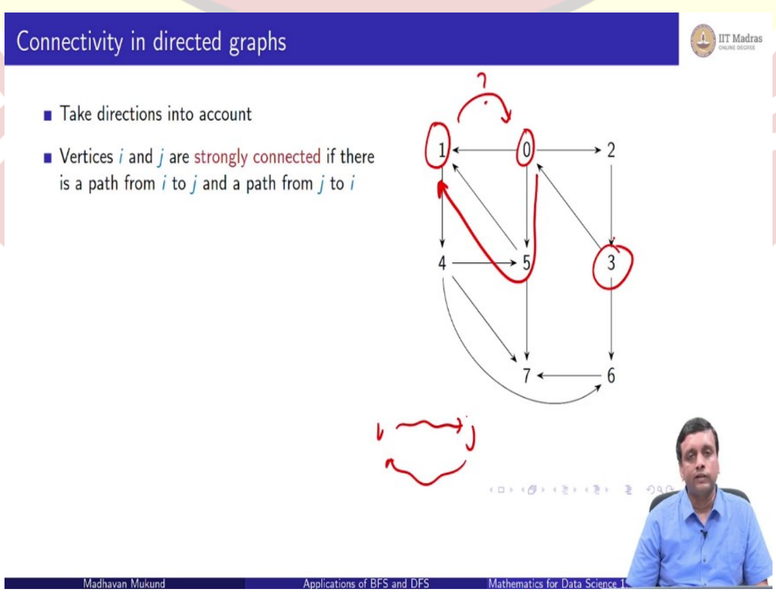
So the, if I am going back from v to u, from u to v, so then I want that this interval is smaller than this interval. So, the ending interval is bigger than the starting interval.

(Refer Slide Time: 9:37)



And finally if I have a cross edge, you will see that these are actually disjoint, because they are happening on different things, so I finished processing one path, so I came down this thing I finished it and then I went back and started here. So, when I started on the right hand side path, I had finished the left hand side path, so all these numbers are exhausted. So, basically the out has happened before the in there, so the two intervals will be disjoint. So, this is how we can use this pre and post numbers with the vertices to discover which of the non-tree edges are back edges and therefore, we can decide whether or not are directed graph actually has a cycle.

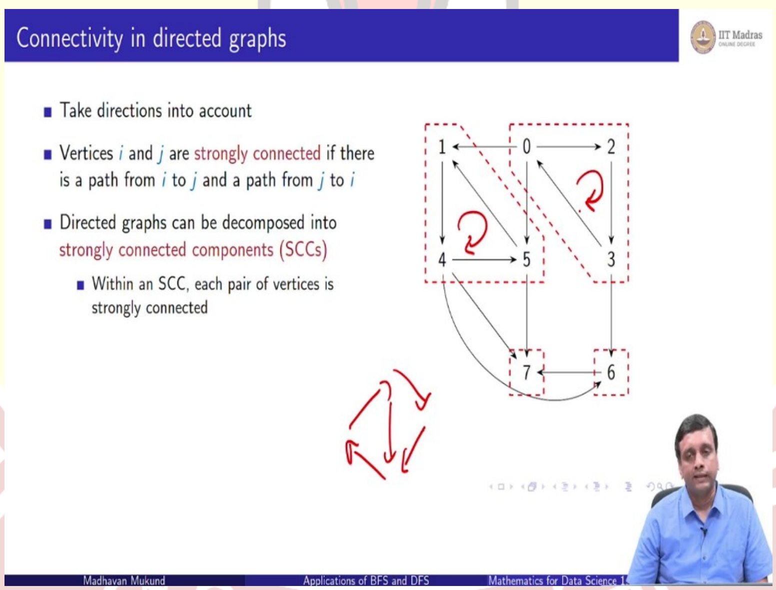
(Refer Slide Time: 10:09)



So, just like cycles have to take directions into account, so does the notion of connectivity. So, we said that in an undirected graph, we said a graph is connected if every vertex can be reached from every other vertex. And now in a directed graph we have to ask whether I can go following the directions. So, I say that a pair of vertices i and j are strongly connected if I go from i I can go to j and then I can come back from j to i by a different path. So, in this case I say i and j are strongly connected, if they were not strongly connected, it is possible I can go from i to j but I cannot come back.

So for instance, if I look at, say for instance 0 and 1 in this case, I can go from 0 to 1 by following this path, but there is no way to go from 1 to 0. Because I cannot go from 1 except 4 and I cannot, basically the only way to come back to 0 in this graph is to come by a 3, because that is the only incoming edge to 0 and there I cannot reach from 1 to 3 by any path. So, 0 and 1 are connected in one direction but not connected backwards and therefore they are not strongly connected.

(Refer Slide Time: 11:25)

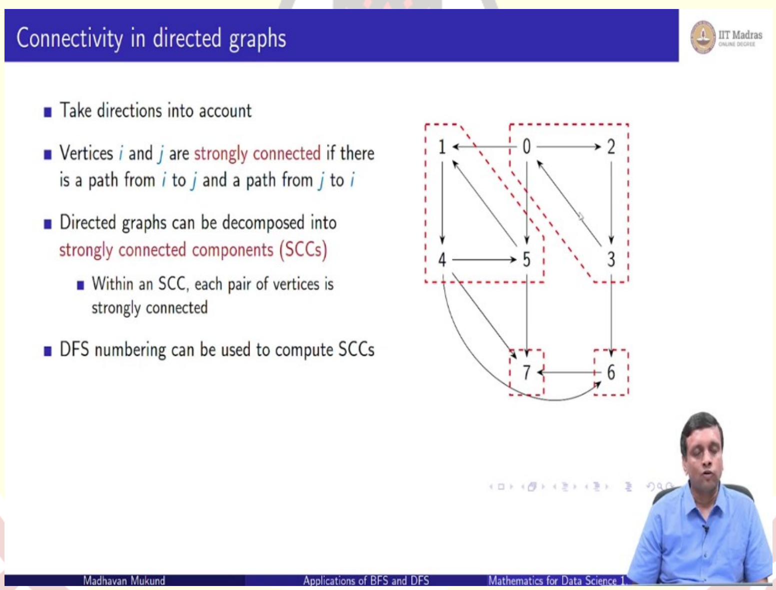


So, therefore, a correct notion of component that we need for a directed graph is one where not just that every pair of vertices is connected but every pair of vertices in that component is strongly connected. I can go from everywhere to everywhere and come back, so I can go anywhere in that component without worrying about where I am starting. So, this is what is called an SCC or a Strongly Connected Component.

So, you can see that for 3 vertices, like we have in this a 3 vertex strongly connected component is just a cycle. So, basically if I have a, if I have a directed cycle, if I have something like this, then this will be a strongly connected component. But I could have more edges, I could have something like this and so on. It does not matter if there are more edges, but there should be a minimum number of edges so that I can go from anywhere to anywhere in both directions.

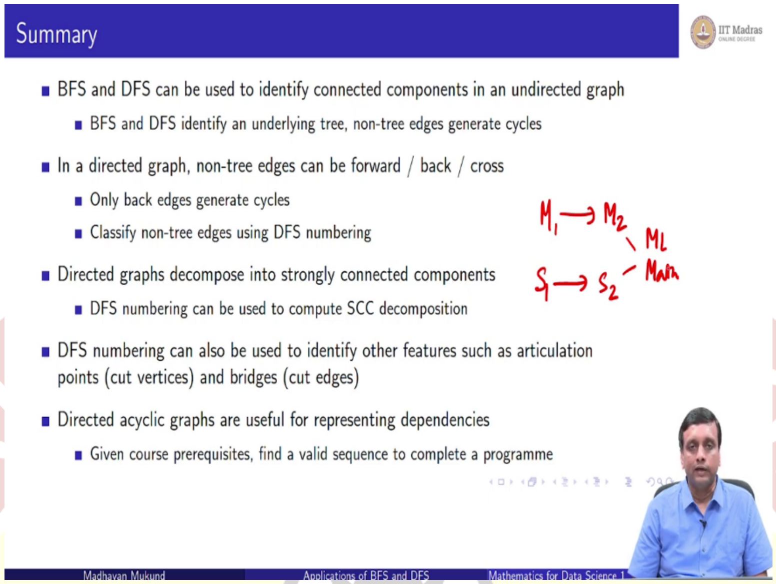
So, in this particular graph 1 4 5 forms a cycle because I can go around this in this direction and reach anywhere from anywhere. Similarly, 0 2 3 forms a cycle, but 7 and 6 now are stuck on their own because if I leave 6 I cannot come back to 6 from any of these paths, if I leave 7 I cannot come back to 7, I cannot leave 7 at all in fact because 7 has no outgoing edges. So therefore the strongly connected components in this graph are the ones which I marked in red.

(Refer Slide Time: 12:40)



So, what we are not going to cover in this particular course but maybe at a later stage is that this DFS numbering that we just did can also be used to compute these strongly connected components. So, we saw that it can be used to compute back edges and detect cycles, but it can also be used to detect strongly connected components.

(Refer Slide Time: 12:58)



So to summarize, we saw that BFS and DFS are primary strategies for reachability in a graph, but what we have seen now is that we can do much more with BFS and DFS. So, the first thing we can do is identify the connected components in an undirected graph. So, by doing BFS or DFS we first uncover a tree, so we identify some edges in the graph which we process during the BFS and DFS, this form a tree. And any edge which is outside this tree must form a cycle with the edges in the tree, so any time there are non-tree edges in the graph after I finish my BFS or DFS, we can generate, we can say that there is definitely a cycle in the graph.

However, for directed graph we saw that this is little bit more complicated, so we have 3 types of non-tree edges; forward edges, back edges and cross edges. And of these only the back edges generate cycles, so this is one instance where we used this DFS numbering this pre and post numbering for vertices in order to detect which of these non-tree edges are back edges.

And finally, we described the notion of a strongly connected component, and although we did not actually calculate them, we claim that the DFS numbering that we have done can also be used to identify the strongly connected components. So, a strongly connected component, remember has one where every pair of vertices is reachable from each other. So, we mentioned last time and we will not elaborate on this, but DFS numbering can be used for many other things, so one thing it can do is identify the so called critical vertices.

So, cut vertex or an articulation point is one which if I destroy it, it will disconnect the graphs. So, if I remove it from the graph, the remaining graph falls apart, so this is critical for instance if you are looking at a network of say a power network, if there is a power station which is relaying power and if it goes out of action and the power network now disconnects, so two parts do not get power from each other, then that is something that we have to be extra careful to protect.

So, these are also important things to discover in your graph. Similarly, there could be cut edges, if I cut a connection between two nodes, then the graph falls apart, so these are called bridges. So, articulation points and bridges can also be identified in a graph using this notion of DFS numbering. And finally, this idea that we are looking for cycles in a graph is particularly important in the directed sense, so there is a very important class of graphs which we will look at next week, which is called directed acyclic graphs.

So, a directed acyclic graph as it suggests is a directed graph which has no cycles in it. Now a directed acyclic graph is very often used to represent these kind of prerequisites or preconditions or dependencies. So, supposing I want to describe for instance the course contents of this program and I say that you have to do maths 1 before you do maths 2, and if I have to do stats 1 before you do stats 2 and maybe there is no correlation between maths 1 and stats 1, so you can do them in any order and you can postpone one to the other, but you cannot do maths 2 before maths 1.

So those are, so I have M_1 and I have an edge saying M_1 must be before M_2 and I have one saying that S_1 must be before S_2. And now maybe I have something which says that in the third semester there is an math for ML in which I must have completed both M_2 and S_2. And now I ask in what order you can do these things? So, clearly you can do math for ML only after you finish all these 4 courses, but these 4 courses you can be a little flexible, you can do S_1 for instance, after M2 or you can do M_1 after S_1.

So, these kind of properties about which sequences can be compatible with a set of dependencies are used very often and this can be done by analyzing this directed acyclic graphs which we will do later on.

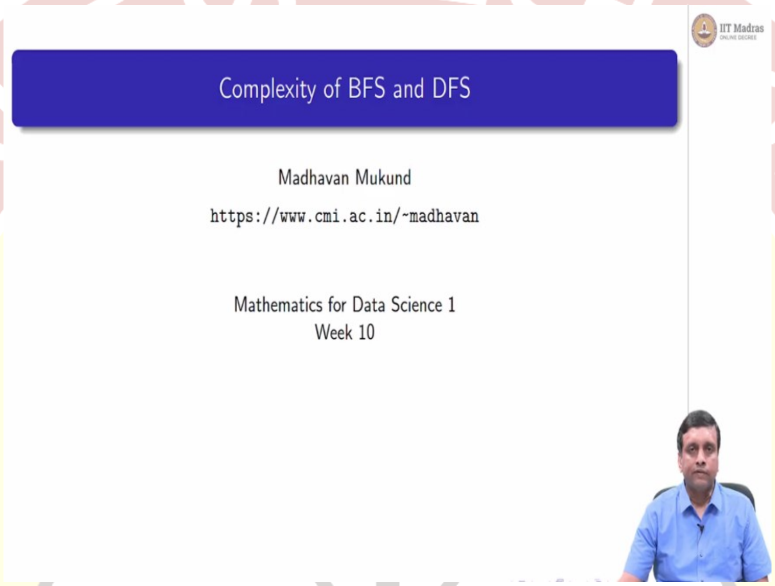
INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

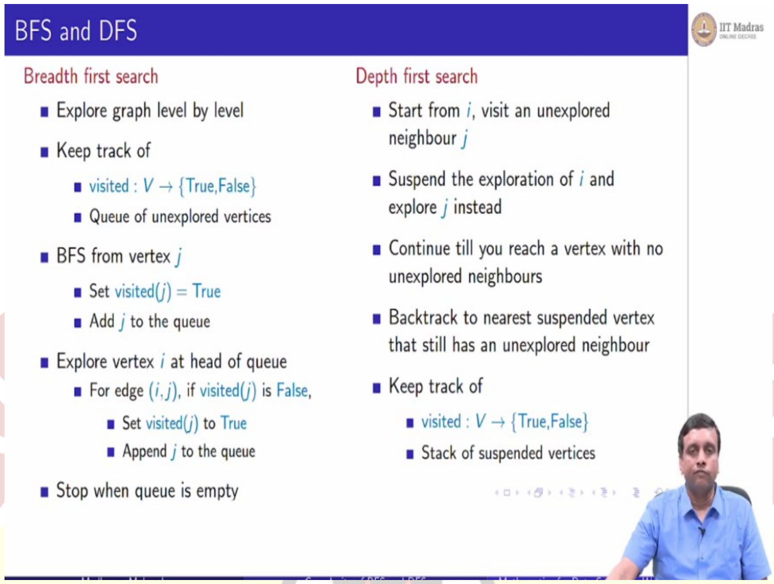
Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Indian Institute of Technology, Madras
Lecture No. 66
Complexity of BFS and DFS

(Refer Slide Time: 00:14)



So, having seen some of the applications that we can achieve using BFS and DFS. Let us go back a little bit and look at the connection between BFS and DFS and these representations that we talked about Adjacency Matrices and Adjacency Lists.

(Refer Slide Time: 00:28)



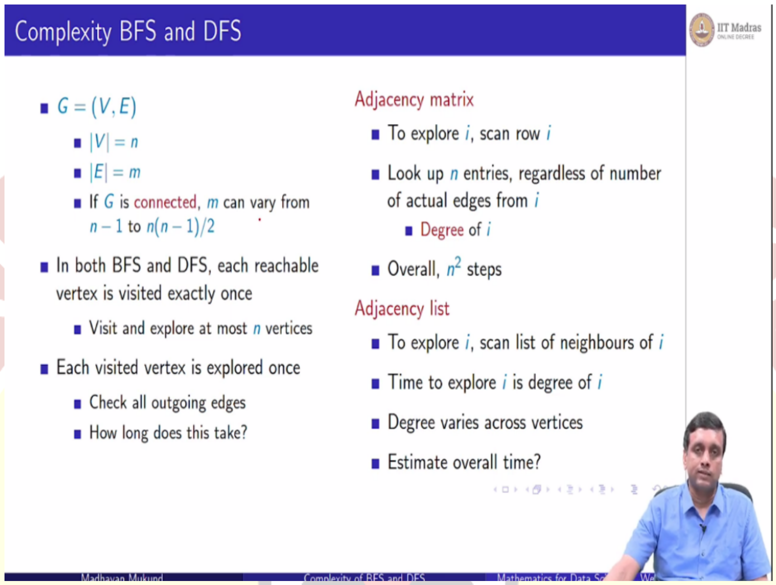
So, let us just formally remember what BFS does. So BFS explores a graph level by level. So, we maintain 2 pieces of information we maintain this flag called visited, which indicates whether a vertex have been visited or not. And we keep this queue of unexplored vertices. So initially, we mark all vertices as unvisited, we start from a vertex j. So, we start by setting that to be visited, set its value of visited to true, add j to the queue, and then we repeatedly process the queue.

And by processing the queue, what we mean is we take out the first element in the queue, look at its neighbors, and if there is any neighbor, which is unvisited, we mark it as visited, and push that neighbor back into the queue, so it will get processed later on. And finally, when the queue gets empty, BFS terminates.

On the other hand, with Breadth first search, we do a kind of aggressive or, impatient traversal. So, we start with i, and we visit 1 neighbor j. And now once we visited 1 neighbor, j, instead of going back to i and continuing with another neighbor of i we, we suspend i, and we go to neighbors of j. The first time we find a neighbor of j again, we will suspend j, and go to a neighbor, that neighbor k and continue.

And at this point, when we finish, and we cannot go further, we go back. And then we have traversed back all the suspended vertices until we find 1 which is not finished, and look for another neighbor, and so on. So here, we keep track of this visited information like in BFS, but we also have the stack, which remembers the suspended vertices, so we can go back in the correct sequence Last In First Out.

(Refer Slide Time: 02:05)



So, what I want to talk about is how much time this takes as a function of the size of the graph that we are trying to explore. So typically, in a graph, there are 2 parts, there is the vertex set, and there is the edge set, which is a subset of the pairs of vertices. So, edge relation is a subset of V cross V. Now, the vertex set is usually denoted as having size n. So, this tells you how many nodes or vertices there are in the graph.

Now the same set of vertices, you can draw many edges, or we can draw a few edges. So, actually, the parameter of the number of edges is independent, in a sense, as a measure of how complicated the graph is. So, this is usually denoted by a small m. So, m is usually the number of edges, n is the number of vertices. So, we saw that, if you have a tree, that is a minimally connected graph, then you will have n minus 1 edges. So, we can have interesting graphs in which the number of edges is roughly the same as the number of vertices, you have n vertices, you have n minus 1 edges.

On the other hand, if you connect everything to everything, then for every pair of vertices, you have an edge. And this gives us n into n minus 1 by 2, which is n choose 2 vertices. So, this is about n square so some something like n square, so we can have either order n vertices or order n square vertices. So therefore, the number of edges, so the number of edges forms and somewhat independent parameter in our calculation.

So, now let us look at both BFS and DFS and see how they do this. So, the first thing is that they visit and explore every vertex exactly once, that is the purpose of that flag visited, it makes sure that we never go back to a vertex which is already visited and try to visit it a second time. So therefore, we will visit and explore each vertex exactly once. So, that whole thing happens in n times. There are n times when we visit and explore vertices. Now, what is exploring a vertex mean?

Exploring a vertex essentially means looking at all its neighbors. So, if we are looking at all the neighbors of a vertex, we are looking at all the edges which are outgoing from that vertex. So, the question is, how long does it take us to do this, when we are actually doing it computationally not doing it by looking at the picture and doing it by hand. So, we said that there are 2 representations that we have of the graph.

The first is this adjacency matrix in the adjacency matrix, the entries are 01. And the ij th entry indicates whether there is an edge from vertex i to vertex j. So, if there is an edge, it is 1 if there is no edge is 0. So, if we want to look at the outgoing edges of a vertex i, the only way we can do it in an adjacency matrix is to walk down the entire row for i. So, these are also look at a i,1, i,2, i,3 and so on up to a,i,0 i,1 up to i,$(n-1)$ if the vertices are numbered 0 to ,$(n-1)$.

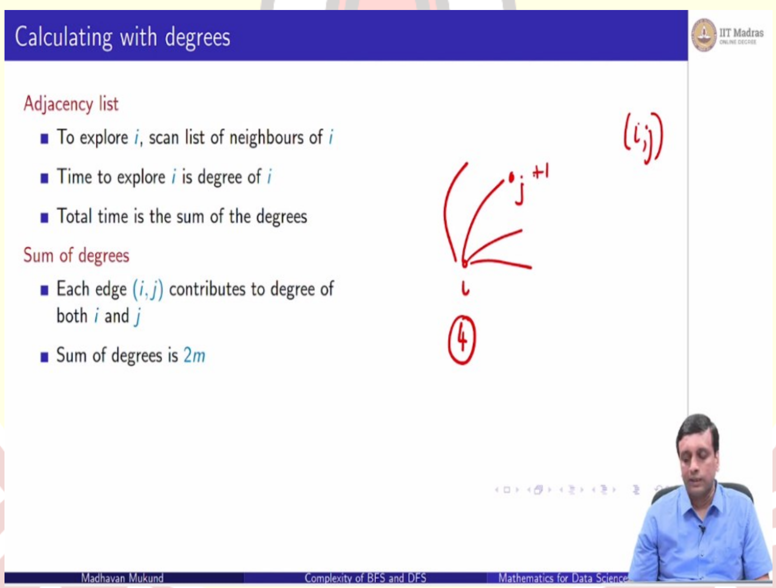
So, we will have to look up the entire row so, whether or not i has many neighbors, or few neighbors are no neighbors at all. I mean, of course, in an undirected graph, it must have at least if we had reached i during this thing, it would have an incoming thing. So, at least one neighbor, but if it is a disconnected vertex from where we are starting, like we saw a connect a component, which has only a single vertex, then it may have nothing at all.But we would not know that until we see the entire row for i.

So, regardless of how many neighbors i actually has, we have to spend time proportional to n, to discover all these neighbors. So, this means that I have overall and processing n vertices. And for each vertex, I have to scan n entries in this matrix. So, $n \times n$, so I have to do something proportional to n square, in order to do Depth first search or Breadth first search.

On the other hand, if I use an Adjacency list, then I have for each vertex an explicit list of its neighbors. So, if it has a lot of neighbors, at least to be long, if it has very few neighbors, and this will be short, but I will spend no more and no less time than I need to scan this list. So, if I have k neighbors, I have to look at k entries. But I do not need to spend more, I do not need to spend n steps looking for k entries when k is small.

The now, the problem with this is that the degree of the vertex as we call it, right degree is the number of vertices which are number of edges, which are incident at a vertex, the degree of a vertex varies. So, maybe for vertex 1, I had a small degree vertex 2, I had a big degree and so on. So, if I want to count how many steps it takes across these n vertices, I have to add up the degrees, I have to look at how much time it takes degree 1, how much time it took for degree 2, and so on. So, how do I get a good way of estimating what this adds up to?

(Refer Slide Time: 06:48)

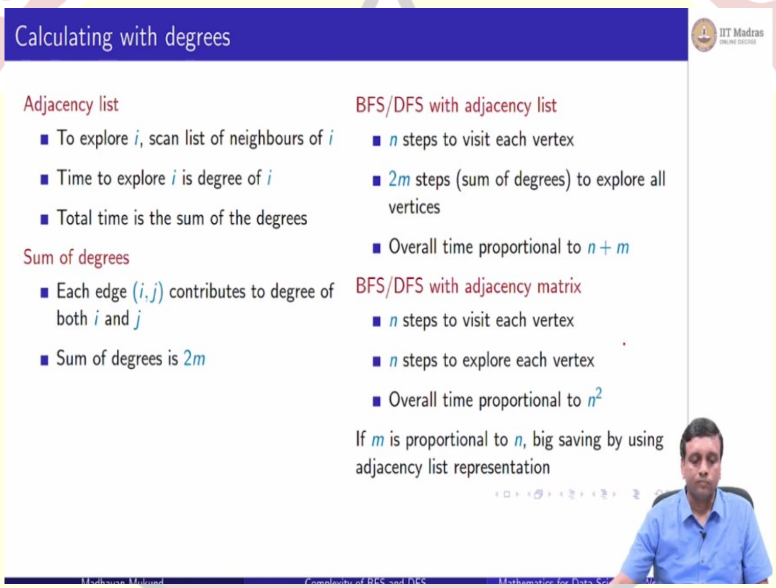


So, the question is, when I am processing an Adjacency list, I do work proportional to degree of each vertex added up. When I process vertex 0, I will look at all its neighbors. So, I will spend time proportional degree of 0 and I do vertex 1, the same thing with vertex i the same thing. So, we are really interested in identifying what the some of the degrees in an undirected graph represents. So, what is a degree?

A degree indicates a number, which is the number of things going out. So, if I have i, then if I have so many things going out, then I get a contribution of 4 from i to the sum of the degrees, because it has 4 outgoing things. But each of these things will go and terminate also in j. So, if I look at this edge i,j, it adds 1 to the degree of i, it also adds 1 to the degree of j. So, each edge contributes to degree of both i and j.

And every number that I get in the degree must come from some edge. So, since there are m edges, and each edge contributes to the degree count of both the starting point and the ending point, the sum of the degrees must actually be $2 \times m$. For each, each vertex each edge, 1 to m, it contributes 1 plus 1. So, it is 2 times.

(Refer Slide Time: 08:16)



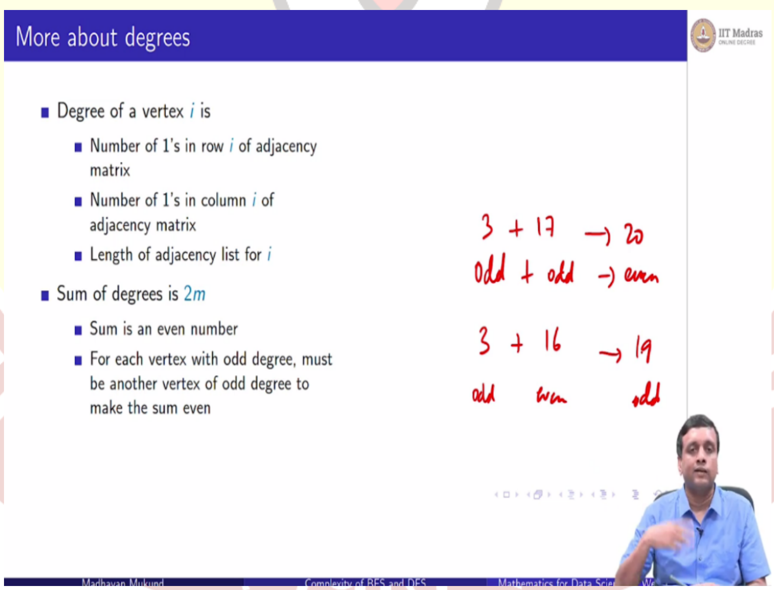
So, now we can see that if you have BFS or DFS with an adjacency list, then you make n steps, you take n steps to visit each vertex. And then across the n vertices, you take the, the sum of the degrees or 2 m steps to explore all of them. So, this says that the time that overall that you spend is proportional to n plus m, you have to visit all the vertices because for instance, if the entire graph is disconnected, there are no edges at all, m is 0, but it does not mean that you will finish your DFS or BFS in 0 time, you have to visit all the n things.

So, you have to spend n steps looking at all the vertices and across the vertices, the contribution of that vertex to your work is the degree of that vertex. So, across the vertices is the summation of the degrees, that is where this m term comes. Whereas as we saw, if you did this with an adjacency matrix, it does not matter really what the degrees are, and what m is, you will end up having to spend n steps for every vertex, because you cannot find out the neighbors of a vertex without looking at the entire row. So, you end up taking time n square.

So, the whole distinction is between n square and n plus m. So, we said that m could be small. So, m could be like n, like in a tree, we have only n minus 1 vertices, or m could be large, it could, in principle be n square. If it is n square, then these 2 are roughly the same. But if it is small, then we have a difference between something which is proportional to n, and something which is proportional to n square.

And that is why adjacency lists can be beneficial for doing our thing. So, if m is proportional to n, then we have a big saving by using adjacency list representations. So, though adjacency matrix is fine in terms of understanding what is going on in terms of mechanically computing something, it can actually cost us not just because it is a large thing, and we are keeping a lot of 0s in it. But also, because in order to extract this interesting information about what are the neighbors, we have no better way than to walk down the entire row.

(Refer Slide Time: 10:19)



So, here are some more interesting things that is useful to remember about degrees. So, remember that the degree of a vertex can be obtained from an adjacency matrix by just looking at its row. So, if you just count the number of 1s in the adjacency matrix in a row, you get the degree of the vertex. And if it is an undirected graph, you can also get it from the columns, because it is the number of incoming edges and the number of outgoing is the same, because it is a symmetric kind of edge graph.

And if you have an adjacency list, then the degree is just the length of the list associated with i, it is all the neighbors of i. Now, we already calculated that the sum of the degrees is 2 n. So, 2 times anything must be an even number. So, the sum is an even number, because it is 2 times the number of edges. So, the number of edges themselves will be odd, but the sum of the degrees is going to be 2 times m, and therefore the sum of the degrees is an even number.

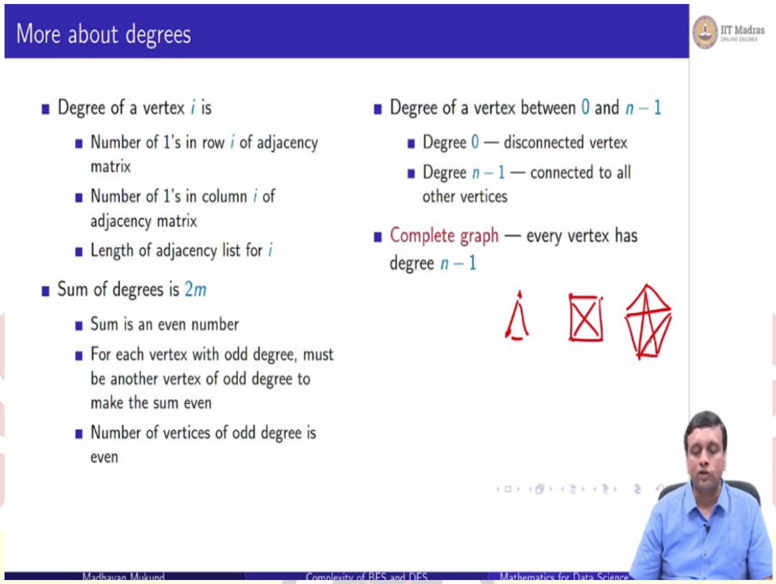
So now, you should remember that, if you have an odd number plus an odd number, then I will get an even number. So, for example, if I do 3 plus 17, then I get 20. But if I do an odd number plus an even number, then I get an odd number. So, you can think of the fact that if I am pairing up an odd number in 2's, then I get 1 leftover element, that is why it is odd. And if I am pairing up an even number in 2s, I get no leftover elements.

So if I take them together and pair it off in 2s, then I have 1 leftover element, that is why it is odd. Whereas if I have 2 odd numbers, then each of them contributes 1 leftover element, I can take those leftover elements and pair them up and I get a new pair, and so it is even, so, odd plus odd is even or plus even is odd. So, this means that if I spot an odd degree vertex in my graph, it cannot be the only 1 there must be another 1 right because otherwise the sum of the degrees will become odd everything else is even plus 1 odd number will be an odd number.

So, if there is an odd number here, there must be an odd number somewhere else, so I can pair them off. Similarly, if I find a third odd vertex there must be a fourth odd vertex. So, for every odd degree vertex, there must be another odd degree vertex. So in other words, the number of degrees or the odd vertex or degree vertices must themselves be even, if I have an odd number of vertices with odd degree, then the overall sum of the degrees will be odd which is not possible.

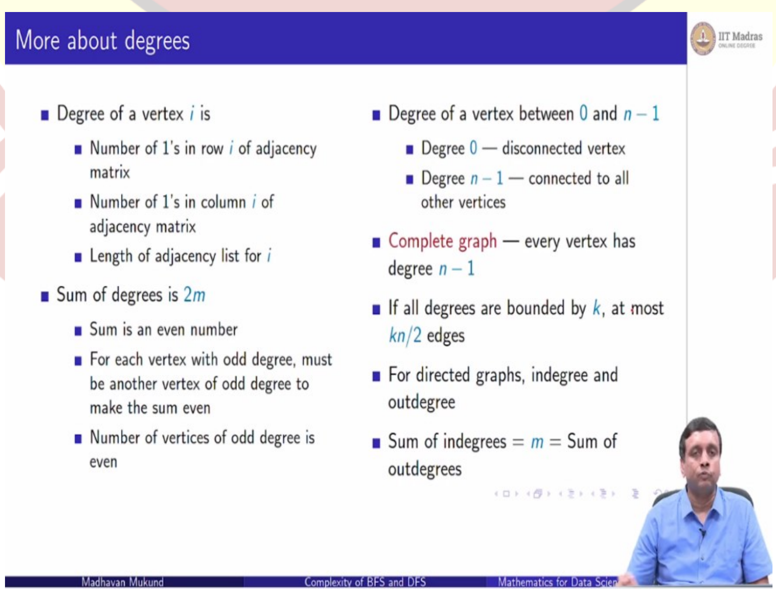
So remember also that the degree of a vertex can be any number between 0 and n minus 1. So, 0 happens when this vertex is actually disconnected from the entire graph, n minus 1 happens when it is connected to every 1 of the remaining n minus 1 vertices. So, it is not connected to itself obviously, we said no self loops, but it could be connected to everything else. So, the special case where every vertex is connected to every other vertex is what is called a Complete graph.

(Refer Slide Time: 13:15)



So, for instance, on a 3 vertices graph, a complete graph is a triangle, if a 4 vertex graph, then I must connect everything to everything. So, I have a square and then I also have the diagonals. Similarly, if I have a 5 vertex graph, then I will have a pentagon with all these diagonal things. So, this is what is called a complete graph. So, in a complete graph, the degree of every vertex is n minus 1 and I will actually have n into n minus 1 by 2 because every pair is connected. So, I have n choose 2 edges.

(Refer Slide Time: 13:44)



Now many graphs that we encounter in practical problems like the 1s we discussed about graph coloring or vertex cover, so on. In many reasonable situations, you can actually say that a particular node will have no more than a certain number of neighbors for example, remember the graph coloring problem for the security cameras? So, clearly, if I give you a particular building, then I know that a given intersection will not have more than a certain number of corridors fixing I cannot have a large number of corridors at some point.

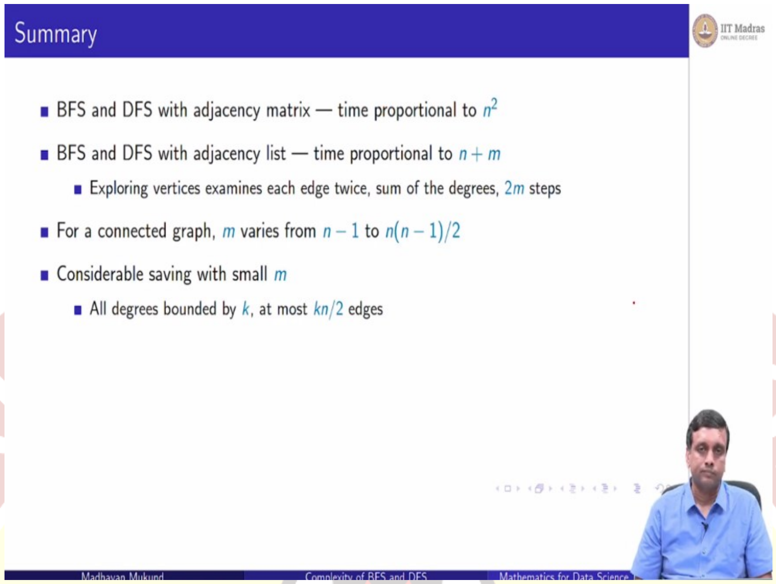
So, there will be some upper bound saying that no corridor, no intersection has more than 5 corridors which meet there or if we are looking at some, some other problem, which is say for instance, placing ambulances at an intersection, then you want to know how many roads meet at that intersection. Or if you are looking at say, the timetabling problem you want to know how many different courses can be scheduled in the same slot. Now, obviously, if you have a fixed number of courses in your curriculum, not more than that many can be there.

So, very often the degree is actually independently bounded by some number, you do not have arbitrarily large degree. So, the number the degree bounded by an external constraint. And if you have now a constraint on the degree, it says that each degree is k, then the total sum of the degrees can be at most k times the number of vertices k times n. And since this is twice the number of edges, the number of edges must be $k \times n/2$.

So in other words, if you have a bounded degree graph, where every edge has a bounded degree, which is independent of the size of the graph, then the total number of edges cannot be more than linear, cannot be more than some function of n. So therefore, you should be working with adjacency lists. And finally, if you have directed graphs, we said that it is no longer enough to talk about degree because there are edges coming in, and there are edges going out, which are quite independent of each other. So, we must talk about the in degree and the out degree.

So, in this case, because each edge contributes to 1 in degree and out degree, the sum of the in degrees is the number of edges and the sum of the out degrees is also the number of edges. So, together the in degrees plus out degrees is 2m, but they get partitioned into 2 quantities which add up to m each.

(Refer Slide Time: 16:06)



So, what we have seen is that if we do an analysis of how the BFS and DFS work with respect to exploring the neighbors of a vertex, if we use an adjacency matrix, it turns out that regardless of how many neighbors a vertex has, we must scan the entire row and therefore the time taken by BFS and DFS becomes proportional to n times n, I have to process n vertices and for each vertex, I have to scan n elements in that row.

On the other hand, if we use an adjacency list, then the time taken to process a vertex is exactly the number of neighbors it has. So, it is across the vertices is the sum of the degrees and this gives us an overall timing, which is proportional to n plus m. And we also saw that there is a large variation in m. So, m can be linear, like in a tree, or it could be quadratic, like in a complete graph. So, it is important to be able to distinguish and use the appropriate representation in particular use adjacency list whenever we can.

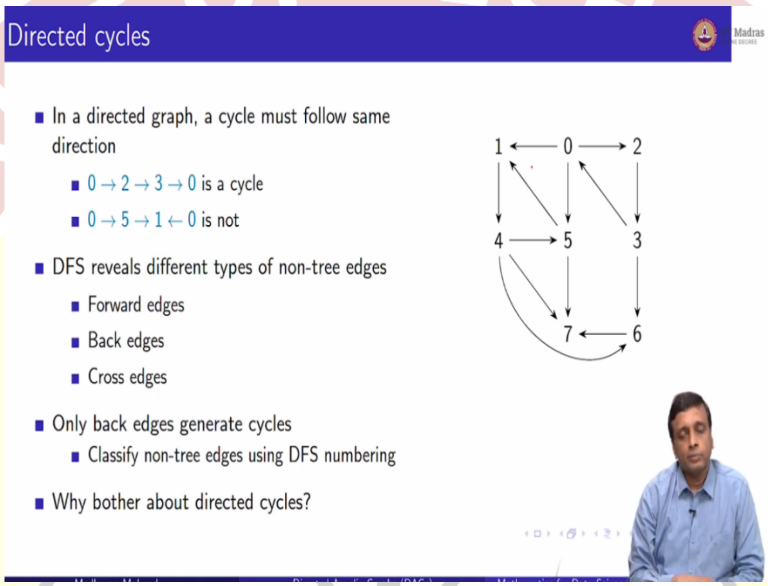
Another situation where we get small number of edges is when we have a bounded degree. So if we, if we have some other constraints on our problem, it says that the number of edges coming out of a given node cannot be more than a certain number independent of the total number of edges. Then we have necessarily a graph which has only a linear number of edges. So again, an adjacency list representation would work best.

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Indian Institute of Technology, Madras
Lecture No. 67
Directed Acyclic Graphs

(Refer Slide Time: 00:14)



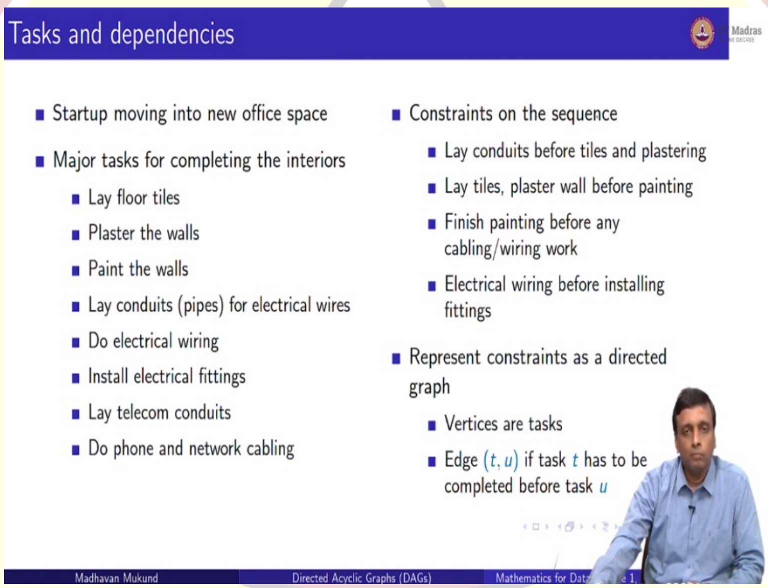
So, last week we looked at graphs with cycles. So, we saw that we can use our depth first search and breadth first search to find cycles and graphs. And in particular, we looked at directed cycles. So, we said that if in a directed graph, you have a cycle, then the cycle must follow a uniform direction. So for example, in this graph here, we see that this is a cycle because you can go from 0 to 2, 2 to 3 and back to 0.

But although this other one here on the left looks like a cycle, it is not because if you go from 0 to 5, and then 5 to 0, then this edge from 0 to 1 is in the opposite direction, so we cannot follow that direction. So, that is not a directed cycle. Whereas for instance, this is a direct cycle. And so this is what we looked at last time. And what we said is that we can use DFS, Depth First Search to find these directed cycles. Because when we do DFS, we will construct a tree to begin with.

So, we have all the edges which are passed are part of the tree which have been used during DFS. And then among the non tree edges, we have 3 different types, so we have these forward edges, which go forward in the tree, we have back edges, which go up the tree from a later node to an ancestor. And then we have cross edges which go across branches. And we said that only the back edges actually generate cycles.

And by using this DFS numbering by recording the time at which we enter each vertex to process it, and we exit after finishing processing the vertex by looking at these DFS pre and post numbers, we said that we could analyze the tree and look at all the non tree edges and decide which category they belong to. So in this way, we can find out all the non tree edges. And the back edges in particular, which are the cycle forming edges. So, now the question is, why were we so worried about cycles in directed graphs to begin with?

(Refer Slide Time: 02:00)



So, let us look at a general problem, where we have some things to do some tasks, and there are some dependencies between the tasks. So, as an example, suppose there is a startup, which is trying to move into some new office space. So, there is some brand new office space, and the startup needs to set up this office before it can move in. So, this office space is completely unfinished. It is a new building just constructed, just the bricks are there.

And so what we need to do is a number of things, we need to lay the floor tiles, we need to plaster and paint the walls, we also need to lay pipes, these conduits as they are called. In order to take wires from here to there. So, there are wires of two types there are electrical wires, but also there are networking cables for computers, there are also telephone cables and so on, and these cannot go in the same conduit because they interfere.

So, we will have separate conduits for electrical wires, separate conduits for telecom equipment, then of course, you have to put in the wiring. And you have to also after you finish the wiring of the electrical things, you have to put in the fittings, you have to put in the lights, the fans, the switches, and so on. So, now, these are all activities which need to be done, but clearly they cannot be done in arbitrary order. So for instance, we have these constraints.

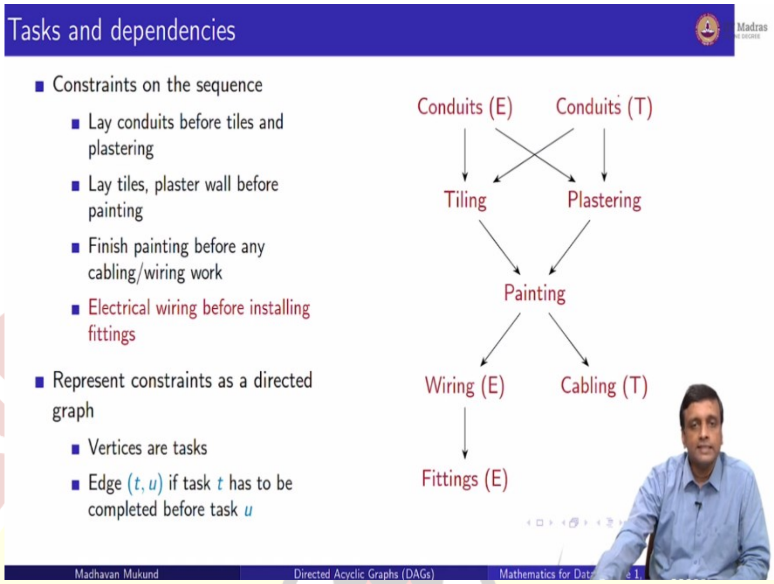
So, we need to put the conduits into the wall and the floor. So typically, these conduits run along the wall to the plug points and to the ceiling. And then across the room, they will flow travel underneath the floor. So, before you put the tiles and before you plaster the walls, you need to put the conduits otherwise, obviously you have to break the tiles or break the walls, which is not a good idea.

Then before you paint the walls, you must of course plaster through the wall. But typically you also like to paint before you lay the tiles because you expect that the person who is laying the tiles might mess up the walls by putting cement when they are laying the tiles. Whereas of course, tiles are usually washable. So, if you are painting it and some paint falls on the tiles, It is not a problem. And clearly you would like to finish the painting before you start putting in the wires into these conduits.

The reason is that if you have the wires hanging out loose when they are painting then the paint will go and gum up the wires and then you will have also paint going into these cracks. So, normally you seal up these conduits with something and then you paint it and then you open it up and put push the wires through these conduits. So wiring cabling happens after the paint and clearly you cannot put your fittings in or you cannot put plug points and you cannot put lights and fans unless the wires are there. So, you can finish the electrical wiring only after installing the fittings.

So, we are going to model this as a directed graph. So, the vertices are going to be the tasks that we have to perform. So, all these tasks on the left laying the tiles plastering the walls and so on. And an edge is going to denote a dependency. So, an edge from t to u says that t has to be finished before you can be started. So, in our case, for instance, you have to lay the tiles before you paint the walls.

(Refer Slide Time: 04:53)

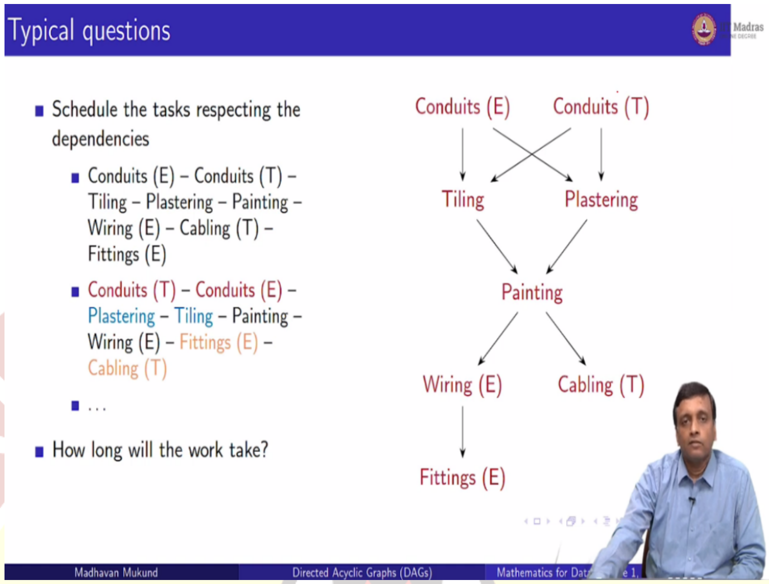


So, if we take this particular thing and draw the graph, so first we have the vertices so we have these vertices corresponding to the activities we have two different types of conduits, electrical and, and Telecom. And we have tiling and plastering. We have painting. We have two kinds of wiring, we have electrical wiring and telecom cabling. And finally we have the electrical fittings. So, these are all the activities. These are the nodes in my graph. And now I have these constraints.

So, the first constraint says that I must lay all the conduits before I do tiling and plastering. So, from each of the conduit nodes, vertices, I have an edge to each of the other two vertices one to tiling and one to plastic. Then it says you must finish plastering and prep and doing the tiling before you paint the walls. So, from both the tile and the plaster edges, we have nodes we have an edge to painting.

Similarly, from painting, we have an edge wiring and cabling because these happen only after painting and finally we can do the fittings only after we do the wiring. So, this is a directed graph, which we have constructed from the task constraints given to us.

(Refer Slide Time: 05:56)



So, what is our problem now our problem is we need to complete these tasks in a way that respects the dependency, so we must make sure that the painter comes only after the tiling and the plastering are done. So, here is a possible way of doing it. So, we start with things which do not have any dependencies, the conduits can be laid right at the beginning. Once they are both done, we can do tiling and then plastering.

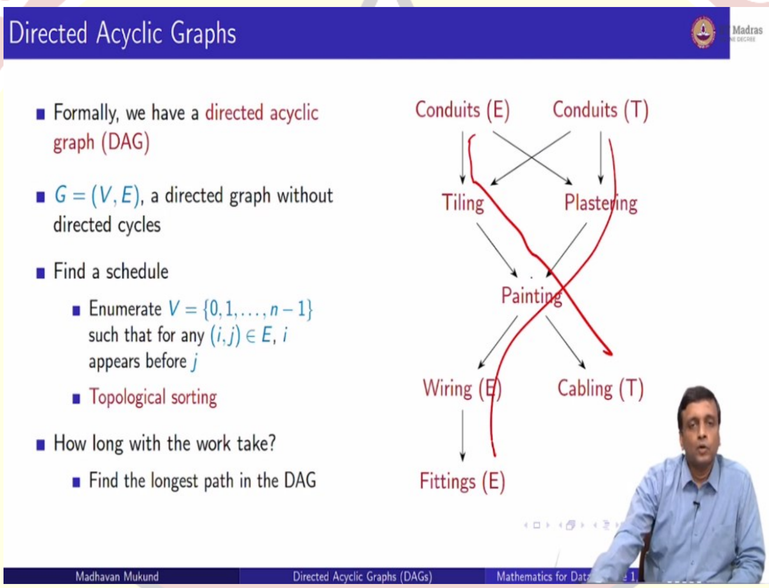
Then once both of these are done, we can do painting and then we can do the wiring first the electrical wiring then the telecom cabling, and finally we can put the electrical fittings. So, this is a sequence in which we can complete these tasks so that whenever we come up to take up a task, all the tasks which needed to be done before that are already done. But this is not the only such sequence of course.

For instance, we could have done the waiting in the opposite order. It does not matter whether we do the telecom cabling (bef) conduiting before or after the electrical conduiting. Similarly, we could take up the tiling after the plastering, because plastering and tiling do not depend on each other. And similarly, we can do the electrical fittings even before we do the telecom cabling, because they go through different conduits and they do not interfere with each other.

Another question we might ask so, this first question is how do we sequence these in a way that does not violate these constraints? The second question is, what is the best way to do this? Supposing we could do things which are independent of each other at the same time. For instance, we could ask the person who's putting the tiles to work alongside the guy who is plastering because we said that tiling and plastering can happen together.

Similarly, we can have the person doing the electrical conduiting, working alongside the person who is doing the telecom conduiting. Similarly, we can have the wiring and the cabling done at the same time. So, if we can do all this, if we can optimize this so that things which are not dependent are done in parallel, then how soon can we finish this? How many days will it take to complete all these tasks, following these dependencies?

(Refer Slide Time: 07:47)



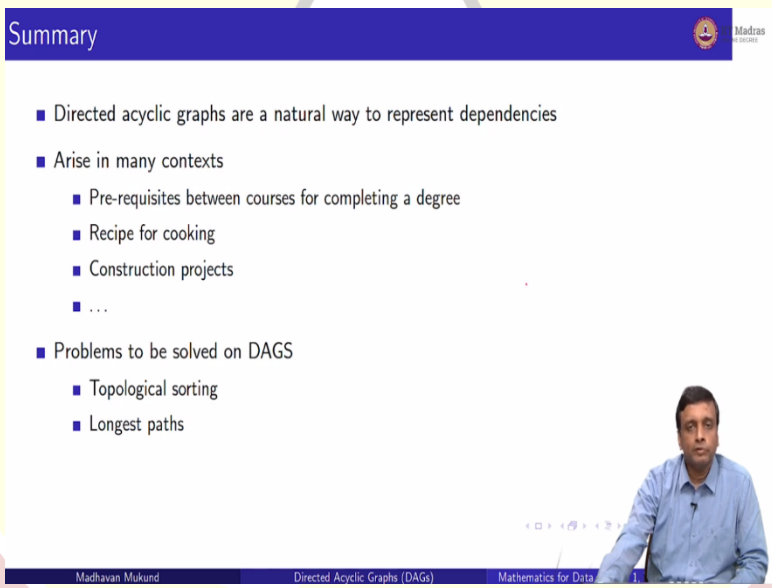
So, if we look at this graph, formally, it is a directed graph. But more importantly, it is acyclic, we do not have any cycles in this because cycles represent dependencies, if a depends on b and b depends on a then which do you do first. So, what we are trying to do is to find a scheduled, which enumerates these vertices in an order, such that in that sequence in the list in which we enumerate these vertices.

If a task i must be done before a task j according to the dependencies, then it must appear before j in the sequence. So, every time we have a dependency, an edge in the graph, that edge, the starting point of the edge has been listed before the ending point indicating with the starting task finished before the second task began. So, this problem formally in a directed graph is called a Topological Sort. So, what we want to do is topologically sort this.

And the second thing is to discover how long we need to move. And then this way, I do not find essentially the longest path. So for instance, we could say that if we start from here, then it is going to take us four steps from starting the conduiting to finishing the cabling. But if we go along this path, for instance, this actually says that we need to do 5 things in a sequence. We cannot do these any faster.

Because plastering can be done only after the conduiting, painting can be done only of the plastering, then wiring and then the fittings. So, we are trying to find the length of this longest path. So, these are the two formal problems that we have with DAGS, topological sorting, and longest path.

(Refer Slide Time: 09:19)



So, to summarize, directed a cyclic graphs are a natural way to represent dependencies. The direction of the edge indicates the direction of the dependency, what must come before what. The fact that it is acyclic follows from the fact that if you have a cycle of dependencies, if i depend on u, and u depend on somebody else, and that person depends on me, then we all depend on each other, so we cannot get started.

So, if I am waiting for you to finish and you are waiting for somebody else to finish and that person is waiting for me to finish, who goes first. So, these cycles cannot be these dependencies cannot form a cycle therefore, it must be a directed and acyclic graph. And these arise in many contexts.

So, we saw this context where we had to finish a room. It could also represent for instance the sequence in which you take courses to complete a degree.

So, courses usually come with prerequisites. So, you cannot do Maths 2 before Maths 1, maybe you cannot do the ML for computation or Computing for ML course, unless you have finished Python programming and both the math courses and both the stats courses and so on. So, now, if you have prerequisites like this, then find a sequence in which you can take the courses to complete the degree.

Cooking is another constraint, a place with a lot of constraints, you need to first of course, make sure you have the ingredients. So, there will be typically a list of ingredients, then there is some processing to be done before you have to chop some things you have to make some things you have to grind some things and so on. And then after that there is a specific sequence in which things go into the pot. So, you put some oil and then you do something else and so on.

So, there are certain things that can be done in parallel 1 person can be chopping the vegetables while somebody else is grinding up something, but there are some things which have to follow a sequence. So, cooking recipes also impose a natural dependency on the tasks in order to prepare a dish. And finally, the kind of problems that we looked at is like a typical project a construction project or any other large project which has many phases, and these phases, some of them can be done in parallel some have to be done in sequence.

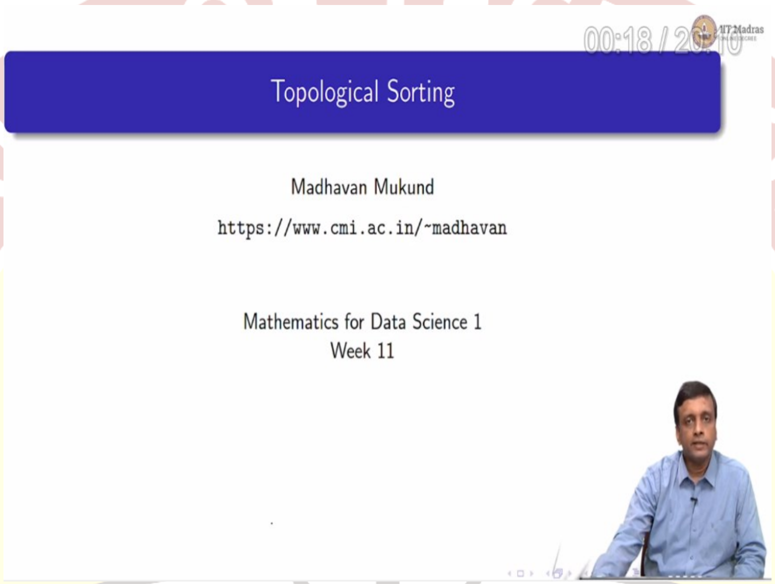
And once we have modeled these things as DAGS, we can solve a similar problem that arises across all these different applications by a uniform problem on DAGS, namely topological sorting and longest paths. So, this is what we will be looking at.

IIT Madras

ONLINE DEGREE

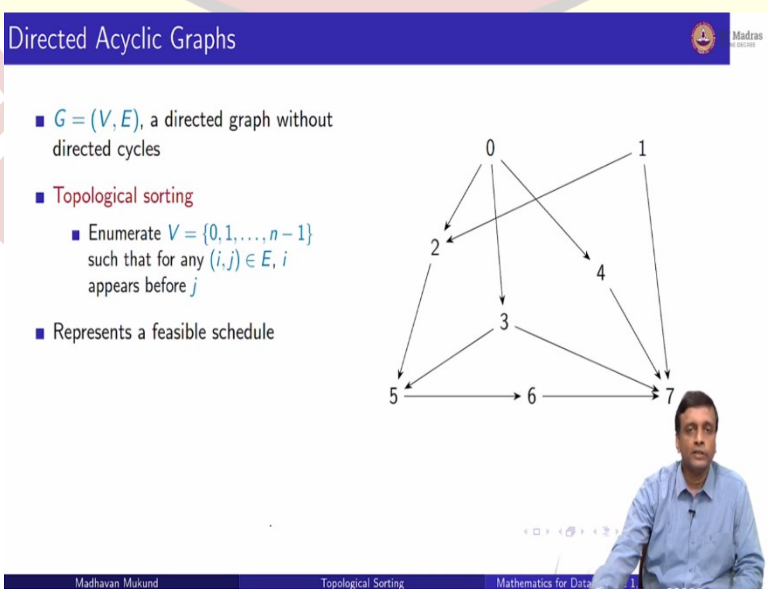
Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Indian Institute of Technology, Madras
Lecture No. 68
Topological Sorting

(Refer Slide Time: 00:14)



So, we have motivated the use of DAGS by saying that they are useful for representing tasks and dependencies.

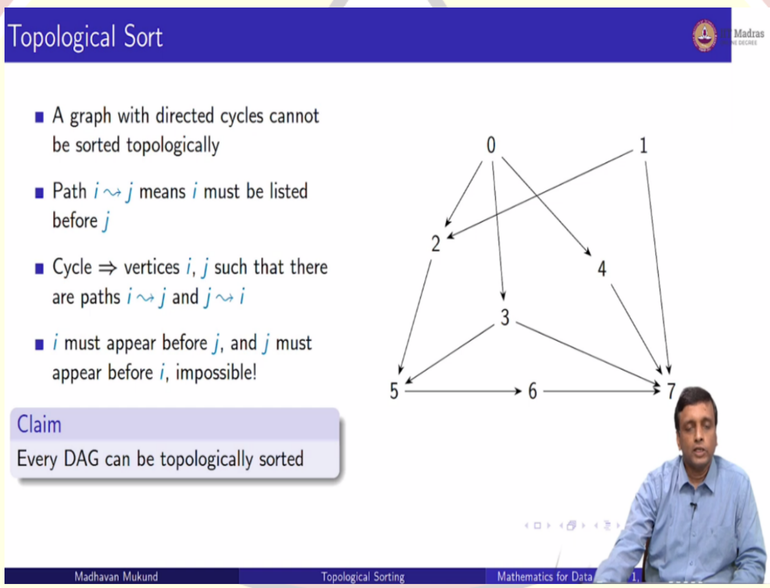
(Refer Slide Time: 00:21)



And 1 of the things that we need to do in a DAGS in a directed acyclic graph is to arrange the vertices in a list that respects the dependencies. So, we said this is a Topological sort. In a Topological sort of the vertices, we sequentially list out the vertices in such a way that every time there is a directed edge from i to j, i must appear before j in the sequence that we list out.

So, in terms of the applications that we are thinking of, if we think of these as tasks and dependencies, then this type of a topological sort represents a feasible schedule. A schedule in which no task appears before all the tasks it depends on are already completed. So, when you come to do something, everything you need to do before that is already done.

(Refer Slide Time: 01:04)



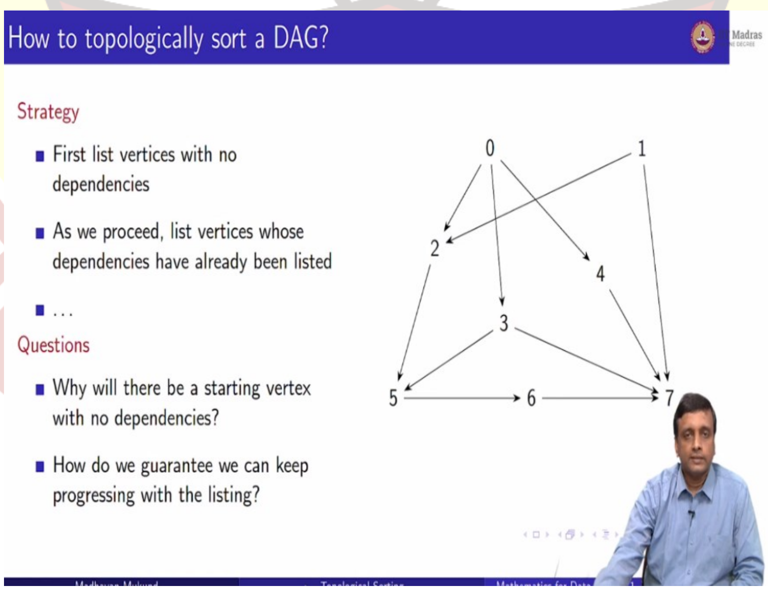
So, the first thing to notice is that if you have cycles, then you cannot do this. If you have a graph with directed cycles, we informally argued that if i depend on somebody else to finish, and somebody else depends on me to finish, then we cannot do anything because we are waiting for each other to finish. But formally, let us see what it means. So, in the topological sort, it is not just for every edge, for every edge i comma j, it is clear that topological sort requires i to appear before j.

But in general, if I have a path of dependencies, if i depends on k and k depends on j, then i must come before k and k must come before j. So therefore, transitively i must come before j. So, anytime there must there is a path from i to j it must be listed before j. So now, if I have a cycle, in a directed graph, it means that I can go from some i to some other j. So remember, a cycle cannot be just something which goes from i to i without going to any other vertex, we need to go at least cross 1 edge. So, I must go from i to some other vertex j and then come back. So, there is a path from i to j and a path back from j to i.

Now, by the previous requirement, if there is a path from i to j, then the topological sort is obliged to put i before j. But since there is also a path from j to i, then we have to put j before i. But clearly in a given sequence either i can come before j or j can come before i we cannot have both these constraints satisfied in the sequence, and therefore this would be impossible. So, that is why if we have a cycle of dependencies, there is no way to order them in a feasible sequence, such that each task appears only after everything it depends on has happened before.

So, what we are going to do now is to show the other side of this, so what we said is if it is not a directed acyclic graph, so it is directed, but with cycles, then topological sorting is always impossible. On the other hand, what we need to argue is that if you give me a feasible set of constraints, in the sense that there are no cycles, there are no cyclic dependencies, it is a DAG, then I will always be able to complete it in some reasonable order. So, every DAG can be topologically sorted.

(Refer Slide Time: 03:14)



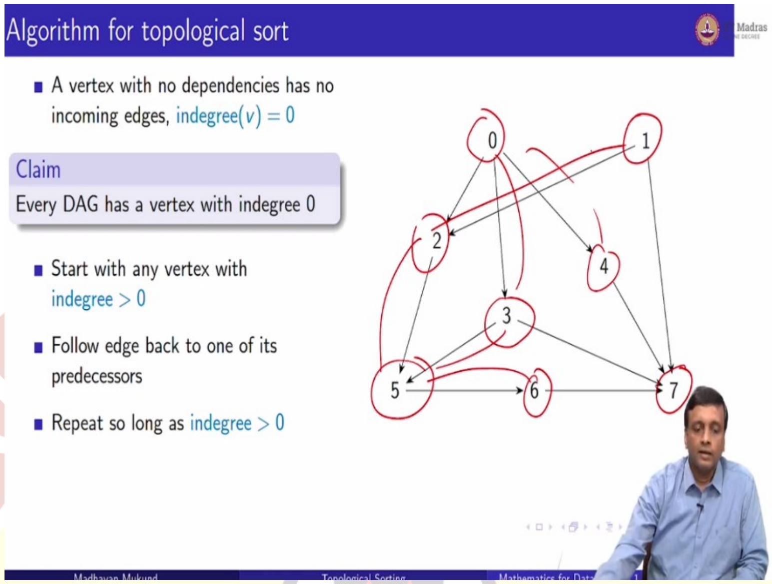
So, how would we go about doing this? So, clearly I have to begin somewhere, so the first thing I have to do is list out a task, which has no dependencies, it does not require anything else to be done. So, there must be a vertex with no dependencies, there may be more than 1. If you look at this graph on the right, we see that 0 has no incoming edges. So, 0 does not depend on anything. 1 has no incoming edges. So, I could start with 0 or I could start with 1. But in general, I need to find such a vertex which has no incoming dependencies to start with.

As I complete the dependencies, a 3later vertex with depends on a few things now becomes available, because everything that needed to be done before that is done. So, as long as we can find vertices whose dependencies have already been listed, we can then list these. So, this is our general strategy. So, we first start with something which has no dependencies. And every time we have a dependency satisfied, we strike it off the list. So, then vertices, which do have dependencies, eventually, all the dependencies have already been listed, and then I can list them.

So, in order to apply the strategy, we need to of course, guarantee that there will be a starting vertex with no dependencies. Otherwise, there is no way we can start. And then we also have to guarantee that eventually, every vertex which has dependencies will find all those dependencies listed out. So, we have to make sure that every vertex, remember that we said that whenever we have a DAG, we can do a topological sort, this is our claim.

And this strategy above says that we are going to do this by starting with a vertex with no dependencies. So, first we have to show that such a vertex exists. And then we have to argue that as we progress, we are definitely going to eliminate all the vertices in our into our list and finally finished by doing all the tasks or listing them all out in this topological sort. So, we need to somehow justify these claims in order to proceed with this strategy.

(Refer Slide Time: 05:09)



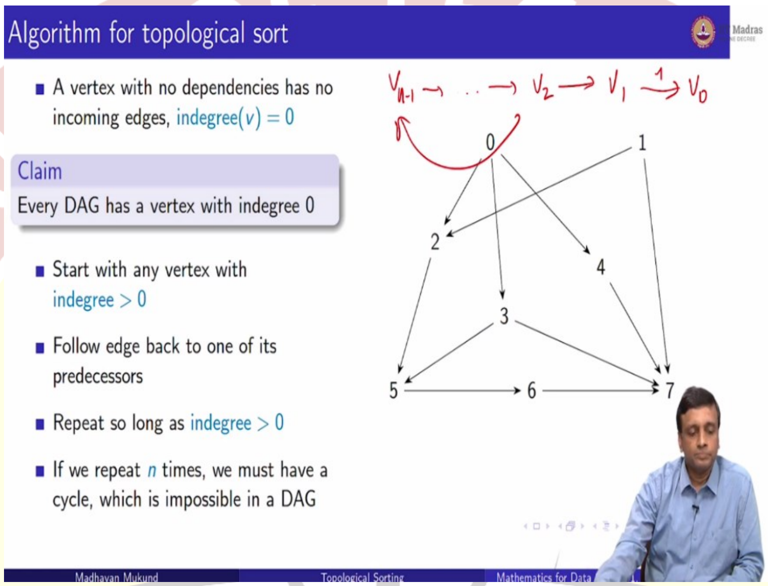
So, remember that in a directed graph, we talk about the indegree and the out degree of a vertex as opposed to just the degree. So, in an undirected graph, the degree of a vertex refers to the number of edges which are incident on that vertex, how many edges have that vertex as an endpoint. But in a directed graph, these edges could either come in, or they could go out. So, the indegree is how many edges are pointing into a vertex v.

What we are looking for is a vertex with no dependencies, that means nothing is pointing into it v does not depend on anything, so there is no edge of the form u comma v. So, why must there be such a vertex? So, the claim is that every time must have such a vertex within degrees 0. So, let us suppose we have vertices within degree not 0, so pick any 1. So, we start with the vertex v, which has indegree greater than 0. Since it has indegree greater than 0, there is at least 1 edge coming into that vertex. So we can follow that edge backwards and go to a preceding vertex. So, we can go back from v from a proceeding vertex.

So, let us say supposing we start here, we say that this vertex, has some incoming edge. So, I go across this vertex, I go backwards and I come to this vertex. Now I say this vertex also has a nonzero number of incoming things it has integrated in 0. So, I will keep doing that. So, I will start here, then I will go back here and then maybe I will go back here. And then maybe I will go back here. And then I stop, because I cannot go back any further.

So, in this particular case, this graph is acyclic and I have stopped at 0 by starting at 6. If I started at 7, I could follow a different path. For example, I could go from 7, I could go to say, 4 and then 0, or from 6, I could have gone from 5 to 2 to 1 and so on. But whichever way I do it, eventually all these paths will have to stop. And why is this the case?

(Refer Slide Time: 07:11)

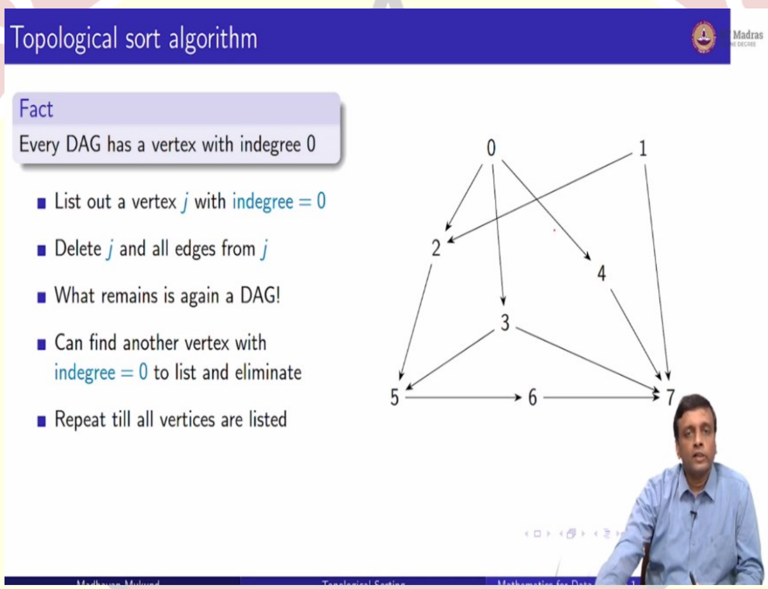


Well, supposing I started some v, let me call it v_1, or v_0, if you want it, then I come back to another vertex, which is v_1. And then I cannot stop because v_1 has indegree greater than 0, so I have to go back to another vertex which is v_2. Now, if I hit the same vertex again, then I have a cycle. So, let me assume that v_1 is different from v_0 and v_2 is different from v_1 and v_0. So I am continuously hitting new vertices as they go along. If I do not hit a new vertex, every time I go backwards, I have already found a cycle and I know this graph has no cycles, but on the other hand, this graph has only n vertices.

So, if I started v_0 and I do 1 step, I get back to v1, if I do 2 steps, I get to v_2. So, if I do $n - 1$ steps, I have reached $(v_n - 1)$. So, after I have done this, if I do it one more time, if I do an nth step backwards, then I cannot find a new vertex anymore, which I have not seen before, because all the n vertices in my graph have already been traversed somewhere in that path that I have seen so far. So therefore, the new vertex must be going back somewhere here. So, it must be one of the vertices already seen. So, there must be a cycle.

So, there is a directed cycle. And since it cannot be a directed cycle, this cannot happen. So, this is a complicated way of proving something is called proof by contradiction. So, you say assume that everything has an nonzero indegree, then I can find a path, which is arbitrary length, in particular of length n, which will visit $n + 1$ vertices. And since n + 1 vertices must repeat a vertex, there must be a cycle and this cannot happen. So, this is why we will always have a starting point, we will have always have a starting point, which is an indegree vertex with a vertex with indegree 0.

(Refer Slide Time: 08:56)



So, this claim is now a fact. So, that was a proof that we have a vertex which is guaranteed to have indegree 0. So, in this particular graph, as we said, we have the vertices labeled 0 and 1, which both have no edges pointing into them. Now, what do we do? Well, we list it out, because we start from there. And once we list it out, we kind of pretend that it is no longer a constraint because it has been done.

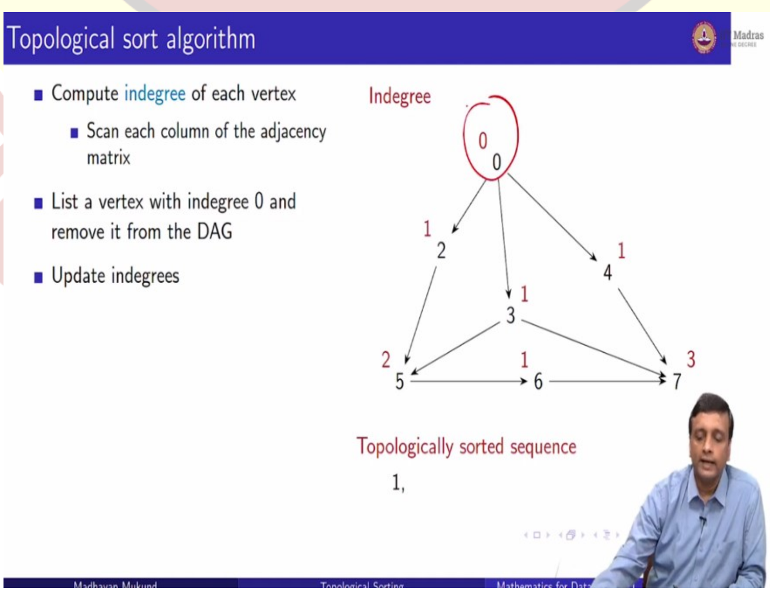
So, if it was a constraint for somebody, for example, if we look at vertex 4, for instance, so vertex 4 requires vertex 0 to be completed, but if I list out 0 saying 0 is done, now, 4 has no longer any constraints, because this constraint is gone. So, this was a claim that as we go along, the constraints will go away until you can list out the later vertices. So, if I delete that vertex that I just found, and all the edges from j, what happens?

So, if I delete this, and then I delete all the edges that point out of 0, then I am left with a smaller graph in which there is 1 less vertex, and a few small, fewer edges depending on how many edges I had connected to that original vertex. But notice that in this process, I have only removed edges from a directed graph. And if the original graph had no cycles, this must also have no cycles, because I have not put back any edge between 2 vertices which are not already connected. So, what remains after this is again a DAG. So, I take a DAG, I remove any vertex from it, it remains a DAG it may not be connected for if I have done it badly, but at least it cannot have any cycles. So, it will be directed and it will be acyclic.

And therefore, by the same argument in the new DAG, that is, after I have believed this constraint has been satisfied, I must again have some vertex with indegree 0. So, at every stage, I have a DAG, whenever I have a DAG, by that earlier argument, there must be a vertex with indegree 0. So, at every stage, there are at least 1 vertex which I can remove. And I keep doing this and after n minus after n stages, I must have removed all the vertices. So, this is how the procedure works.

So, we repeat this process until all the vertices are listed. And we are guaranteed that at every stage, start with a DAG, remove a vertex of degree 0, I have another DAG, therefore, I have another vertex of indegree 0, remove that I have 1 more, and so on. So, that is why this process is guaranteed to make progress, and is guaranteed to exhaust all the vertices in my DAG.

(Refer Slide Time: 11:24)



So, the first step to implement this as an algorithm is to compute the indegree. So, assume that we have the graph presented to us as usual as an adjacency matrix, then we know that the incoming edges are in the columns. So, remember that the row i has all the outgoing edges from i, and the column i has every entry of the form j comma i, which is edge from j to i. So, if I look at the column i it has all the entries for the incoming edges.

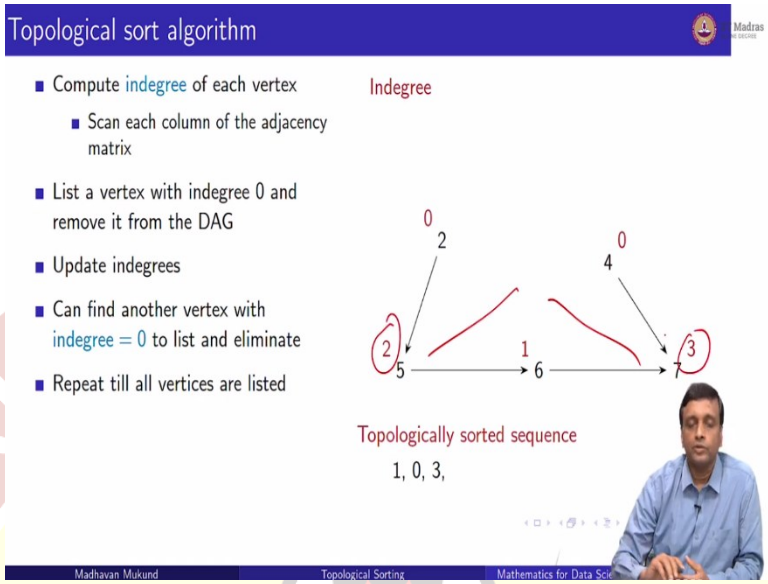
So, if I just walk down column i and add up the 1s, remember that this matrix has 0 1 entries, if I just add up the 1s, I will get the indegree. So here, we have got this graph and by doing this one scan, although we can do it pictorially in this particular case, by doing this one scan, we can count the incoming arrows. So for instance, this has degree indegree 2, because there are 2 edges coming in this as indegree 4, because there are four edges coming in. But this is not something we have to do with the picture, we can actually just mechanically do it using the adjacency matrix.

So, now that we have this, now we can compute. So, we have an alternative, a second list, in some sense a list of indegrees of every vertex, we are, we know in advance that there must be at least 1 of these vertices which has indegree 0, we do not know which 1, but we can find it by just scanning down the list and looking for a 0. So, we go down the list and we look for a 0 and perhaps we decide to pick so there are 2 in this case, as we know.

So, we have both what is 0 and 1. So, let us suppose we choose to do this one. So remember, our procedure says list it out and remove it from the graph. So, we list it out here to the bottom is our list, and we remove it from the graph. So, the edges, which are now pointing out of vertex 1 have been removed. And in this process, the targets of those edges indegree has reduced. So, when I do this, I had I had an edge I claim like this.

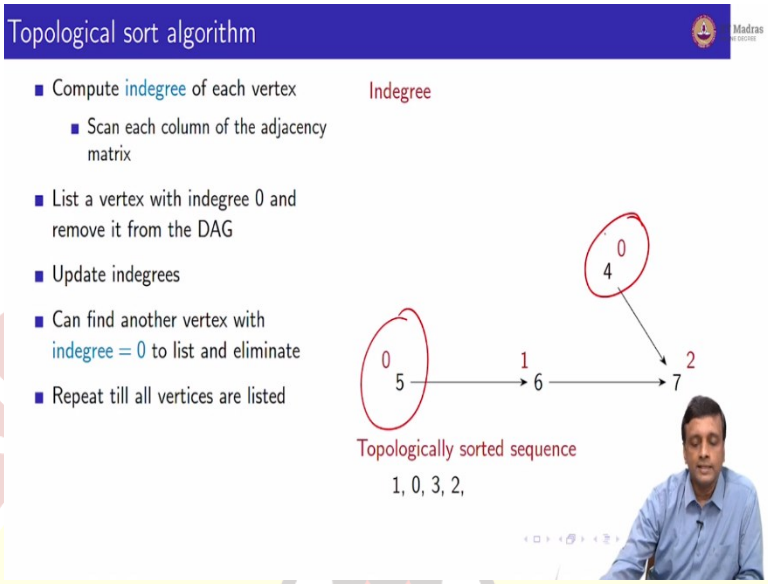
So, this indegree and this indegree now will change. So, when I remove that vertex, I must also simultaneously update the indegrees, I do not have to scan all the indegrees. Again, I only have to look at the row for i the vertex i just deleted as i, I only look at the row for i and every ij that I have as an edge, I look at the degree of j and I reduce it by 1. So in this case, I had 1 to 2 and 1 to 7. So I go to vertex 2 and reduces indegree by 1, I go to vertex 7 and reduces indegree by 1. Now again, I have a DAG a smaller dag, again, I must have a vertex, at least 1 of indegree 0 here, I have no choice, I have only this one. So, I remove that and list it out. And now again, these are the three vertices which were getting edges from the vertex 0, which has this just deleted.

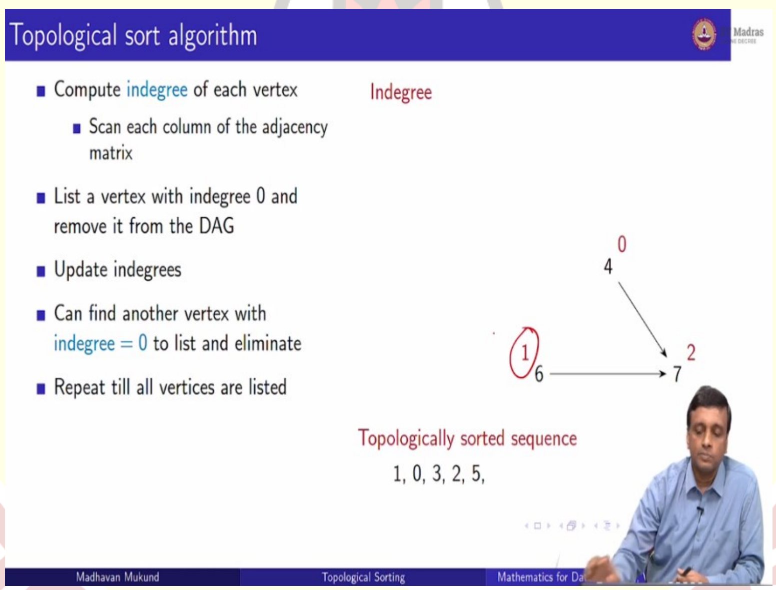
(Refer Slide Time: 14:10)



So, now I have to reduce they are indegrees by 1. Now I have a wide variety of vertices with indegree 0, which I can enumerate next, because all of them depended on 1 and 2. So, I pick any 1 of them. So, for instance, I picked 3, the one in the middle, I picked 3 and I remove it, so I list it out. And now 3 was pointing in this direction and this direction, so this 2 will have to reduce and this 3 will have to reduce. So, I reduce them. Now again, I have 2 choices, 2 and 4. So, perhaps I do 2 next.

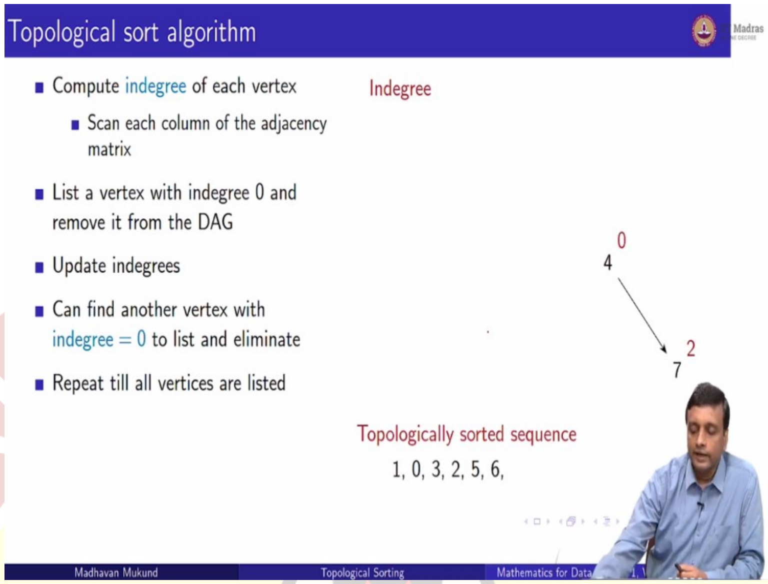
(Refer Slide Time: 14:42)

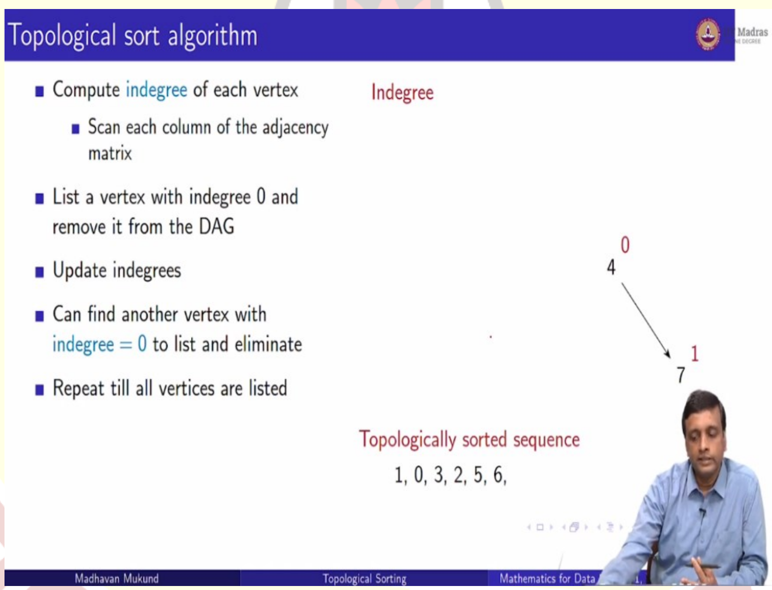


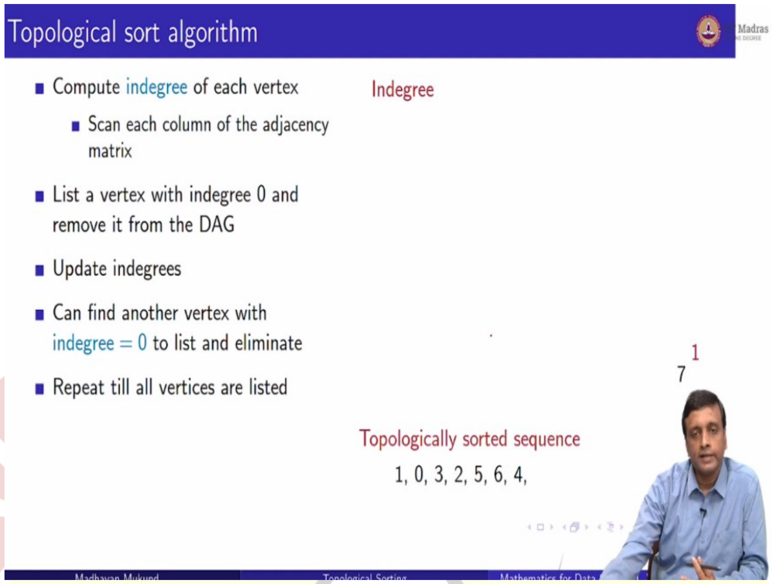


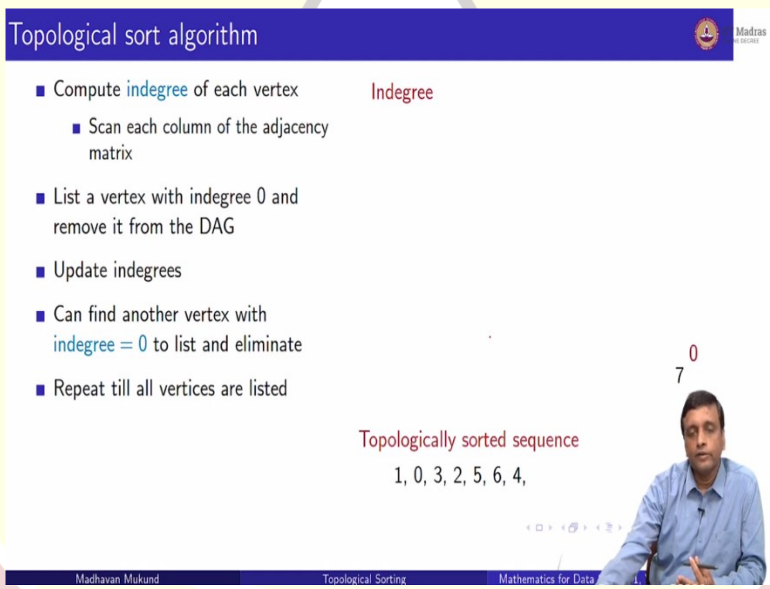
So, I take 2, and now I reduce the indegree of 5 by 1. Now I have these 2 candidates to enumerate next, so perhaps I seek vertex 5. And then I have to reduce the indegree of 6 by 1. And now maybe I do 6 next.

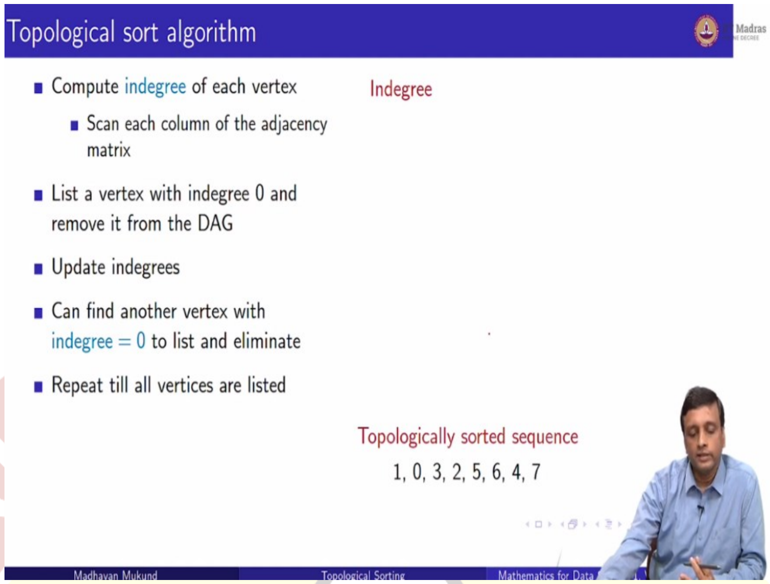
(Refer Slide Time: 15:02)





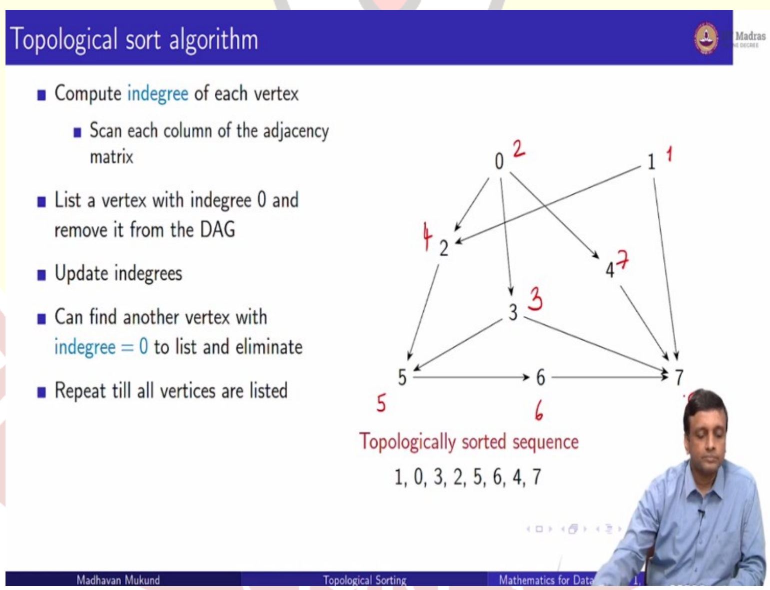

Topological sort algorithm
Madras
Compute indegree of each vertex
Scan each column of the adjacency matrix
List a vertex with indegree 0 and remove it from the DAG
Update indegrees
Can find another vertex with indegree = 0 to list and eliminate
Repeat till all vertices are listed
Indegree
1
7
Topologically sorted sequence
1, 0, 3, 2, 5, 6, 4,
Madhavan Mukund
Topological Sorting
Mathematics for Data


Topological sort algorithm
Madras
Compute indegree of each vertex
Scan each column of the adjacency matrix
List a vertex with indegree 0 and remove it from the DAG
Update indegrees
Can find another vertex with indegree = 0 to list and eliminate
Repeat till all vertices are listed
Indegree
0
7
Topologically sorted sequence
1, 0, 3, 2, 5, 6, 4,
Madhavan Mukund
Topological Sorting
Mathematics for Data



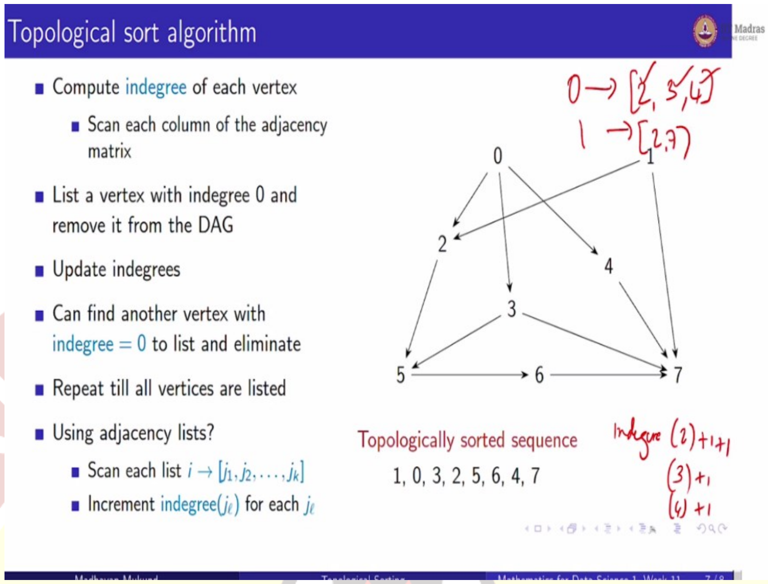
And then I reduce the indegree of 7 by 1. But I still cannot enumerate 7 because it has indegree 1, but 4 is left with indegree 0. So, I can enumerate for next, reduce the indegree of 7 to 0. And finally, I can enumerate 7.

(Refer Slide Time: 15:22)



So, if we look at our original graph, this is what the graph looked like. So, what we have said is that we did this first, then this, then this, then this, then this, and this, then this. And then. So, this is the sequence in which we enumerated it, perhaps not the most obvious sequence that you would have thought of, we might have thought of doing it top to down to 01, and then maybe 234, and then 567. But this is a valid sequence.

(Refer Slide Time: 15:52)



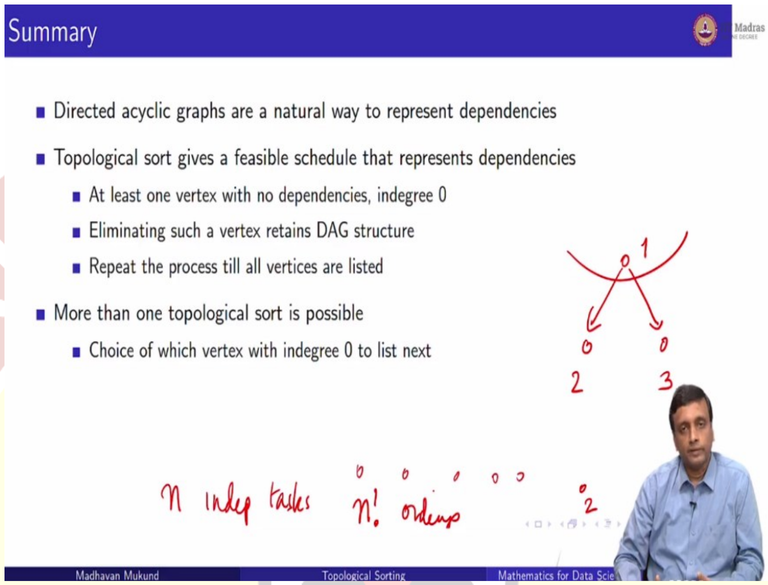
So, what if we had adjacency lists instead of adjacency matrix, so we said did an adjacency matrix representation, we can find the incoming edges by looking at the relevant column, we look at column i, and we have done it. In adjacency list, we only have outgoing edges. If I look at the list for i, it has only edges pointing out of i. So, how do I get the edges pointing into i. Well, you do not do it in one shot.

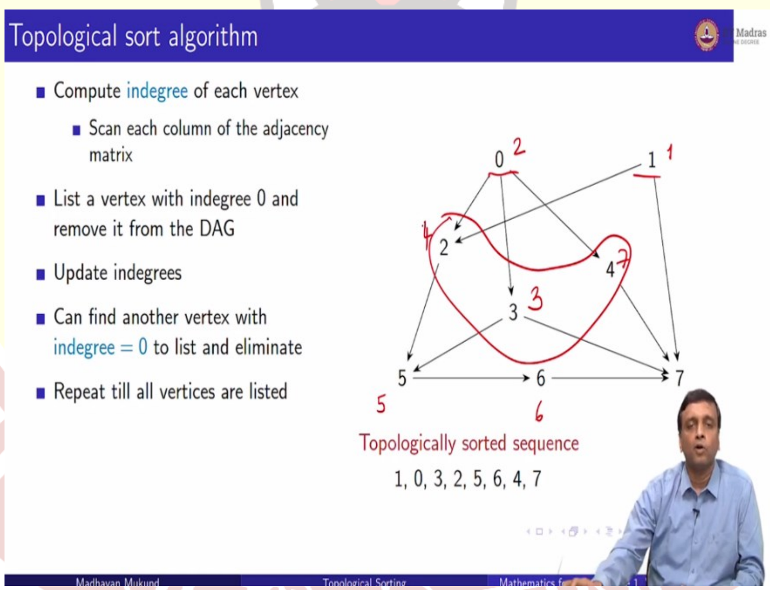
So, for the column thing, you compute the indegree of i in one shot by looking at the column for i, here you look across all the lists, so you will start with vertex 0, and look at the list of things that 0 is pointing to. So, 0 is pointing to, in this case, 2, 3, and 4. So, what you will do is you will have separately indegree of 2, indegree of 3 and indegree of 4. So, when I see this 2, I will do a + 1 here, when I see this 3, I will do a + 1 here, when I see this 4, I will do a + 1 here.

Now I come to 1, so 1 has outgoing 2 and 7. So, now I will do another + 1 here, and in 7, I will do a + 1, and so on. So, you basically scan all the lists from top to bottom. And for each outgoing as you see, you go to the corresponding target indegree and incremented by 1. So, even if you have an adjacency list, we can do a simple scan, and update all the indegree to start with. After that it is only a matter of checking the indegree as you are deleting them. So, it does not matter whether it is incoming or outgoing.

So, we have the usual caveats as we had before, that is if you are doing an adjacency matrix, everything takes order n time because you have to look at all the outgoing edges or all the things but as we are doing something, which is the adjacency list, you can do it in time proportional to the total number of edges.

(Refer Slide Time: 17:20)





So to summarize, we have seen from the earlier lecture, that directed acyclic graphs are a very natural way to represent dependencies. And one of the fundamental problems in such a situation is to find a feasible schedule, find a sequence in which you can perform the tasks or do whatever we need to do, which does not violate any of the constraints. So, something must be listed before something that that depends on it.

So, what we observed is that in any DAG, there has to be at least one vertex which has no dependencies, there is something within the graph version on the directed acyclic graph representing the dependencies has indegree 0, so we can list it because it has nothing that has to come before it. And eliminating a vertex from a DAG gives us back another DAG smaller DAG, possibly disconnected.

So, we could have a DAG for instance, we look like this, this is a DAG. So, it says that I must do this, say this is my first task. And I must do this before 2 and 3. Now once I have done this, now I have a DAG, which consists of just 2 and 3. So, we might end up with a disconnected graph, but it is still without cycles. And therefore, by the same logic, it must have something of indegree 0 and so we can keep repeating and that is why this process works.

Now, the other thing to notice is that more than one topological sort is possible. So we saw that when we looked at that example of how to set up the room, we said that for instance, tiling the floor and plastering the walls can be done in either order. And in particular, what happens is that when we end up with multiple choices for indegree 0 vertices.

So, if we look at our previous example, for instance, I could have started with 0 or with 1 we chose to start with 1 if we started with 0 we would have got a slightly different sequence. Similarly, we had a situation when we had 2 3 and 4 all available to us with indegree 0. So, we chose to do 2 first we could have done 3 first, we could have done 2 first we could have done 4 first.

So, whenever we have multiple degree vertices with integral 0 topological sort does not necessarily force us to take one or the other we might choose to take the smallest one in which case we get one particular order, but there are multiple orderings possible. So, this is a thing that we need to remember that Topological sort produces a sequence which is compatible, but this is by no means the only sequence there are multiple topological orderings possible.

In particular, if you have no dependencies if I have all the tasks are independent, that any ordering is possible. So, if I have this is basically if I have n independent tasks, then I would have $n!$ orderings. So, the number of topological orderings can be very large so we are not really interested in computing at this point the number of topological ordering we are also interested in finding one of them.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

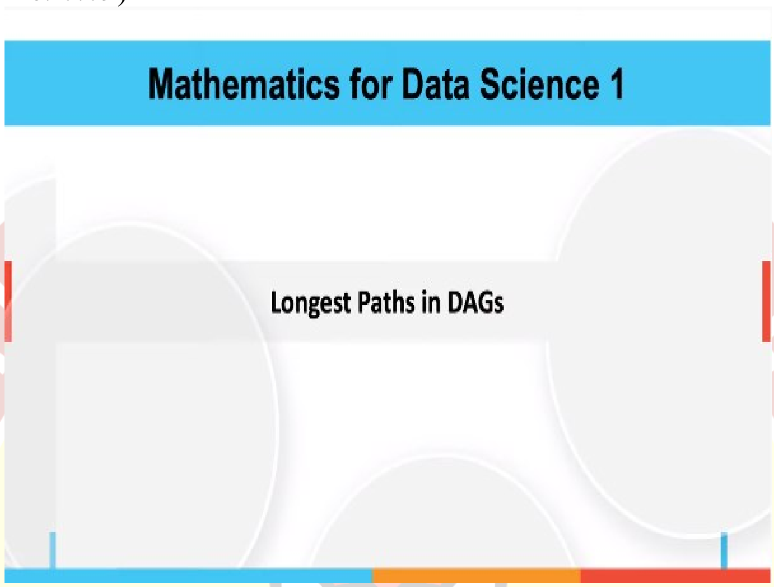
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

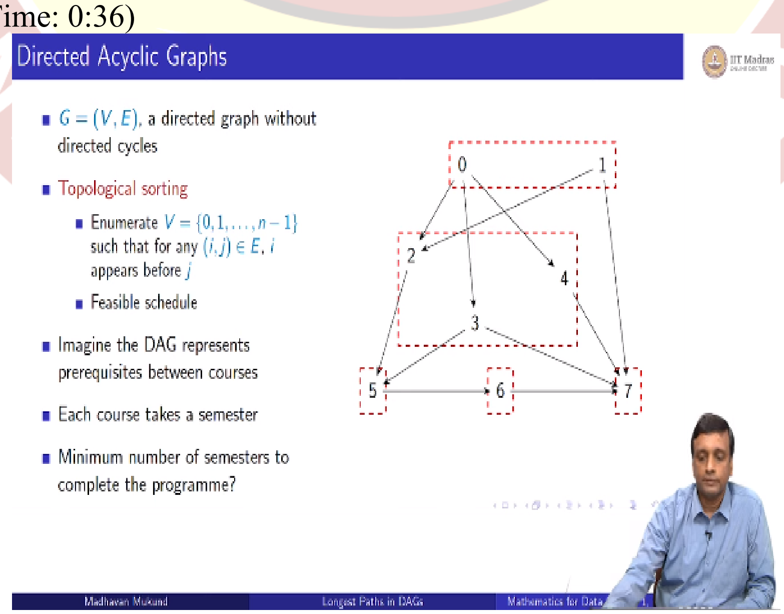
Mathematics for Data Science 1
Professor Madhavan Mukund
Indian Statistical of Technology, Madras
Lecture 69
Longest Paths in DAGs

(Refer Slide Time: 0:09)



So, we are looking at directed a cyclic graphs as representations of sets of tasks with dependencies. And we said that there are two natural problems on this. One, is to find a sequence of feasible sequence in which I can do the tasks and this was topological sort. The other one was to try and find out how many steps I need to perform these tasks in an optimal way.

(Refer Slide Time: 0:36)



So, this is my DAG. So, if I do a topological sort and I list out the vertices such that whenever there is an edge from i to j, i appears before j and this gives me a feasible schedule. Now, we may be interested in finding out how fast we can do this if tasks which have no dependencies can be done together.

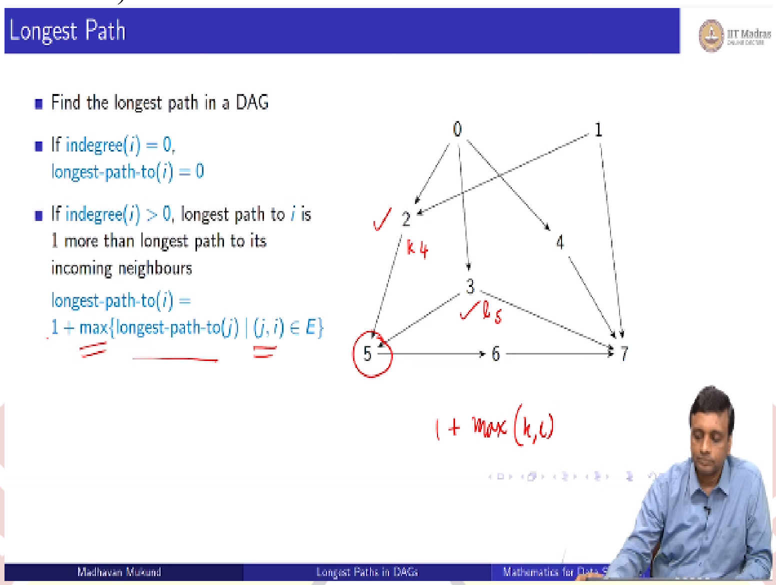
So, for instance, in a topological sort, we would have to put 0 before 1 or 1 before 0 as a sequence. But if we had no dependencies, and it is possible to do 0 and 1 parallelly, then we could do 0 and 1 at the same time, if there are resources to do both. So, a good example is when you are taking courses, so when you are taking courses, you do not do just one course in a semester, you can do many courses. And as long as the number of courses which are available for you to take are reasonable, you can do all of them, say three, or four or five, maybe in a semester.

So, if the DAG represents prerequisites between the courses, and each course takes a semester, so you can finish and go to the next course only in the next semester, then the natural question to ask is how many semesters do I need to complete the remaining requirements? So, I have a set of requirements, and they have some prerequisites between them, how many semesters do I need from now to finish the program satisfying these requirements. So, this is the problem that we want to solve now.

So, in this particular case, for instance, as I said, we can do 0 and 1 together. So this can be done in the first instance, then I can do 2, 3 and 4. In these are courses, then in the first semester, I can do 0 and 1 course, in the second semester, I can do courses 2, 3, and 4. But now I am stuck because I cannot do 7 until I finished 6, I cannot do 6 until I finished 5, 5 I can do because all the prerequisites are 5, namely 2 and 3 are done. So, in the third semester I can do 5, then in the fourth semester I can do 6 and finally in the fifth semester I can do 7.

So, if this was my sequence, I mean, this was a DAG representing my prerequisites, then I cannot do better than a sequence of 5 semesters. So this is the problem that we want to compute.

(Refer Slide Time: 2:41)



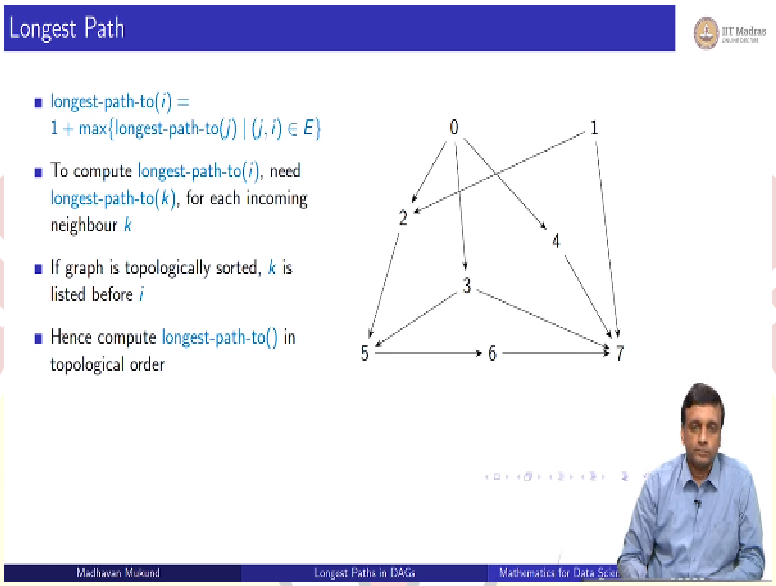
So formally, it consists of finding the longest path in a DAG. So, what we want to do is really compute the longest path to every vertex, and then if we have that, then the longest path among these will be the longest path overall, if I compute the longest path from one of the starting points, so when can I do course 6, when can I do course 3? If I know this, then among all these, if I have the maximum, some course requires me to do it in the fifth semester, then I know overall, I need five semesters. So, one way to solve the longest path problem is to solve this question of computing the longest path to each vertex.

So, with the assumption that the longest path to an initial vertex, so the path is how many semesters before in the course that example, how many semesters do I have to wait? So, with 0 waiting, I can do anything which is indegree 0, so that is a good starting point.

So, in this case, I could do these two things right to begin with, so this is indegree 0. Now what happens next. So, if I have indegree which is not 0, then supposing I look at this vertex, it can happen only after 2 and 3 have happened. So, if I know that the longest path to 2 is some k and the longest path to 3 is some l, then I must wait maximum of k and l, I cannot finish 2 and 3 until max of k and l happen.

So, supposing this is fourth semester, fourth semester, and this is going to happen in the fifth semester then, only in the sixth semester can I get to course number 5, so, this + 1, so, the longest path to i is going to be 1 + the longest path to every incoming neighbor of i. So, remember this set comprehension notation so, this is the set of all the numbers longest path j for j, i in the edge set, all the incoming edges coming into i, I take that, take the maximum of all of those and add 1 because I now have to go to the next semester.

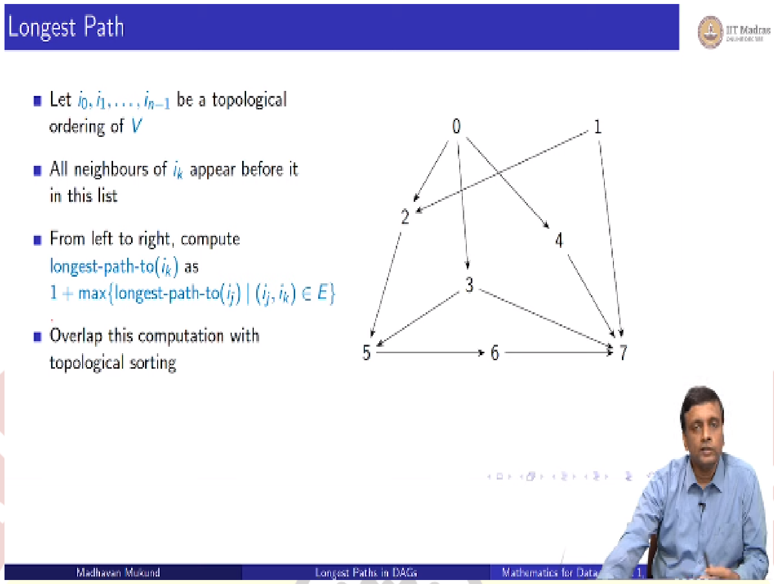
(Refer Slide Time: 4:54)



So, the longest pathway to i is 1 + the maximum of all the longest paths to the incoming neighbors. So, to compute this, I need to know the neighbors, longest paths for all the incoming neighbors. So, if I know that, then I can take the maximum of those and add 1 to it. But how do I know those? Well, I know those if I have already calculated those before this, and I would have calculated those before this if I calculate them following the topological sort.

So, if I sort out the sequences according to the dependencies, then by the time I come to i, and I want to compute the longest path to i, all the incoming neighbors of i should already have been listed before i. And if I am computing longest path as I listed, okay, then that information about all these incoming neighbors will be available to me when I come to i. So, that is the strategy that we are going to do, we are going to compute longest path in topological order. So, we are going to compute the longest path to every vertex as we compute topological sort, in fact.

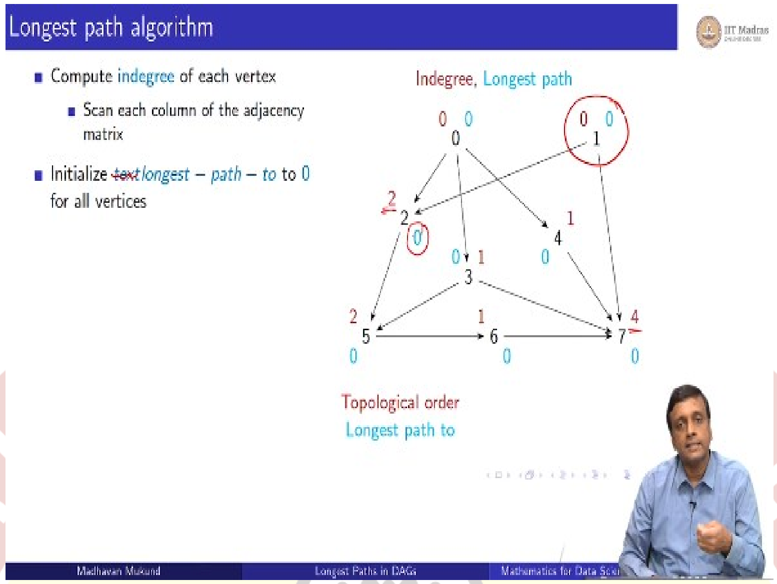
(Refer Slide Time: 5:59)



So, more formally supposing this is a topological ordering of i, of V, right. All the neighbors of some vertex ik, so i0 to in - 1 is some reordering of the numbers 0 to n - 1. So, the vertices are 0 to n - 1 and I have rearranged them in some sequence, and that sequence is i0 to in - 1. If I look at a particular entry, the kth entry in this was a k + 1 entry i_k, then all its neighbors in the graph must appear before it.

So, if I compute from left to right, then I can compute for each of these ik, the longest path based on the values that I have already computed to the left. And since I am doing this as I am going along, I do not have to actually enumerate all the vertices before I compute, as I come to a vertex I already have the information to compute its longest path information, so I can do the overlapping computation of the topological sort and the longest path, I can do them at the same time. While I am computing topological sort, I can simultaneously compute the longest.

(Refer Slide Time: 6:57)



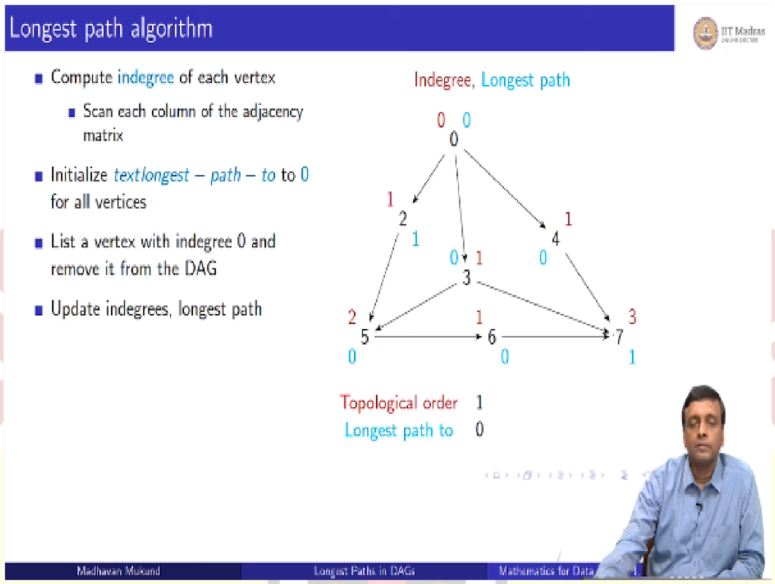
So, how do we do this? As before, we compute the indegree of every vertex, right and we also initialize this longest path, okay to be 0 for all vertices, because initially I do not know anything. So, the blue number indicates my current knowledge about the longest path, right, and the red number is the indegree which we are using for topological sort.

So, now we do this topological sort, as we go along and as we go along we update, so the first vertex, remember we did last time, we will do the same order, we picked 0, I mean, pick 1 which has indegree 0. So, if you have 1 which has indegree 0, then two things happen. One is we are going to update the indegree of everything which is pointing out of it. But we are also going to now update, we have now, we know definitively information about vertex 1. So, this maximum has a kind of monotonic property.

If I have the maximum of a set, it will only get more if I go along, so if I already know that the incoming vertex 1 has longest path 0, I cannot have of course, 0 is in this case a very simple value because a path cannot have negative values, but I cannot, the maximum cannot go below 0.

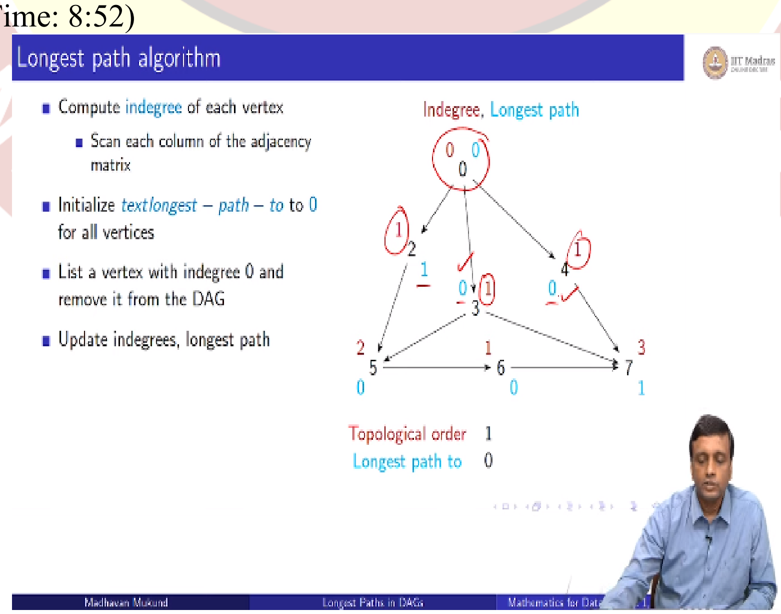
So, I already know that if I know that the, if I have frozen the value for the incoming vertex 1 as having longest path 0, then I already know that the value for 2 must be 1 + that and I can keep updating this. So, I can compute the max of those things incrementally and keep moving on.

(Refer Slide Time: 8:30)



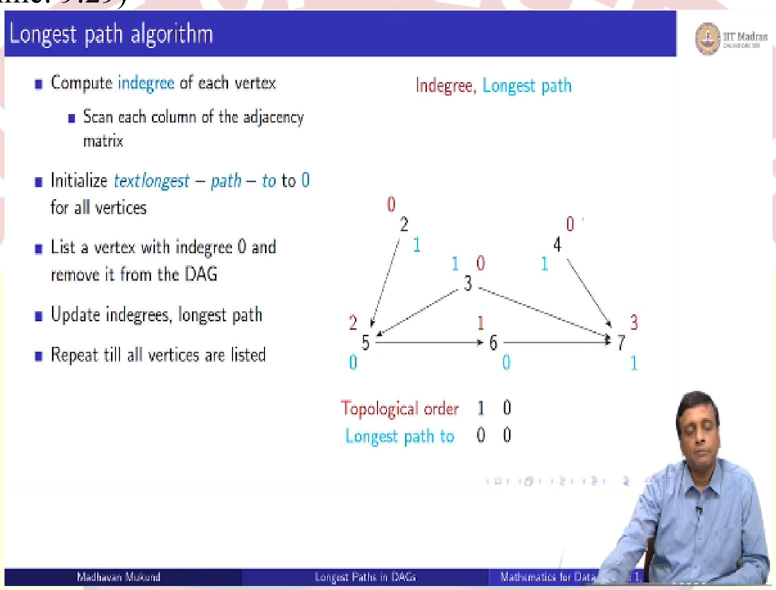
So, what I will do is I will remove that vertex as before and now I will update as I said, these two entries. So, both of these entries, so the 2 will become 1 for the incoming indegree of vertex 2, the 4 will become 3 for indegree of vertex 7. But for both of them, the longest path will now go from 0 to 1 because I know that it is at least 1.

(Refer Slide Time: 8:52)



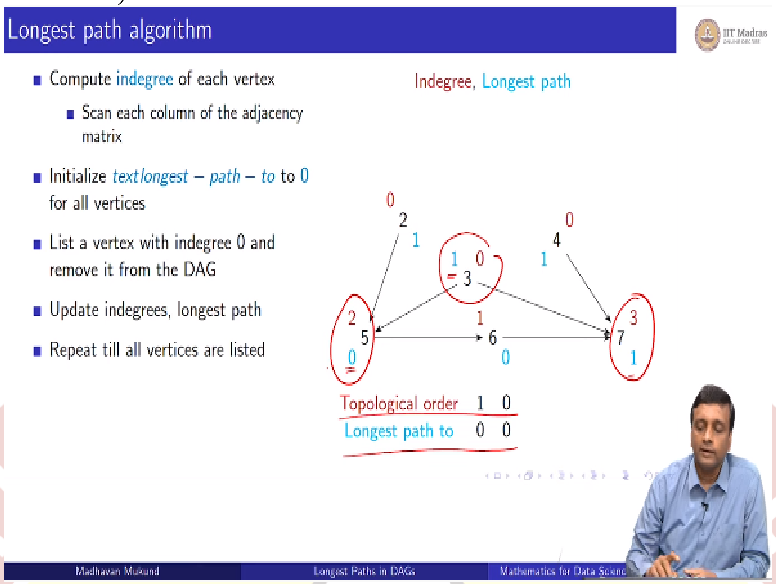
Next, we did this one. So, when we do this one, we will now update as before the indegrees of this, but we will also update these. Now notice that for 2, vertex 2 it already believes its longest path is 1 and if I took the new information from 0, it is 1 + 0, which is 1, so nothing is going to happen, so 1 is going to remain 1, but for 3 and 4, where I had previously believed, I mean this without any justification that the longest path is 0. I am now going to update it to 1 + 0 as 1.

(Refer Slide Time: 9:29)



So, when I remove the 0, the next step is I make all the indegree 0 because they have all now got no vertices pointing into them, but I also update the longest path for 3 and 4. I have also updated for 2 but no change happened because 2 already knew that it was of longest path 1.

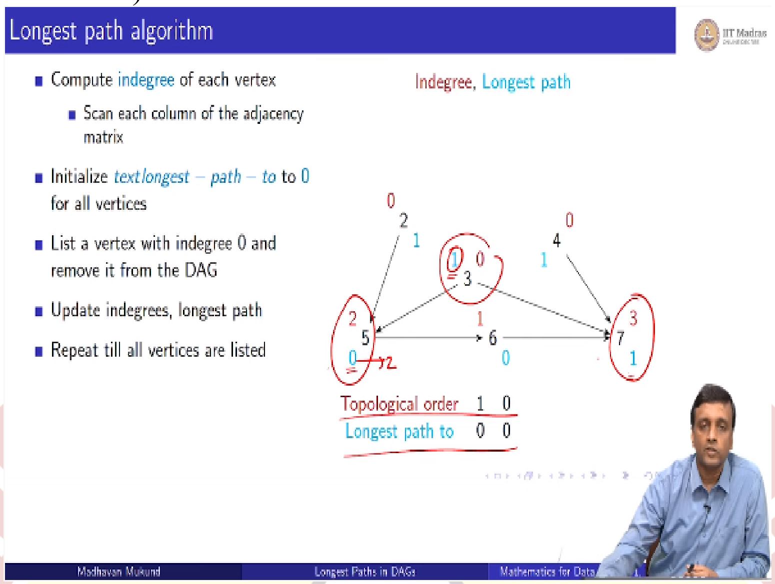
(Refer Slide Time: 9:49)



Next, I will pick 2 in this, next I pick 3, sorry, this is what we did last time. So, notice that on the bottom we are keeping track of this information, right, so we are keeping track of both the topological order, the order this, so the top row indicates the vertex number as I enumerated, so I first enumerated vertex 1, then I enumerated vertex 0 and in the lower row we are keeping track of the longest path, which incrementally we are getting a final value for every vertex as we enumerate.

So, now output vertex 3 and this is going to update these values, so 3 has longest path 1, 7 believes his longest its path is 1, but now it must be at least 1 + 1, 2. So, 7 is going to change similarly, 5 is going to now move from 0 to 1 and of course, the indegree are going to reduce.

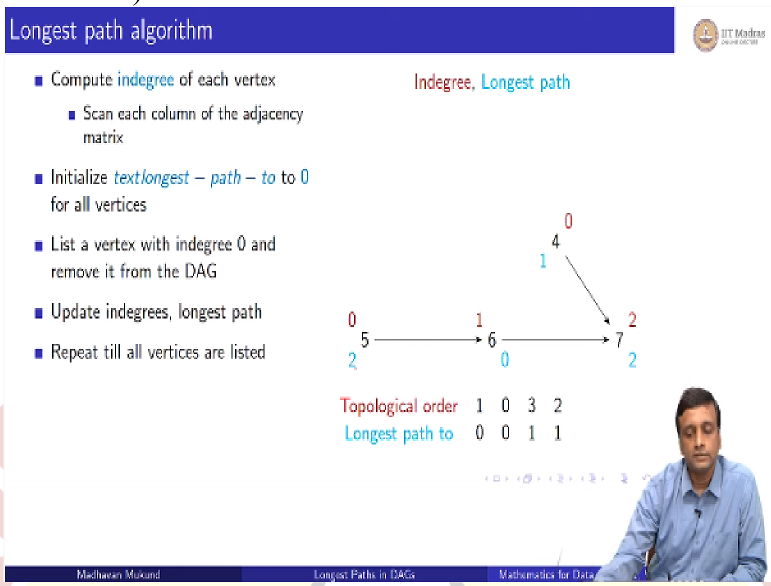
(Refer Slide Time: 10:35)



So, I remove the 3 and then I increment the longest path here, this should be, so this is also notice that see, the path here was 1, so this goes directly from 0 to 2, it goes to something which is 1 + the maximum known incoming thing. So, the maximum known incoming thing is now 1, so it goes to 2 and this was already 1 and it goes to 2 because, so it is not that it is incrementing, it is computing 1 + max, so this goes to 2, this also goes to 2 and the degree has reduced.

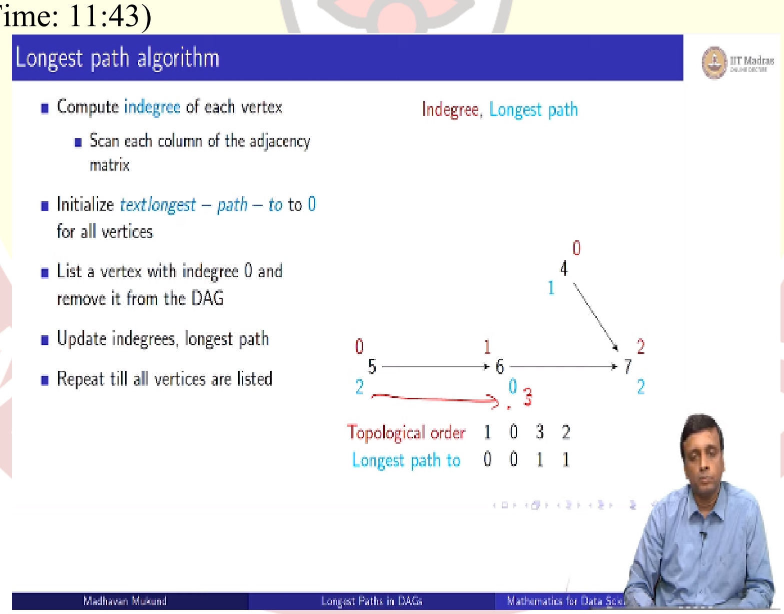
So, we are overlapping this topological sort, with this longest path computation. So, in the topological sort, we are decrementing the indegree each time we remove an edge, in the max, in the longest path computation we are doing a max. So, sometimes it changes, sometimes it does not change, when it does change, it might change by more than 1 and so on, you have to keep track of what was the incoming max that you saw so far and add 1 to it.

(Refer Slide Time: 11:39)



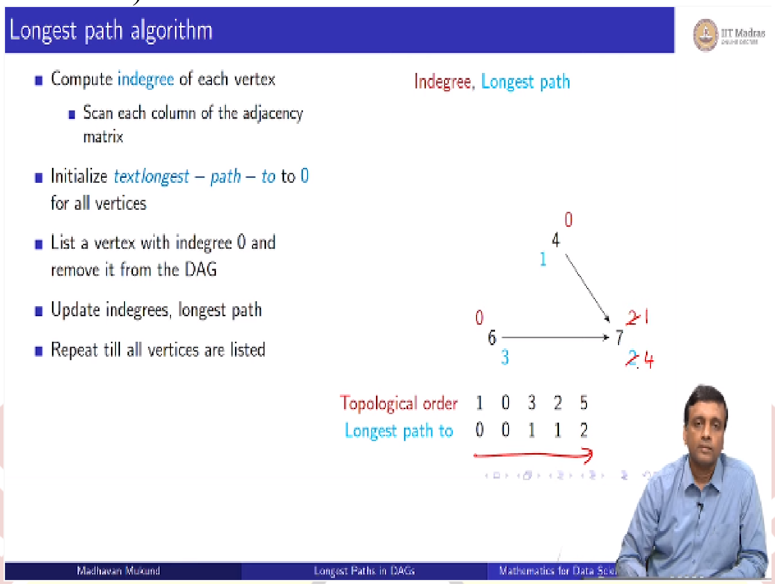
So, next we did 2, so if we do 2, then because 1 + 1 is 2, nothing is going to happen here but the indegree of 5 will reduce, so I remove 2 and now the indegree of 5 reduces but its longest path does not change.

(Refer Slide Time: 11:43)



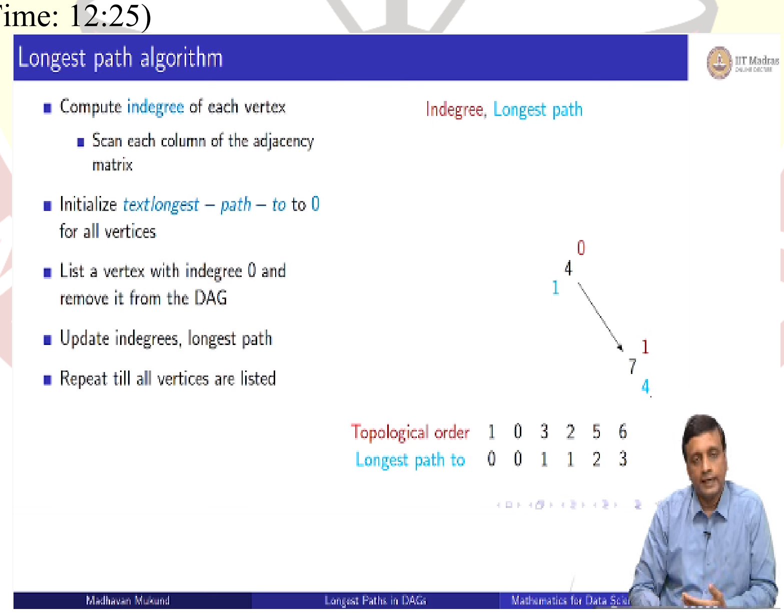
Next, I do 5 and when I do 5, this now influences this, and this says this must become 3, right, because the incoming edge to 6 has longest path 3 at the other end, 2 at the other end. So, the incoming path to 3 must be at least 2 + 1.

(Refer Slide Time: 12:00)



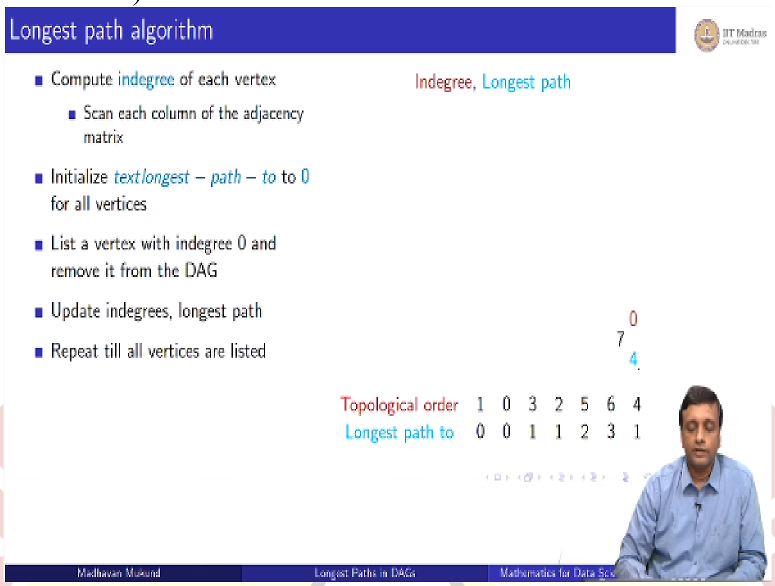
So, now the degree of 6 will reduce from 1 to 0, but the length of the longest path will go from 0 to 3 and notice as before that we are, as we are enumerating the vertices, we are also enumerating the longest path, because now I have enumerated it, its longest path is known. Next, we did 6, so when we remove 6, then this will go to 1, but this will go up to 4, so I remove 6.

(Refer Slide Time: 12:25)



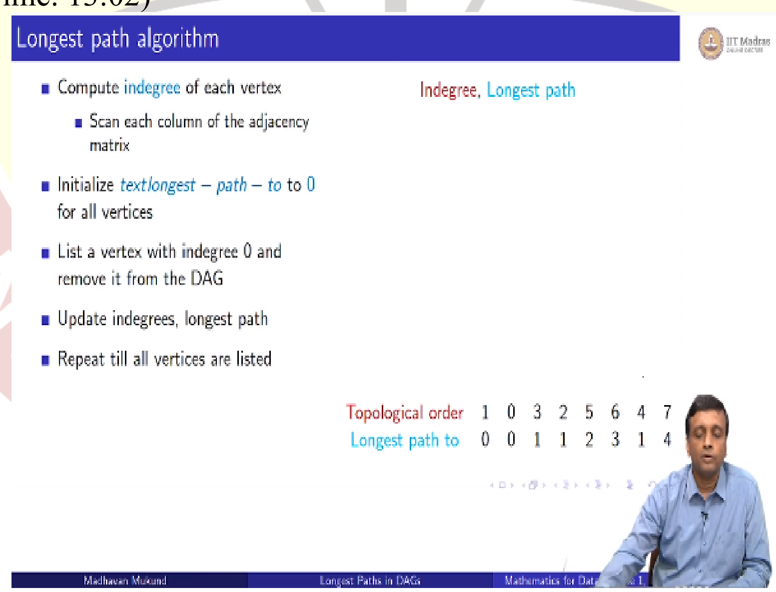
And now 7 jumps from longest path 2 to longest path 4 because I have discovered this path of length 3 coming up to 6, which is one of the incoming neighbors. But I cannot enumerate 7 yet, because its indegree is not yet 0.

(Refer Slide Time: 12:42)



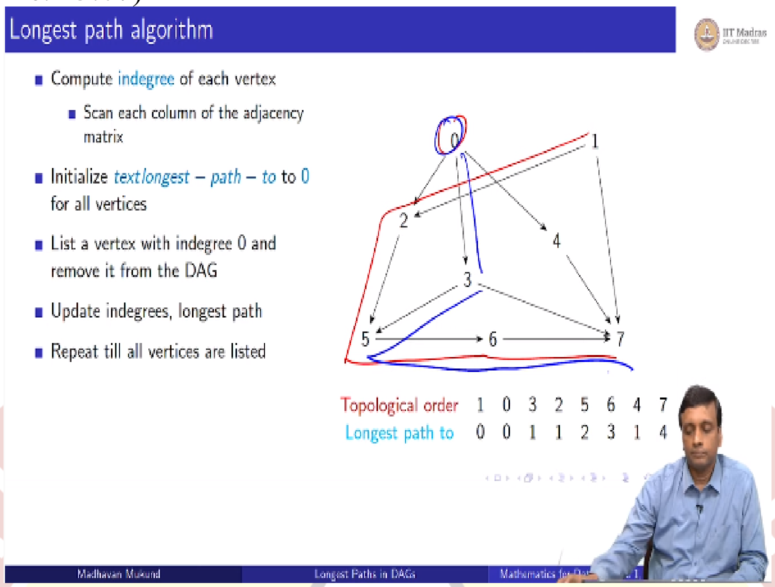
So finally, when I output 4, at this point, the indegree of 7 becomes 0 but there is no change to the length of the longest path, because the longest path does not come through 4 it has already come through 6, and 7 has already discovered that. It is just that I cannot finalize it until I actually reach the enumeration of 7.

(Refer Slide Time: 13:02)



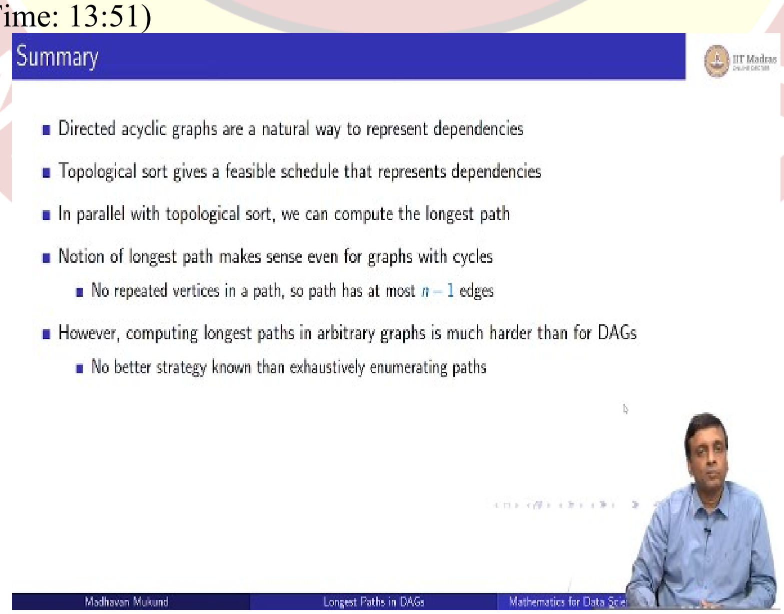
So finally, I list out 7 and now I have the longest path with every vertex listed below.

(Refer Slide Time: 13:07)



So, if you go back to the graph, you can verify for instance, that the longest path to 7 goes through say 1, 2, 3, so this is one longest path so 1, 2, 3, 4. So, basically the longest path is in terms of number of edges, at how many things I have to do before I come here, how many semesters I have to work before I do course number 7, there are multiple longest paths in general, so in this particular case, there is only one longest path to 7, I guess you can find but no, you can also find the longest path which goes this way. So, you can take a path, which goes from 0 to 3 to 5 to 6 to 7, so this is also a longest path of length. So, just the fact that you have a longest path, does not mean that is a unique one.

(Refer Slide Time: 13:51)



So, just to reiterate that directed acyclic graphs are a natural way to represent dependencies. So, we saw before the topological sort is how to get a feasible schedule, right, how to extract a sequence in which I can do these tasks, such that all dependencies are satisfied before I come to a task.

But now, we also saw this problem of how to compute the shortest duration that I need if I am allowed to do tasks in parallel. So, that is this longest path problem and we said that the longest path can be computed in an overlapping way with the topological sort. Because once we process of all the dependent vertices of a vertex, then we have enough information to process that vertex itself.

So, in the same sequence that we enumerated, we can also associate with each vertex, its longest path and incrementally build this up in parallel, so we do not have to take any extra work, while we are doing topological sort we can also compute longest path. Now, you might ask whether longest path makes sense in a graph with cycles because when I have cycles, I can go round and round.

I can go to a vertex and then come back to this place and go ahead but if you remember we made a distinction between a path and a walk. So, we said a path cannot repeat a vertex, a walk can repeat a vertex. So, going around the cycle and then proceeding technically is a walk not a path. So, if you use this literal interpretation of path, so we can have longest paths in cyclic graphs also, it makes sense, what is the longest sequence of edges I can travel before I repeat a vertex? This could be at most n - 1 but is it n - 1? So, that is a question. So, you could ask the same question about longest paths in a graph which has cycles.

So, in a directed a cyclic graph, we came up with a reasonably efficient algorithm which processed each vertex only once, and then depending on how we represented it as an adjacency list or a adjacency matrix, we had to do some scanning of that to find out the incoming and outgoing edges. But beyond that, it is a very reasonable algorithm, it takes time proportional to n or n + m, I mean n squared or n + m depending on how you are doing it.

What happens in this general case, well, in this general case, actually turns out to be surprisingly hard. So, if you have cycles in your graph, you can definitely define the notion of a longest path by looking at only paths in which vertices do not repeat, right. So, you know, that is going to be n - 1, but to know whether it is n - 1 or not is surprisingly hard.

So, actually there is no known efficient way to do this. Of course, you can do it by brute force, by examining every possible path in the graph and counting its length and then taking the maximum and essentially what we do not know right now is there any better way than that to do it, right. So, there is a huge gap between the longest path problem and DAGs, in the longest path problem and directed graphs which are not DAGs.

So, in directed graphs which are not DAGs, the longest path problem is very difficult computationally compared to the relatively simple problem that we have when we have DAGs.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

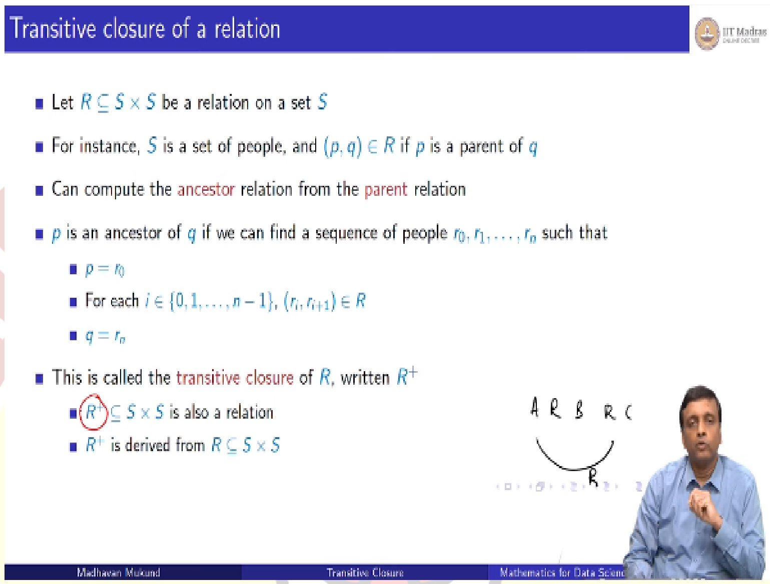
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor Madhavan Mukund
Indian Institute of Technology, Madras
Lecture 70
Transitive Closure

(Refer Slide Time: 0:18)



One of our original motivations for looking at graphs was to visualize relations. So, let us go back to relations. So, supposing we have a relation R on a set S. So, relation R on a set S remember is a subset of the Cartesian project S cross S, so S cross S is all pairs S1, S2 taken from S and some subset of these form a relation.

So, concretely, for instance, supposing we have a set of people, maybe a family, and then we want to find out the family tree in some sense, then we might represent as a relation when two people are related as parent and child. So, we can say that p, q belongs to R so, R is the parent relation, whenever p is the parent of q, so, p, q belongs to the relation R, if p is a parent of q.

Now, given this parent relation a very natural question is, what is the grandparent relation, what is the great grandparent relation and so on. So, in general what is the ancestor relation. So, we have is so and so is p an ancestor of q, is in the family tree is q a descendant of p? So, how would we do this? Well, to find out whether q is a descendant of p or p is an ancestor of q, we have to trace some sequence of relationship, we have to find a child of p and for that child, we have to find child of that child and so on or we have to find a parent of p.

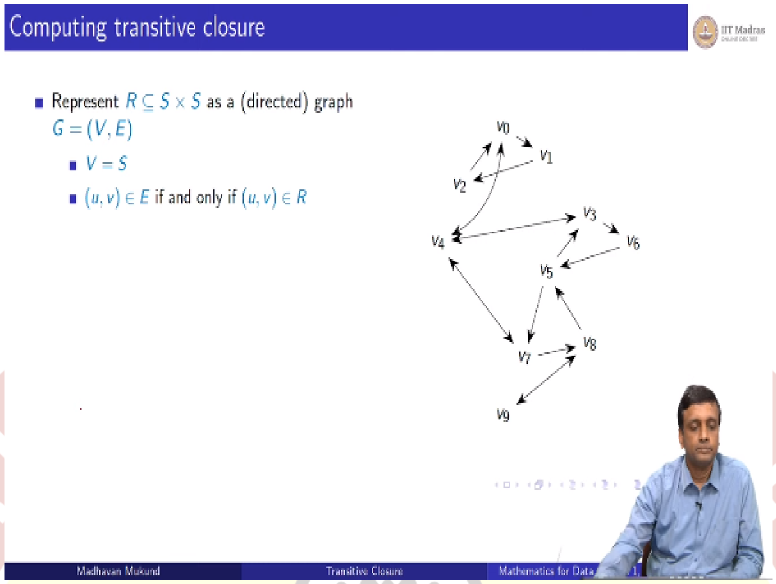
So, in this case, we are looking for ancestor so, p to q is a parent child relationship. So, we start with some p let us call it R_0, we have to find the sequence R_0, R_1, R_n we have to find the sequence of people says that R_0 is a parent of R_1, R_1 is a parent of R_2, R_2 is a parent of R_3 and so on, R_n-1 is a parent of Rn, so this is an ancestor sequence. So, R_0 is an ancestor of everybody to the right in sequence and particular if R_0 is P and R_n is qr, these are our two people we are interested in finding out, then we have established that p is an ancestor of q.

So, this is a new relation. So, this relates pairs of people who are connected by a sequence of parent relations. So, this has a name, this is called the transitive closure. So, transitivity says that if A is related to B and B is related to C, then A must be also related to C, this is the definition of a transitive relation.

So, parent is not a transitive relation, so what would happen if we made parent a transitive relation, if we forced parent to be a transitive relation, so, we said that whenever somebody is related to somebody and somebody is related to the other person, first person is related to third person, then that is the ancestor relation.

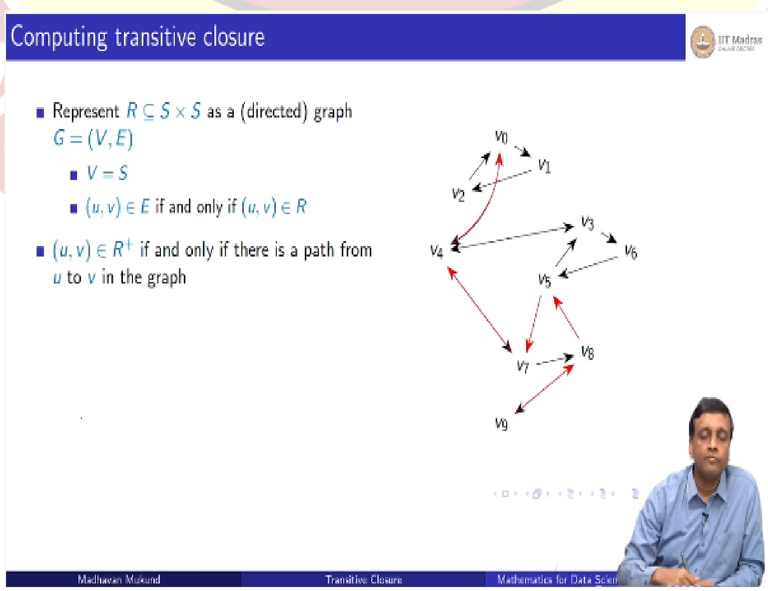
So, this is what happens when you close, we make the parent relation closed under transitivity, we force transitivity onto it, we compute what is called a transitive closure, okay, this is what we have. So, we normally denote this by R^+, R^+ means that we have to apply R one or more times to go from p to q. So this is a transitive closure of R.

(Refer Slide Time: 3:32)



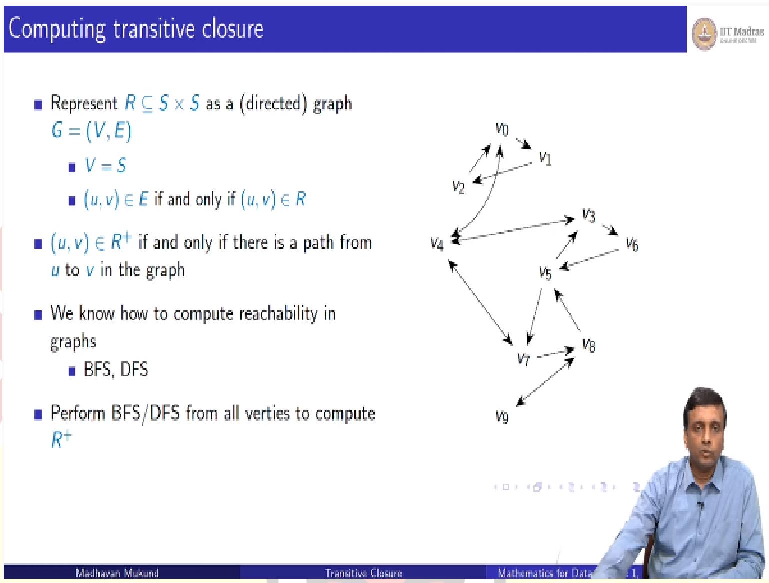
So, our question is, of course, how are we going to calculate this transitive closure in a systematic way? So, recall that we can represent any such relation as a directed graph in general. So, we represent, the vertices represent the set S, V=S and each edge represents a pair in the relation. So, if u,V is a member of the relation R, then we put an edge from u to V and this is the only case in which you put an edge from u to V. So, the edges are exactly the pairs which are related by R.

(Refer Slide Time: 4:07)



So, what we have defined as R^+ can be calculated in the graph as being a path, right, so we can say in this case that V9 is related by R^+ to V_0, because there is a path from V_9 to V_0.

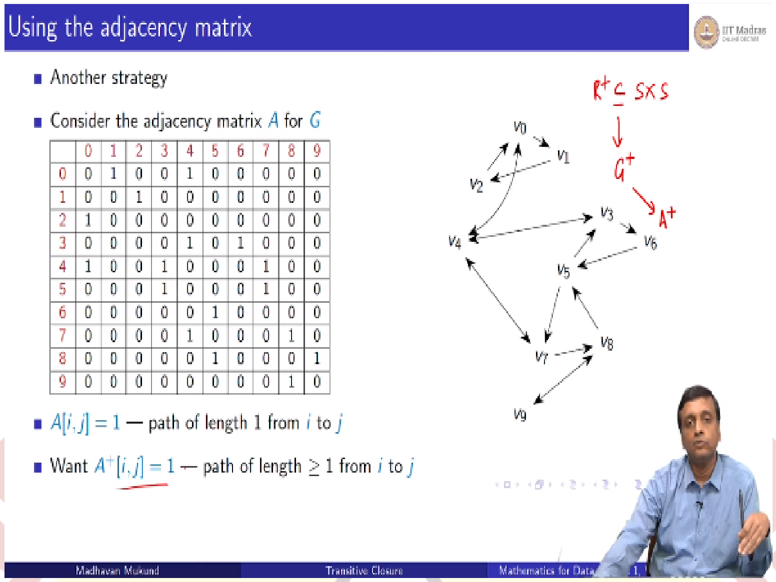
(Refer Slide Time: 4:25)



So, how do we find these paths? Well, essentially, we want to find all such pairs for which V_9 is related and we said that if you do not focus on a single path, we can just calculate reachability, what all is reachable from every vertex and this we can do using breadth first search and depth first search. So, in the R^+ case, we are interested in finding the reachability for every i in every j, right, we want to know for every i and every j, whether it falls into R^+ or not. So, we have to compute this reachability for every starting point.

So, one way to do this is to just perform this BFS, DFS starting systematically we started the 0th vertex, you perform BFS, then you know everything which is of the form 0, i, then you start at 1 and you perform, notice that this is a directed graph. So, you have to do it because if there is a, if 0 can reach j, it is not obvious or is not required that j can read 0 in a directed graph. So, you have to perform this from every vertex to find out what all pairs fall into R^+. So, this is one way to do this using what we already know.

(Refer Slide Time: 5:25)



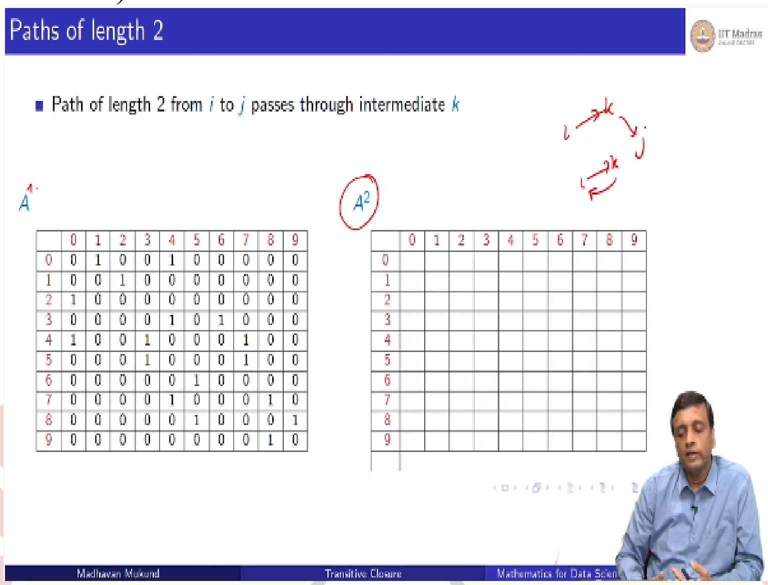
So, here is another strategy, so the other strategy is to look at the adjacency matrix. So, in this adjacency matrix, we put a 1 if there is an edge from i to j. But another way of thinking of an edge is that there is a path of length 1 from i to j. A single edge is a path of length 1 and what we want is to create an similar adjacency matrix for this expanded list.

So, remember that R^+ is also relation, so R^+ we said is also a relation, so this also correspond to some new graph G plus, because every relation we can draw a graph, just add the same set of vertices in this case, but now put an edge if there is an R+ relation value, relationship between i and j.

So, this G^+ will correspond to an adjacency matrix A plus, so, that is this A+ here, we want an matrix A+ whose ij entry is 1 , if there is a path of length 1 or more from i to j, so, if there is an edge directly, so an ancestor, a parent is an ancestor, parents parent is an ancestor path of length 2, parents, parents parent is an ancestor, so that is length of 3 and so on.

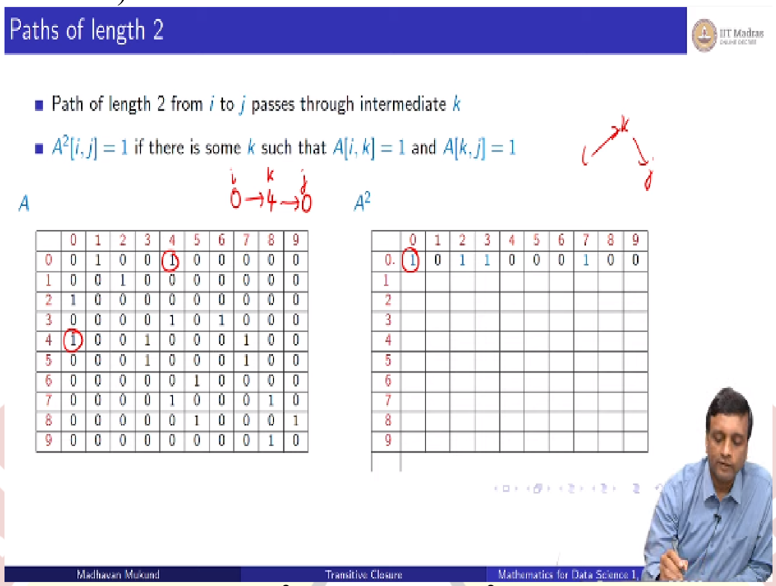
So, therefore we have any path of length 1 or more between i and j that will be an edge in the R^+ relation and in the G^+ graph, and we want to compute A^+ from A, so that is our goal. So, we know one way to do it, which is to go to breadth first search, and then fill in A+ from breadth first search by doing breadth first search from every vertex. But the question is, can we do it directly using just A?

(Refer Slide Time: 7:08)



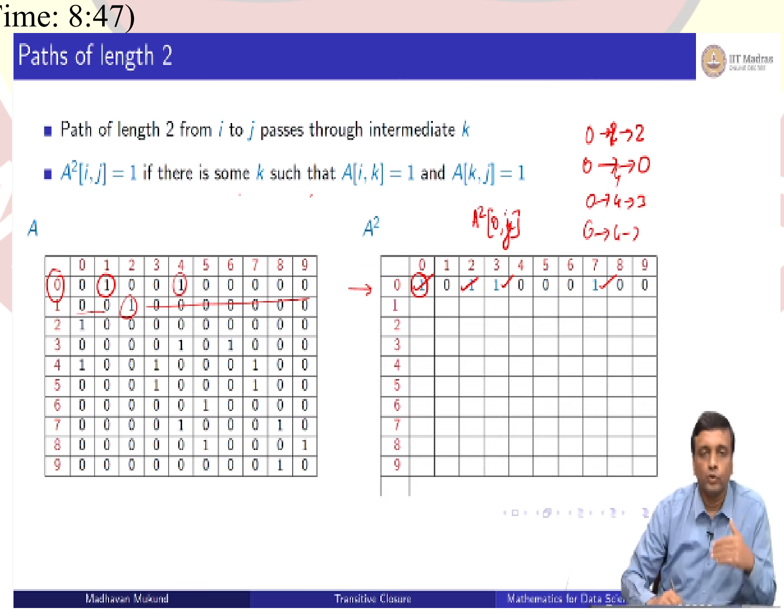
So, we have now on the left, we have A and A denotes paths of length 1. So, as the first step let us try to compute paths of length two, right. So, paths of length 2 go from i to some k, to some j. So, this is a path of length 2, we could also be back to itself, path of length 1 cannot go but we could have a path of length 2, which goes from i to i. So, this is the matrix that we want to call now A^2 for instance, A^2 so A, which we can think of A^1 if you want, represents paths of length 1, A^2 represents paths of length 2.

(Refer Slide Time: 7:55)



So, how do we fill in the entries of A^2 from A? So, A^2 ij is 1 if there is some k such that A_{ik} is 1. So, there is i to k there is a path of length 1 and k to j there is a path of length 1. So, I look at this entry, why is there an A^2 entry from 0 to 0? Well, I claim that there is a 0, 1 entry, not a 0, 1 entry sorry, there is a 0, 4 entry, so I have a path from 0 to 4, because of length 1, and then I have a 4, 0 entry. So, if I choose k to be 4 and i to be 0 and j to be 0, then I get A^2 ij as 1 because for k =4 I have this.

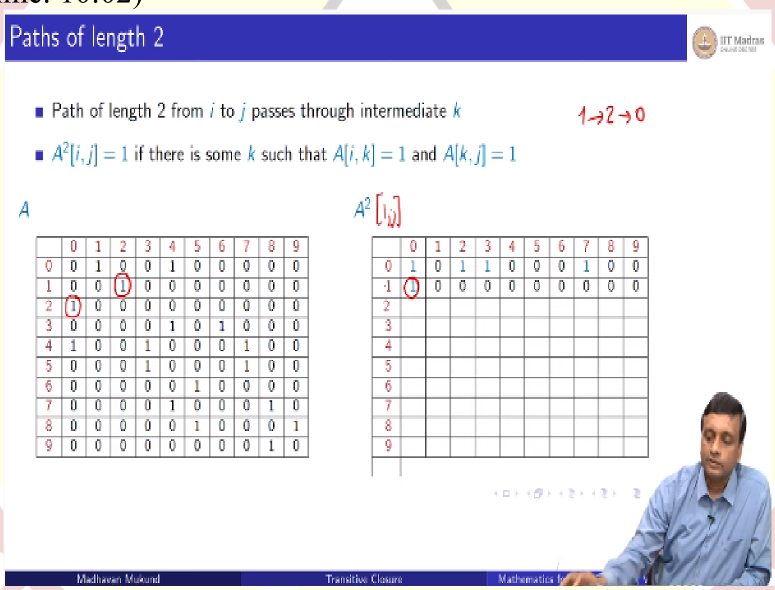
(Refer Slide Time: 8:47)



So, in general, what I have to do is I have to look at, so if I do it systematically, I will start, I want to fill in the row for 0. So, I look at all this, I look at the first entry, I can go, where can I go via 1? So, if I say 0 to 1, then I look at the 1th at the entry, so 1 can go to 2, so therefore I have an entry 0, 2. So 0 to 1 to 2, then 0 can go to 4, so I should have. Sorry, yes, so 0 to 1 cannot go anywhere else, so 1 has only one outgoing edge so through 1 I cannot go anywhere else. So, 0 can go to 4, so I now look at the outgoing edges from 4, so 0 to 4 to 0, right, so that is how I get it.

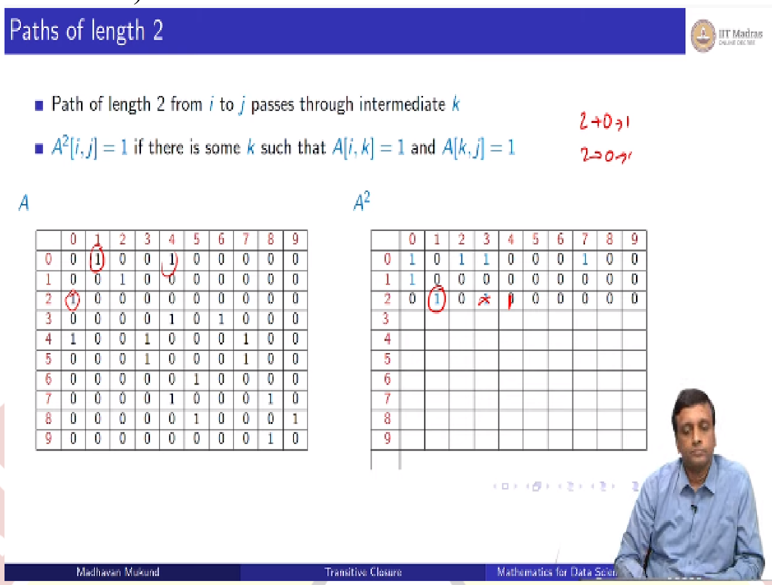
Then I have 4 goes to 3, so I have 0 to 4 to 3, so I get this entry and finally I have 4 goes to 7 so I have 0 to 4 to 7, so I get this entry. So, in this way, I can compute all the entries of the form A^2 0,k, by finding the intermediate values by looking for something that 0 goes, 0 to k and then from k to j.

(Refer Slide Time: 10:02)



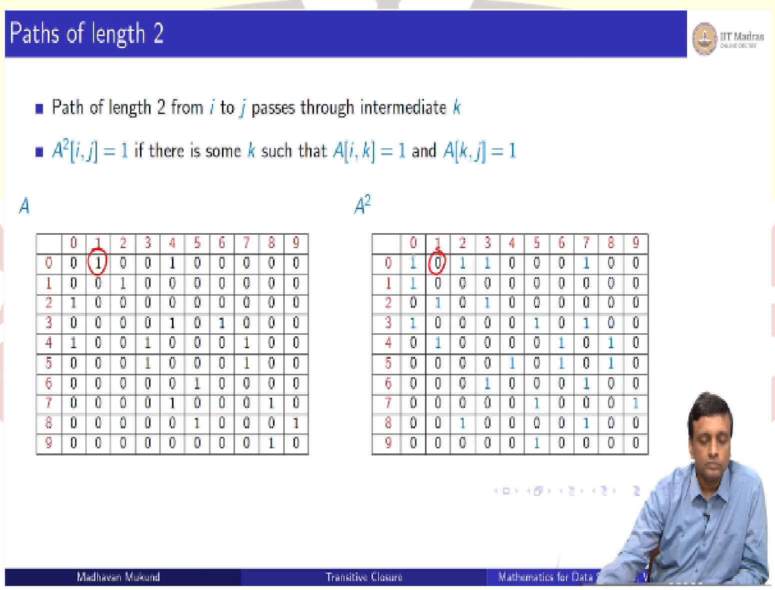
So, I can do the same thing for 1 now. So, I want to find all the entries of the form 1,j, so I look at the, so 1 has only one outgoing thing, going to 2, and 2 has only one outgoing thing going to 0, so the only new thing I discovered is 1 to 2 to 0, right, and that is the same.

(Refer Slide Time: 10:25)



Now, I will do the same thing for 2, so 2 has only one outgoing thing to 0. But 0 has outgoing edges to 1 and to 4. So, I get 2 to 0 to 1 and 2 to 0 to 4. So, I get 1, so this should not be here, this will be here, I get 2 to 1 and 2 to 4.

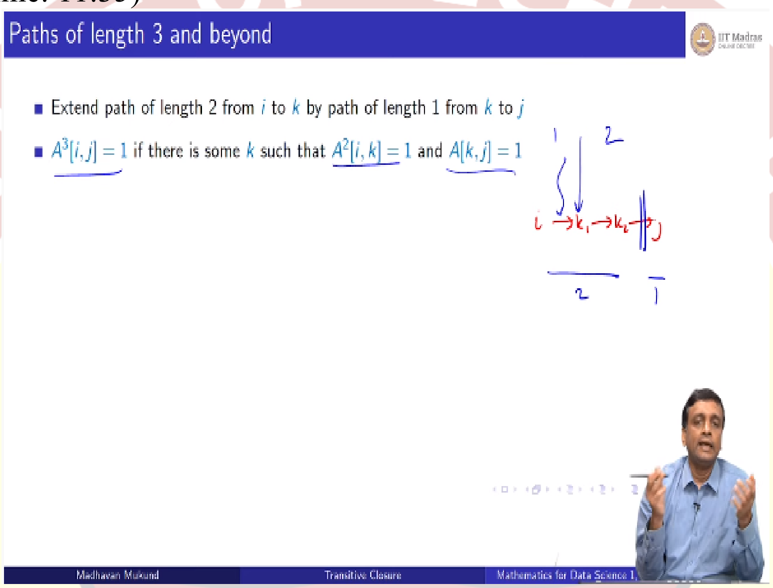
(Refer Slide Time: 10:47)



So, in this way, I can do this for all the entries, I can for every row, I can take the outgoing edges 3 to k and the outgoing edges k to j, and add an entry 3 to j, right. So, by scanning these two rows in this matrix I can compute A^2 matrix. So, A^2 represents all the paths of length 2.

So, notice that the paths of length 2 do not subsume the paths of length 1, right, so for instance, we had a path of length 1 from 0 to 1, but we have no path of length 2 from 0 to 1. So, these are paths strictly of length 2, they are not of length 0 or not of length 1, so A, the first A has edges, the second one has paths of length 2, so I can go in length 2 from 0 to 0. But the fact that I can go from 0 to 1 in length 1 does not mean I can go from 0 to 1 length two. So that is paths of length 2.

(Refer Slide Time: 11:35)

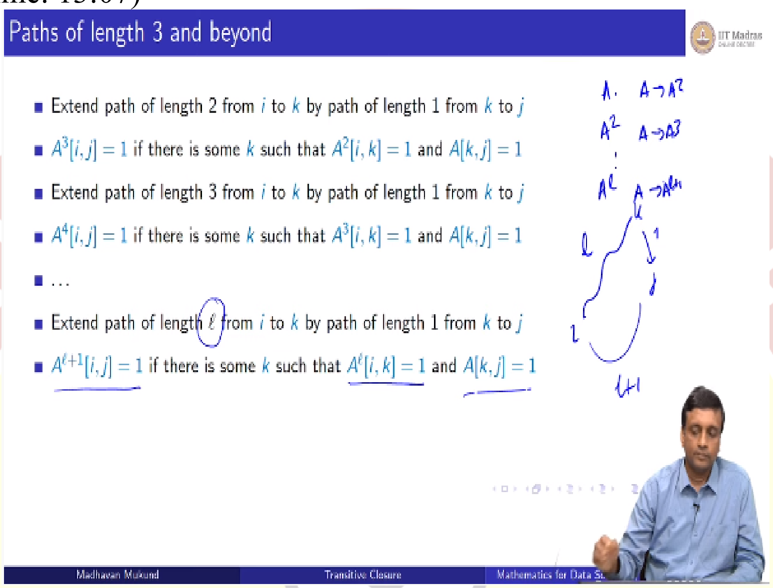


So now, how do we go to path of length 3? Well, if I have a path of length 3, it must be of this form, i, k_1, k_2 to j, right, so there must be some two things in between. So, I can split it up whichever way I want, I can either take this point and say that I have a path of length 2, followed by a path of length 1, or if I want, I could do it the other way, which is I could, I could split it up here and say I have a path at this point.

Sorry, and say I have a path of length 1 followed by path of length 2, right. So basically, a path of length 3 can be decomposed as two+ one or one+ two. And I already have explicit matrices for 2 and 1, I know all the paths of length 1 are represented in A, I know all the paths of length 2 are represented in A^2.

So, I can say now A^3, i,j is 1 if there is some k for instance, where there is a path of length 2 from A to k and there is a path of length 1 from k to j. So, now earlier, I looked at A and within A I look for two entries. Now, I look at an entry in A^2 and I look for an entry in A and I try to match them up, I tried to find a k such that from i to k have an entry 1 in the A^2 matrix, and from k to j, I have an entry in the A matrix. So this gives me A^3, so you can do the same calculation as before and you can come up with a new matrix. So I started with A, I did one pass over A and I got A^2. Now using A and A^2, I can get A^3.

(Refer Slide Time: 13:07)

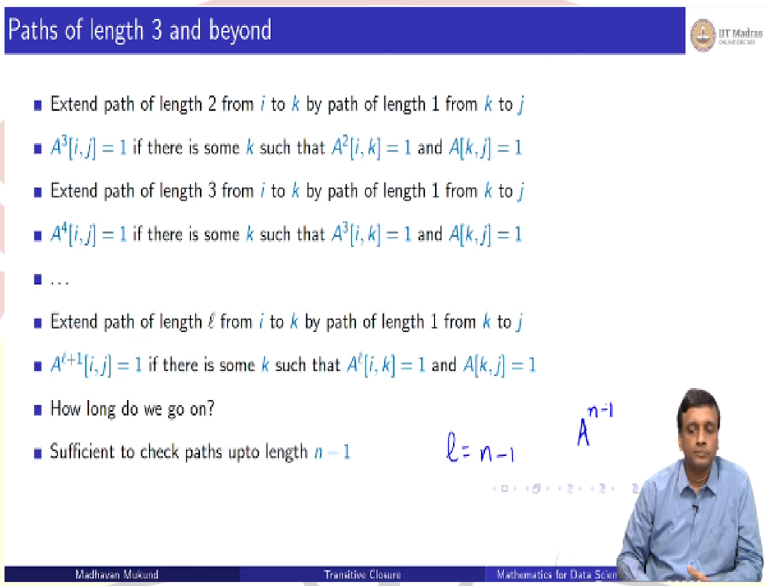


So, now I can go from 3 to 4 in the same way, if I have a path of length 4, then it can be decomposed as a path of length 3 followed by 1 edge. So, I can take the entries in A^3 and combine them with entries in A, so I can look for a k such that A^3 i, k is 1 and A k, j is 1. So, this now gives me A^4 i, j, which will be 1 provided there is a path of length 3 from some i to some intermediate k, followed by a path of length 1.

Now, you could also do this as 1+ 3, you can do it as 2+ 2, but let us just follow this general rule, where we break it up into one less or + one. So, in general, if we keep going, right, so if I want to, I already know paths of length l, and I want to extend it to l+ 1, then I will say that A_l+ 1 i, j is 1. If I already know that there is a k for which there is a path of length l and I can extend it by 1 edge, a path of length 1 from k to j. So, I can go from i following some l steps to k and then I can go from k to k in one step. So, then I can go therefore from here to here in l+ 1 steps.

So, we just do the same thing again and again. The first time we are doing A combined with A, second time we are doing A^2 combined with A, in general we do Al combined with A and each time I will go from here I get A^2, from here I get A^3, from here I get A^l+ 1 and so on. So, I can keep on building this matrix, which captures longer and longer paths.

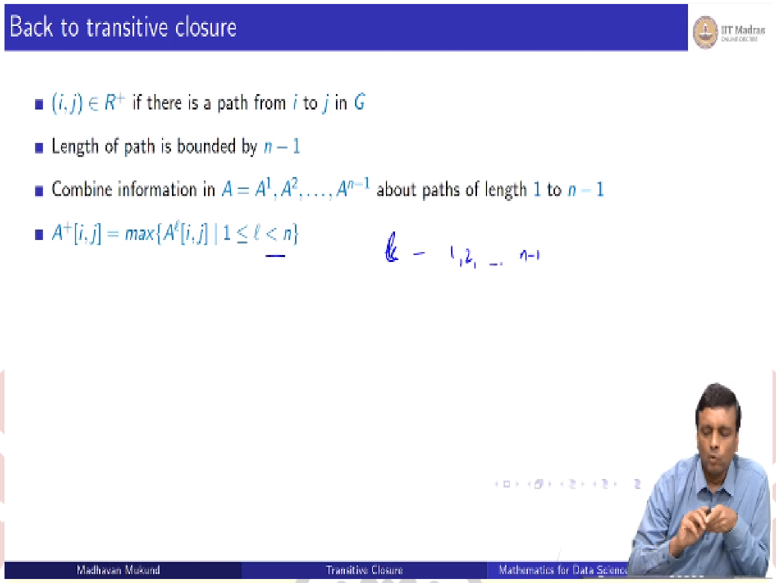
(Refer Slide Time: 14:44)



So, now where do we stop? How long do we go on? Well, here we know that if there is a path at all, then that path cannot have more than n-1 edges, because once I have traversed n-1 edges, I have seen n-1 different vertices other than the starting point and therefore, anything beyond that must repeat a vertex, so there must have been a shorter path.

So, therefore, if there is a path at all from i to j, it cannot have length more than $n - 1$, so I can stop with. So, once I have computed A to the $n - 1$, right, I have got everything of interest, I have got all paths of length 1, 2, 3, 4, up to n-1. And any path which is longer than n-1 cannot be new. I mean, since it cannot contribute any new information to me about whether or not i and j are connected by the relation or not.

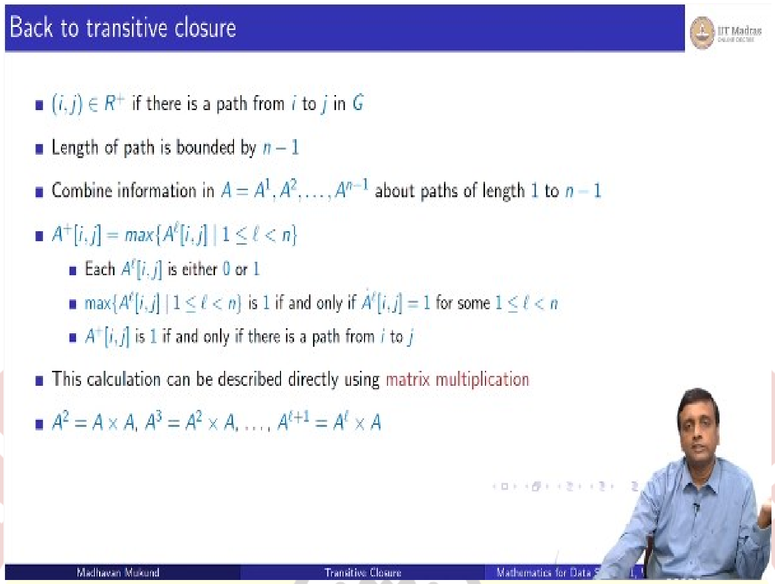
(Refer Slide Time: 15:37)



So, remember that the reason we are doing this is for the transitive closure. So, we said that i, j is in the transitive closure R^+, if there is a path from i to j in the corresponding graph for R. And we have observed many times that the length of this path is at most $n-1$. So, therefore, we can combine all this information, right, so, we have the original A which is the same as A^1, right, path of length 1, then from that we computed as A^2, then A^3, A^4 and so on and up to the A^{n-1}.

So, I have this n-1 matrices, which gives me all information about paths from length 1 to length n-1. So, what do I want to do? I want to say that there is an edge from i to j in the R^+ relation. If there is an edge somewhere in one of these, right, and I am going to write it in this complicated way, I am going to say it is a maximum of the i, j entry in all the matrices from k =1 to k =n-1, notice that this is strictly less than it. So, k goes from 1, 2 up to $n-1$ because it starts at 1 l, A to the l sorry, l goes from one to n-1, right.

(Refer Slide Time: 16:50)



So, this is, so, what does this mean? So, notice that each and each entry is 0 or 1. So, every entry of these, these are all 0, 1 matrices, either there is a path of that length or there is not a path of that length. So now, when I am taking this max, it is basically checking if all of them are 0, that is there is no path of length 1, there is no path of length 2, there is no path of length 3 and so on, the max is also going to be 0.

So, there will be an A+ entry which is 0, there is no path, but if there is a 1 anywhere, right, in any one of those l positions from 1 to n-1, if any one of them is 1, then the max will become 1, if there are many paths, there are path of length 3 and 7, it will still be 1 because max of 1 and 1 will remain 1.

So, by taking max we are just recording is there at least 1, 1 in that sequence or not? Sequence meaning across all these matrices in the ij th position is there at least one of these matrices which has position value 1 at i, j. If so, the max will give me 1, if all of them are 0 max will give me 0.

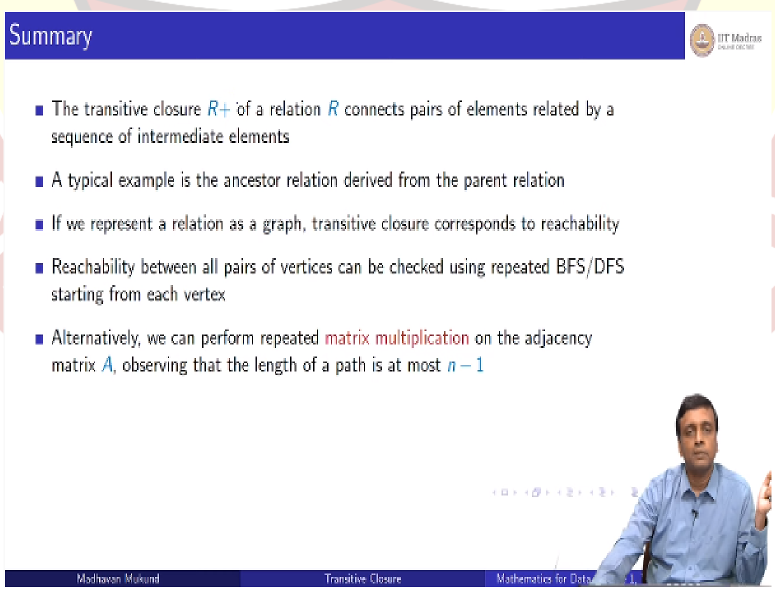
So, in that sense, this A+ entry captures the fact that there exists some length path between 1 and n-1, between i and j and if it is 0, it means there is no such path, right, and we know that if there is no path of length n-1, there cannot be a longer path because if there is a path it must have at most length n-1, anything longer than that will be looping and will be redundant.

So, therefore, A+ i,j is 1, if and only if there is a path from i to j and in particular, this path must always be bounded by length n-1. So, what we have done, we can actually reformulate it in a way that is called matrix multiplication, which we will not do right now.

But it is important to know that this, what we did is a very tedious calculation rows and columns and all that is actually a very standard operation on matrices. So in this form, we can write it as a sort of multiplication of matrices. It is not exactly what I have written here, but for the purpose of this lecture, this is fine.

So, you can actually believe that this operation, the reason we are doing this with matrices is that this operation on matrices is actually a standard mathematical operations on matrices. So, though we have done this column and row chasing explicitly, saying we look for a k here, we look for a k there, actually, you can actually do it directly as a matrix operation. So therefore, it is a very standard operation is not something new.

(Refer Slide Time: 19:09)



So, to summarize, the transitive closure tells us so, this would be an R^+ on top, the transitive closure tells us whether there is a sequence of intermediate elements which connect two elements, right. So, I have to start at p and go through multiple R edges to reach q, an example of this was our ancestor relation. So, the ancestor from the parent, so a sequence of parent edges generates the ancestor edge.

So clearly, since we visualize relations as graphs, this corresponds to a path, right, and in general, these are directed edges because these relations are not assumed to be symmetric, like the parent relation is certainly not symmetric, if A is parent of B clearly B is not a parent of A, right. So therefore, in general, you follow a path in a directed sense, and this is just a reachability question in graphs.

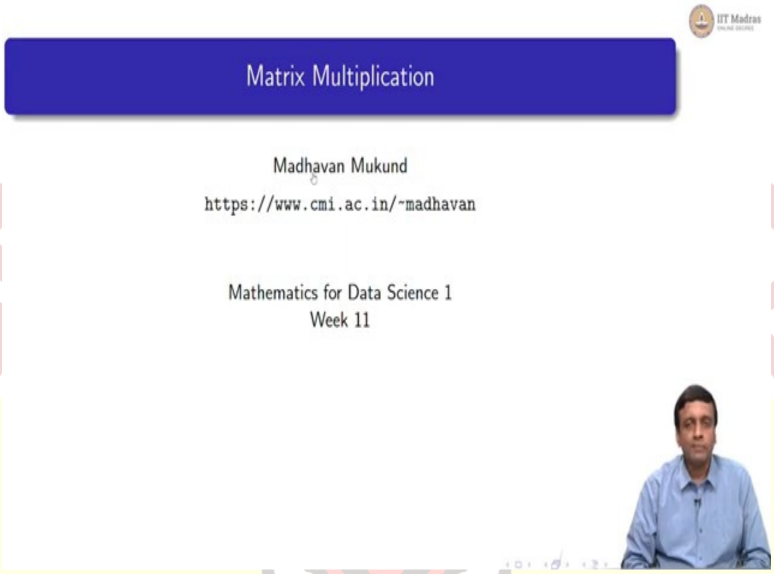
And we know that we can do this by repeatedly doing BFS and DFS from every starting point, but what we have seen in this lecture is that alternatively, we can take the adjacency matrix and do a form of matrix multiplication, we can do a form of matrix multiplication to go from A to A^2 to A^3 and so on and stop with A to the n-1 because A to the n-1 records paths of length n-1 which is the maximum length path, which is useful to us to find out whether two edges, two nodes are connected and once we have got this we can take a look for a 1 in any one of these n-1 matrices and if so, declare A+ ij to be 1.

IIT Madras

ONLINE DEGREE

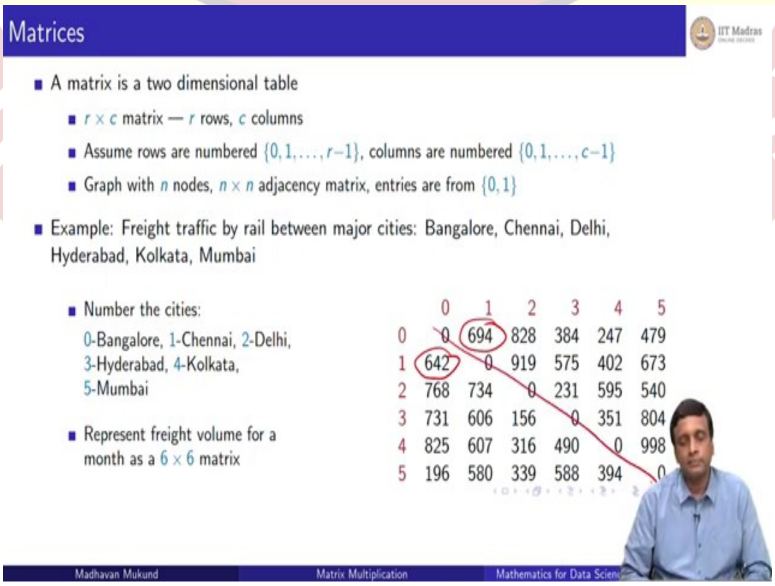
Mathematics for Data Science 1
Professor. Madhavan Mukund
Department of Computer Science
Indian Institute of Technology, Madras
Lecture No. 71
Matrix Multiplication

(Refer Slide Time: 00:14)



We have used matrices to represent graphs. So, we have this adjacency matrix. And when we looked at the transitive closure problem, we suggested that it might be useful to describe the transitive closure computation, using some operations directly on matrices. So, let us look at matrix operations. In particular, we want to lead up to this notion of matrix multiplication.

(Refer Slide Time: 00:36)



So, a matrix in general is a 2-dimensional table of values, usually numbers, and it has a certain number of rows, because it is 2 dimensions. So, it has rows and it has columns. So, usually, it has some r rows and some c columns. r and c could be different numbers. And we write this as $r \times c$, or we call it an $r \times c$ matrix. So, r is the number of rows c is the number of columns. And as usual, we will number the rows starting from 0 and the column starting from 0. So, the rows are 0,1,2, up to r-1, and the columns are 0,1,2, up to c-1.

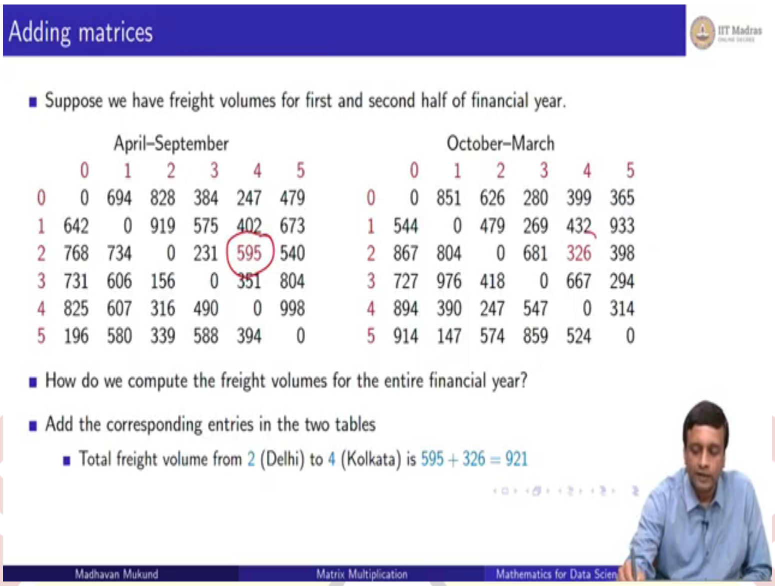
So, in the example that we have been looking at the adjacency matrix, r=c=n, which is the number of vertices in our graph. So, we have 1 row and 1 column for every vertex in the graph. So, we number our columns and rows is 0 to n-1. In this particular case, the entries in our matrix are 0 and 1, but in general, they could be any numbers, even other values in general, but we will look mainly at matrices where the entries are numbers.

So, for a completely different example, let us consider that we want to record some information about freight traffic by rail between some cities. So, let us take say the 6 biggest cities in India, Bangalore, Chennai, Delhi, Hyderabad, Kolkata, and Mumbai. And suppose we have information about the volume of traffic, say in terms of 1000s of tonnes of millions of tonnes, whatever is the appropriate metric for this. So, we have some information. So, we want to represent this metric.

So, the first thing is that every city will correspond to an entry in the matrix in a row and the column. So, because we are normally numbering our rows and columns from 0 to n-1, or 0 to r-1, and 0 to c-1 in this case, again, notice that it is going to be a square matrix. So, it is going to have 6 rows and 6 columns. Let us just number them in alphabetical order. So, 0 is Bangalore, 1 is Chennai, 2 is Delhi, 3 is Hyderabad, 4 is Kolkata and 5 is Mumbai, it does not really matter in what order we put it.

So, this could be the matrix. So, this is not a symmetric matrix, as they say. So, there is a certain amount of traffic say, from Bangalore to Chennai, but there is a different amount of traffic from Chennai to Bangalore, because in general, there is no reason why the amount of freight traffic should be the same in both directions. Notice that along the diagonal, there is no freight traffic recorded from a city to itself, because that does not make any sense for us. So, this is another way of using a matrix where we are representing some information about freight between pairs of cities.

(Refer Slide Time: 03:02)

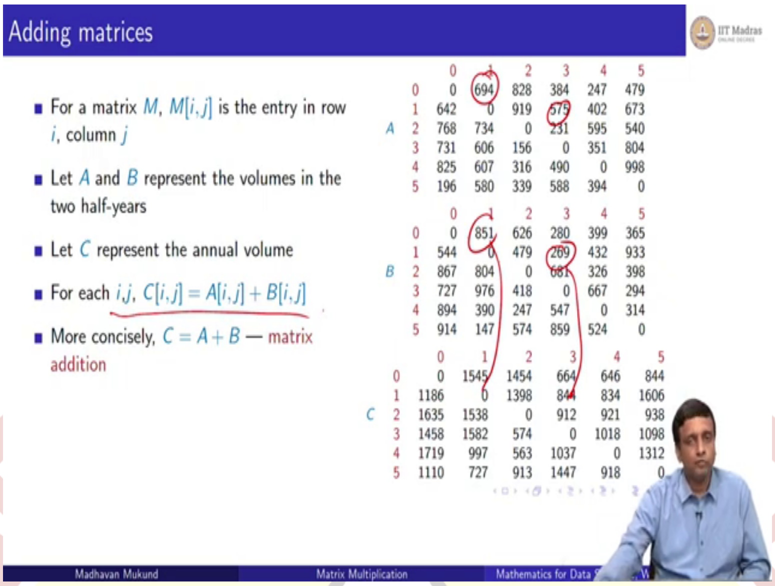


So, now, suppose we have this information, and it is counted over a certain period of time. So, in India, you may know that the financial year starts in April. So, April 1’st is the beginning of the financial year and march thirty first is the end. So, maybe for the first half of the financial year, from April 1 to September thirtieth, we have a certain number of amounts of information about this fr8 volumes, and that is the matrix on the left. So, this is the first 1, so, this is the first half year and then we have a corresponding thing for the second half year. So, the numbers are different.

So, for instance, here, there is only 247 whereas here in the same entry from Bangalore to whatever was 4, you have Kolkata I think, there you have 399. So, now, a natural thing would be to compute the freight volume for the full year given this information about the freight volume in the first half of the year, and the fr8 volume in the second half of the year. And quite clearly, the freight volume for any pair of cities is going to be the sum of the 2.

So, if you want to know how much traffic there has been from Delhi, which is city 2 to Calcutta, which is city 4, you look up the entry for the first half of the year, it says okay, there were 595 million tonnes or whatever it is from Delhi to Calcutta in the first half. And then the second half, there are 326. So, clearly put together it is 595+326, which is 921. And this is how we would compute this information.

(Refer Slide Time: 04:22)



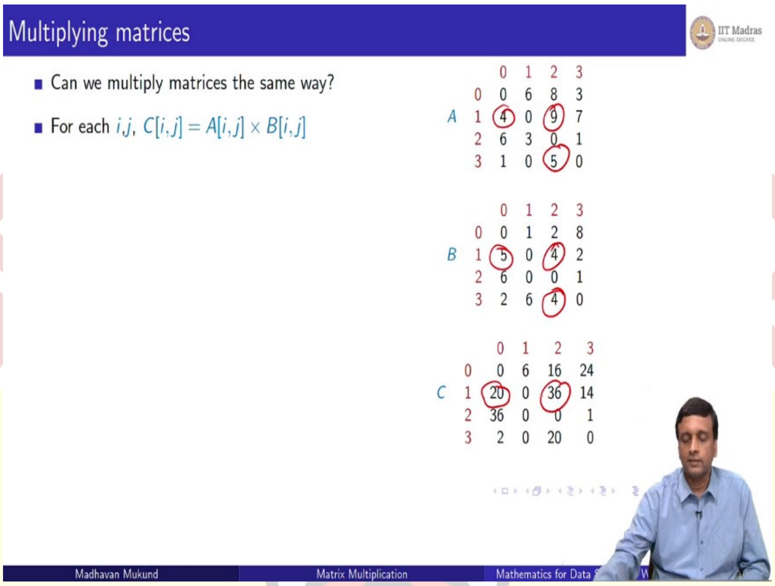
So, some notation. So, if we have a matrix, then remember our rows are numbered 0 to r-1 and our columns are 0 to c-1. So, we will look at the I'th row in the J'th column, and that will give us 1 entry in the matrix, where the row intersects the column and we will refer to that as M i comma j. So, M i comma j is the entry in the matrix at row i and column j.

So, now here is our information about the 2 half years again, presented on the right-hand side. So, these are the 2 matrices that were given. And we want to compute the aggregate volume for the whole year into a new matrix, call it C. So, I have used A for the first half year, B for the second half year, and now we want to compute a matrix C, which represents the total volume for the whole year. And as we said before, the ij'th entry, the entry in row I column, J of C is going to be the sum of the ij'th entries in A and B. It is quite obvious.

So, if we want to look at a pair of cities and find out the total traffic between the 2 cities, you add up what traffic was there in the first half of the year to what traffic was there in the second half of the year. So, when we do this, we are doing this for every element. So, we are taking this element adding this element and getting this element, we are taking this element, this element and getting this element. So, we are doing it in the same way for every element of the matrix. And so, we can just concisely say that we are adding these 2 matrices. So, this is called matrix addition.

So, when we have a matrix A and a matrix B, and they both have the same size, the same number of rows and the same number of columns, then for every position in A there is a corresponding position in B. If I look at row ij, in column J A, there is a row in column J and B because they have the same size. And So, I can now add these entries and put it into a third matrix of the same size. And I will just write C =A+B. So, C =a+B captures this fact that for every row and every column, I am adding the entries in A and B to put them into C. So, remember that we were not so interested in adding matrices but multiplying them. So, what should we do for multiplying matrices?

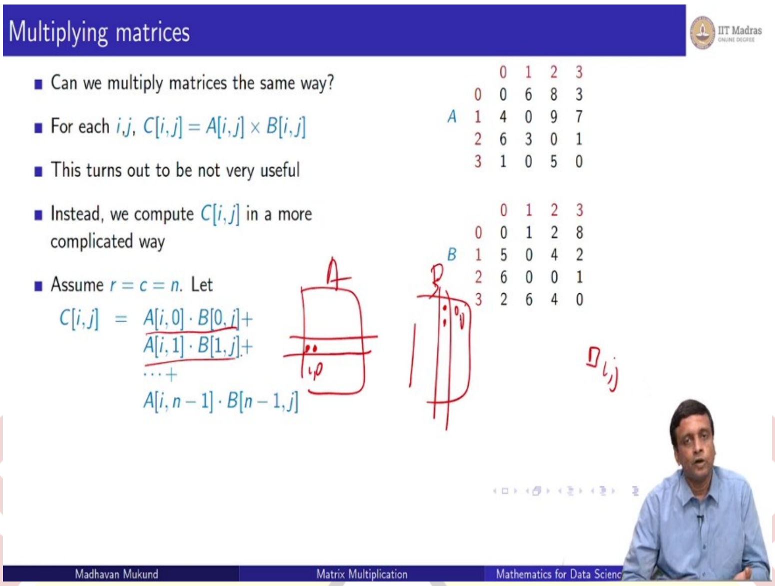
(Refer Slide Time: 06:25)



So, can we multiply matrices the same way? So, let us take simpler numbers. So, here are 2 matrices A and B. And suppose we do the same rule, the same rule would say instead of adding A_{ij}, into B_{ij} multiply A_{ij}, by B_{ij}. So, here, for instance, if I look at C 1, 0, it is A 1, 0, 4 ×B 1, 0, 5, 4 × 5 is 20. Similarly, if I look at, say for instance, A 3, 2, it is 5, B 3, 2 is 4. So, again, it is 20. If I look at this entry, it is 9 × 4 is 36. So, I have just multiplied entries, one at a time.

So, this could be a candidate for the operation of matrix multiplication is the obvious way in which we can take addition and replace addition by multiplication.

(Refer Slide Time: 7:14)

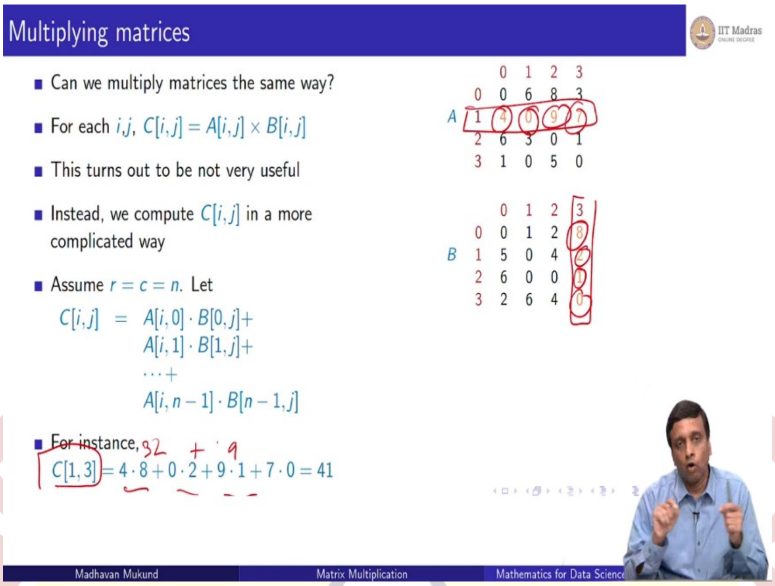


Now, it turns out that this particular way of manipulating matrices, there is nothing wrong with it, it is just that it does not turn out to be very useful in practice, addition made sense for us. Multiplication, like this usually does not give us any useful number. So, there is a different operation, which is given the name matrix multiplication, which is a little more complicated than this.

So, we want to compute the multiplication A × B, where A and B are matrices. So, again, we have to tell us a rule for computing the ij'th entry of the product. So, the product of A and B has, again, entries i comma j. So, what is the value that goes into C ij? So, here is the complicated rule. The complicated rule is that you look at row I. So, you look at row. So, this is my matrix A and my matrix B. So, if I want to compute the ij'th entry in C, I look at row i in A and I look at column j in B.

So, I look at the first entry here, what is the first entry here it is i comma 0. And look at the first entry there. What is the first entry there it is row 0, column j. So, I multiply those 2 numbers. Now I move to the right, and I look at the next number here. That is 1, i comma 1. And I look at the next number in that column, that is 1 comma j. And I multiply those 2 numbers. So, I keep walking down the row and A, A'th row at the same time I walked down the j'th column and in B, and every point I stop and I multiply the 2 numbers that I get, I take all these numbers and add them up. It seems really complicated, but this is how matrix multiplication works.

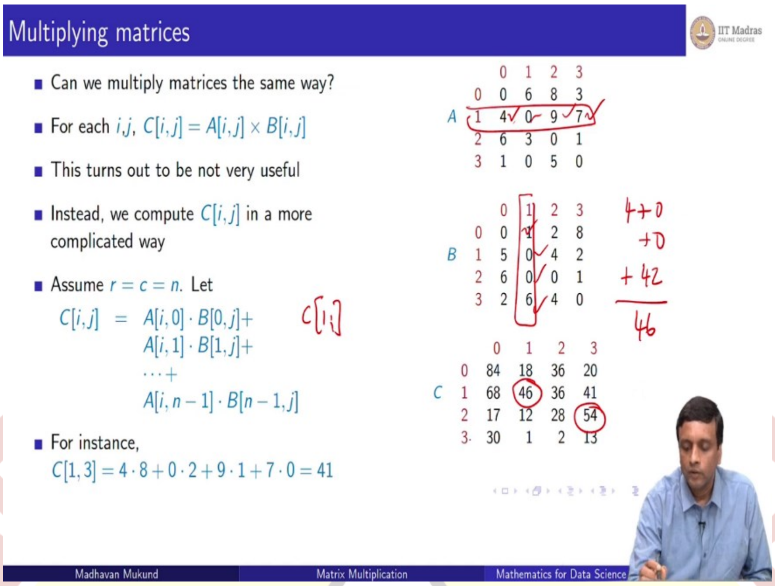
(Refer Slide Time: 8:56)



So, let us do an example. So, supposing I take these 2 matrices, and I want to find out, what is the final entry going to be for C 1, 3, that is row 1, column 3 of the final matrix, what is it going to be? Well, I have to look at row 1 in A, and I have to look at row 3 and B. And then I have to start one by one. So, first, I multiply 4 by 8. So, that is my first component, then I have 0, which is nice, 0 × 2 is 0, no problem, then I have 9 × 1 is 9. And then I have again, 7 × 0, another 0, so, I get 32. And I get 9, so, 32+9, so, I get 41.

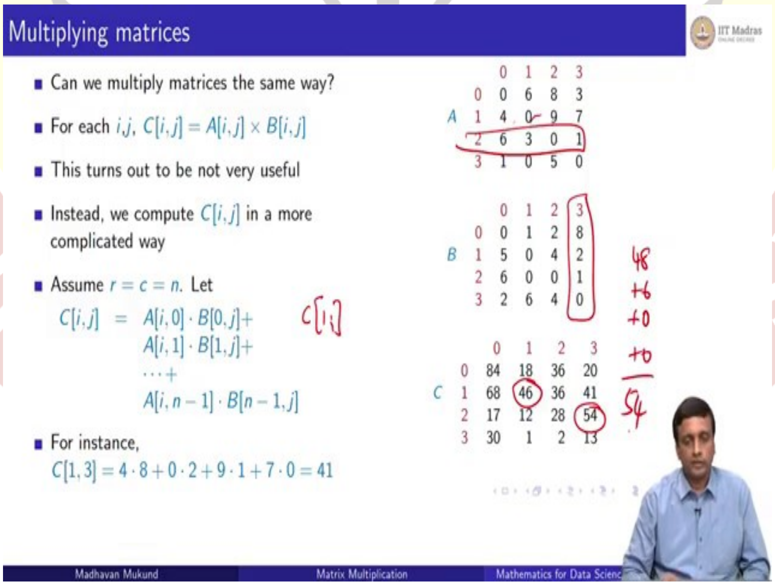
So, this is how I compute 1 entry. And this is just 1 entry. And I have to do this for every position. So, for every position in the output, I have to scan a row of the input and a column of the second matrix or row of the first matrix and add up those that many, I have to multiply each individual pair and add them all up. So, I have to do a lot of work to compute 1 entry.

(Refer Slide Time: 9:55)



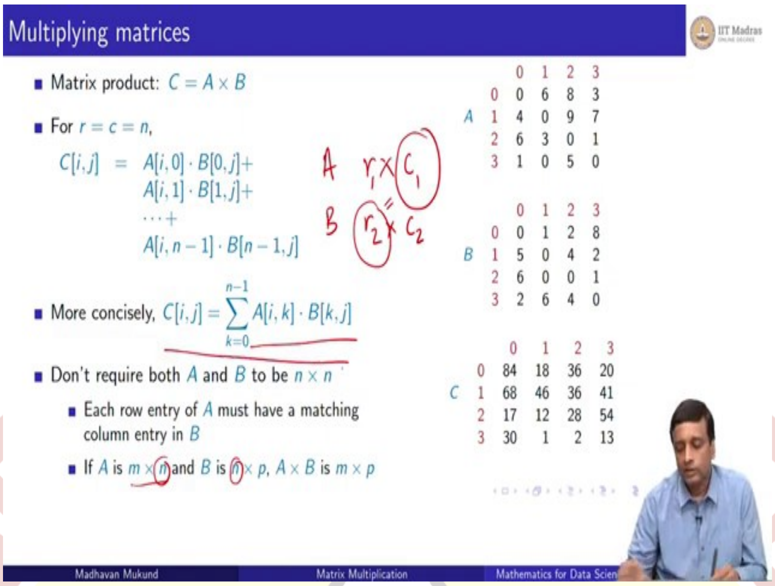
And if you do this, then you get this matrix, let us look at some other entries. And see, how did we get this 46? How did we get C 1 1? Well, So, C 1, 1 means I have to look at row 1. And I have to look at column 1. So, I take 4 × 1 is 4, then I get 0 × 0, then I get 9 × 0 is 0. And then I get 7 × 6 is 42. So, this is how I computed it. And then I finally got 46. So, by any entry, I can do that, so, I can take this 54 and do the same thing.

(Refer Slide Time: 10:50)



So, I want now, if I want this 1, then I want to go to the, I want to go to 2 comma 3. So, I want to go to row 2, column 3, and now I have 8 × 6 is 48,+3 × 2 is 6,+0 × 1 is 0,+1 × 0 is 0, so, I get 54. So, this is how I compute the product.

(Refer Slide Time: 11:05)



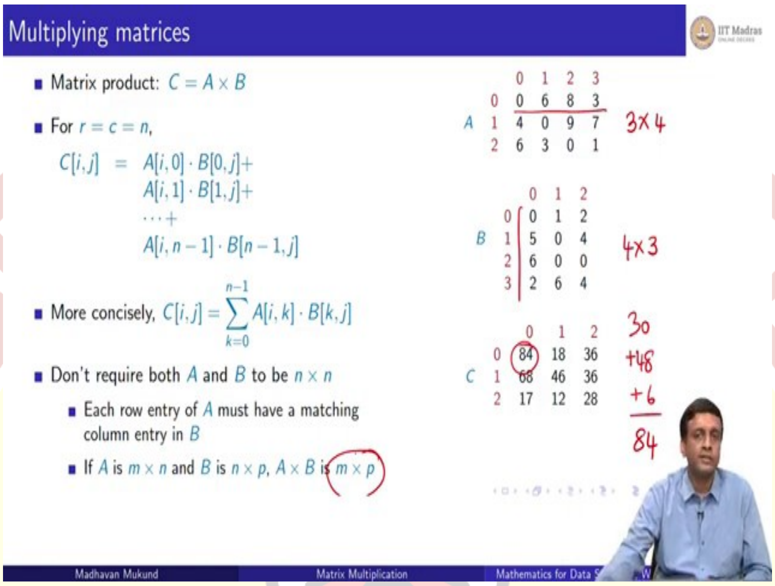
So, what we have described is called the matrix product. So, this thing on the right, by computing for every. So, what we have done in particular is a square matrix. So, r is n, c is n. So, we say C_{ij} is A_{i0}, B_{0j} +, A_{i1} B_{1j} up to $A_{i\,n-1}$, the last column, and $B_{n-1\,j}$, the last row in the column J. So, there is a simple way to write this in mathematics notation, which is that you replace this entire summation by the summation sign. So, we say, let this second position range from 0 to N-1. K. So, we take the summation over all k ranging from 0 to N-1 of Aik × Bkj. So, this is just a shortcut for writing that long sum without having to specify every term and writing dot dot dot.

So, in general, we do not need these matrices to be squared, what we want is? That this thing is well defined, for this thing to be well defined. When I am walking down row i and A, I must be able to take the same number of steps in column j and B. So, therefore, if I look at the length of a row in A, if I process some number of elements in that row, then I must be able to find a matching element for each of them to multiply and then add it. So, the length of a row in A must be the same as the length of a column and B.

So, remember that we write these things as r by c. So, if I look at A. So, it has some $r_1 \times c_1$, and look at B, it has some $r_2 \times c_2$. So, what is the length of a row in a matrix? The length of a row is a number of columns. The number of rows is how many rows there are, but I want to know how far can I go in a row. So, that is the length of a row in A is c_1. What is the height of a column in B? It is the number of rows, I go from the top row to the bottom row. So, that is r2. So, what I want is that r2=c1.

So, what we are saying is that if A is m×n, then this n must be the number of rows in B and B can have any number of columns. So, we can take 2 matrices, which have different shapes, provided the number of columns in A=the number of rows in B. Only then I can do this the summation correctly.

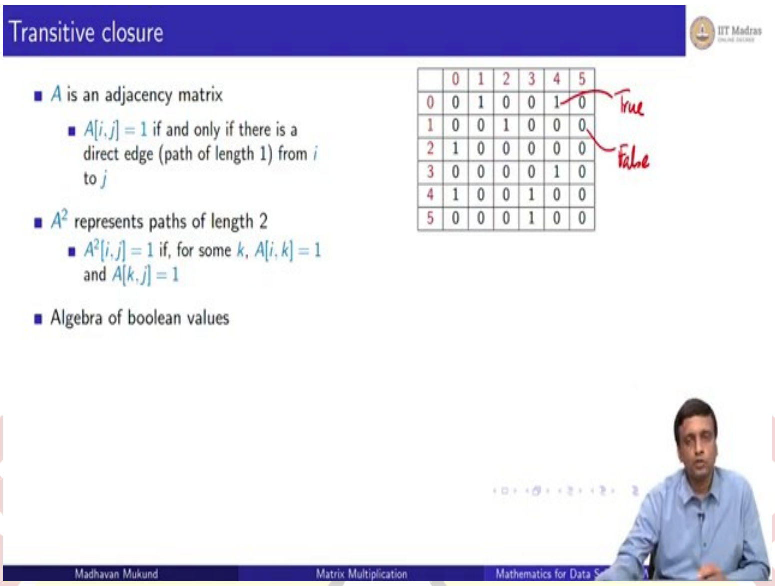
(Refer Slide Time: 13:26)



So, let us look at an example. So, what happens here is supposing now I have something which is 3 × 4, I have 3 rows and 4 columns. And here I have 4 rows and 3 columns. So, now when I take an entry here, and I multiply it by an entry here, I get 1 entry here. So, I get 6 × 530,+6 × 848,+3 × 2, 6, which is 84. So, I have an entry in C for every row and column, which is there as a row of A and a column of B.

So, finally, I end up with a matrix which has as many rows as A and as many columns as B. So, if I start with m by n, and multiply it by an n by p matrix and up with an m by p matrix, in our context, remember that we are dealing with adjacency matrix. So, now very limited context in which we are using it for graphs. This does not matter because we are always going to do square matrices, but matrix multiplication in general does not require the matrices to be square. It only requires this correspondence between the columns of A and the rows of B.

(Refer Slide Time: 14:33)



So, now let us get back to the problem which motivated all this, the problem of transitive closure. So, remember that transitive closure was trying to capture the fact that 2 vertices in a graph. So, it started off with a relation, we said that we have a relation and we want to find out whether 2 objects in our set are related by a sequence of pairs in the relation. And then we said we will model it as a graph. And now every relational pair in our relation modelled as an edge in the graph.

So, finding a sequence of relational pairs, and our relation is the same as finding a sequence of edges in our graph. So, it is the same as the reachability problem. So, can we find out which all pairs are reachable from each other? So, we started with an adjacency matrix. So, this is just some arbitrary adjacency matrix, we do not really need to worry about what this graph represents, because once we have the adjacency matrix, we said we can compute directly with it.

So, the property that we know is that adjacency matrix captures paths of length 1, in other words, direct relationships, which are already given to us and the R relation we started with it. So, the edge relation is the relation R that we are trying to visualize. So, we know all things which are directly related because that is given to us. Now we want to go from that to say things which are related indirectly in 1 step.

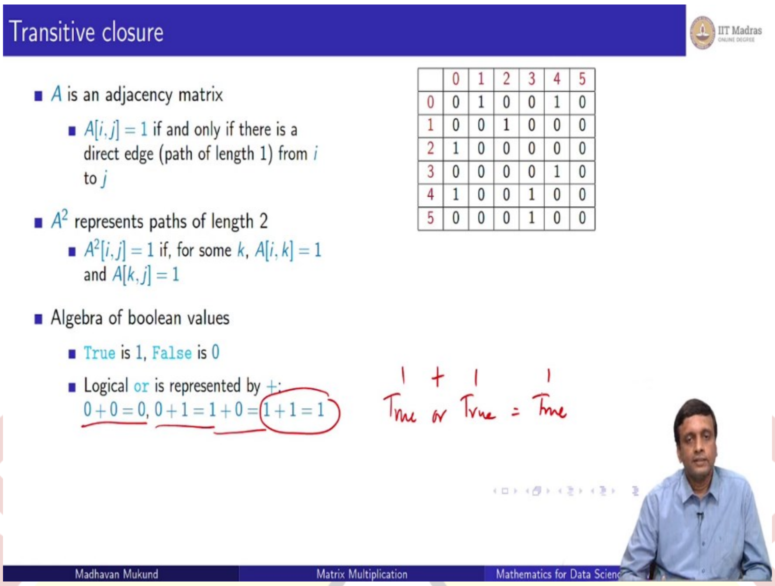
So, we look at what we call A^2, which is all pairs, which are connected by a path of length 2, by length 2, we mean there are 2 edges, I need to traverse 2 relationships to go from i to j. And we said that if you want to go from i to j in 2 steps, that means there must be an intermediate vertex, and we call that k. So, there must be some k, such that I go from i to k, and then I go from k to j. So, that was our A^2.

So, now, how do we put this in the framework of matrix multiplication, that is what we are trying to do. So, we are not doing multiplication addition over our normal numbers. We are doing multiplication and addition, over the values true and false. So, we are doing something which is called Boolean algebra. So, let us just write it down and understand what we are doing. So, remember that these things indicate that there is an edge, I can write this as true or there is no edge, this is false.

So, this is saying that there is an age from 0 to 4, there is not an edge from 1 to 5. So, the 0 is interpreted as the nonexistence of an edge. So, if this is the answer to the question is, is there an edge from 1 to 5? The answer is no, false there is no such edge, therefore, the answer is 0. And is there an edge from say 3 to 4. So, the entry a 3 4 is 1, and therefore the answer is true. So, we are working with these 2 values only. So, this is a 0 1 matrix. So, 0 represents false 1 represent true. So, this is our starting point.

And now the operations that we are interested in over Boolean values are AND and OR. So, if I want to say OR, OR of 2 Boolean values is true provided at least 1 of them is true. So, only if both of them are false is the answer false. So, if I now represent this OR by +, so, this is where now this matrix multiplication idea is going to come. So, this is the algebra. So, we are now representing the values by 0 1. So, first of all, we have removed true and false from our Boolean algebra, from our Boolean set and we have replaced it by 1 for true and 0 for false.

(Refer Slide Time: 18:08)



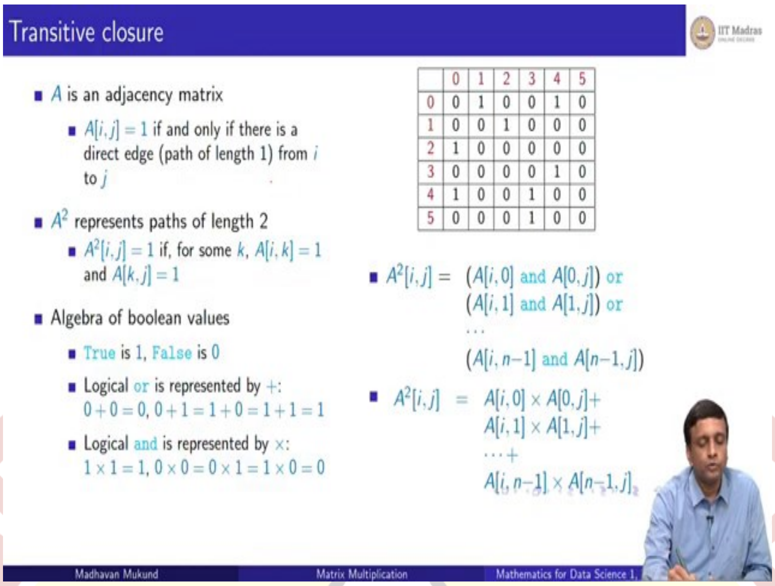
	0	1	2	3	4	5
0	0	1	0	0	1	0
1	0	0	1	0	0	0
2	1	0	0	0	0	0
3	0	0	0	0	1	0
4	1	0	0	1	0	0
5	0	0	0	1	0	0

Now we are taking operations and replacing them by what we think of as arithmetic operations. So, we are taking the logical OR which says false or false is false and false or anything any other pair gives true by saying that, if + is or and 0 is false, then 0+0 is clearly 0. 0+1 is 1 and 1+1, 1+0 is, is 1. This might even follow from the fact that there are numbers, obviously, even as integers, 1+0 is 1 and 0+1 is 1.

What is perhaps surprising from a numeric point of view is that 1+1 is also 1. So, clearly, we are thinking of these as integers, 1+1 would be 2, but 2 is not really a value that we are dealing with, we are dealing with only true and false. So, what this is saying is that true or true,=true. So, we are taking this and making it+we are taking this and making it 1 taking this and making it 1 and saying this is 1. So, this is a kind of, strange kind of arithmetic we are doing but it is justified because the underlying interpretation is in terms of logical values and logical operations.

The other interesting operation for us is AND. So, AND is kind of symmetric to OR in the sense that in AND we need both to be true. So, the only interesting case is 1×1. So, 1 × 1 is 1, that means true and true is true. And if any of them is 0, then the answer is false. So, false in anything is false. So, both are false, or 1 is false, it does not matter. So, now this is our setup. So, we are working with the 0 1 matrix. And now we have these operations multiplication which represents AND, and OR, which is represented by a +.

(Refer Slide Time: 19:34)

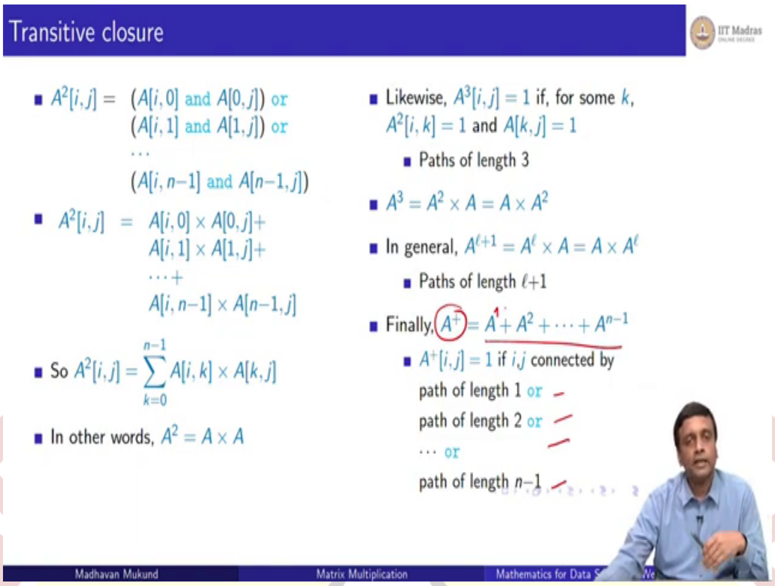


So, now let us look at what this says? It says for some k. So, A^2 ij is 1. If for some k, I can find Ak $_{ik}$ =1 and A kj =1. So, for some case, so, what are the possible values of k? k could be 0, k could be 1. So, k has to be one of the vertices. So, k could be 0, K could be 1, k could be 2, anything up to n-1.

So, this expression, this sentence here translates to saying either k is 0, in which case I have A_1 0 is true and A0 j is true, or I have k =1 in case A $_i$ 1 is true and A1 j is true and so, on. Or finally, the last possibility is that it is A $_{i\ n-1}$ and A $_{n-1\ j}$. So, if any of these pairs A I k, k A j is true simultaneously true, then there is an edge from i to j, If more than 1 is true, it is still true.

If I have multiple ways of going from i to j in 2 steps, it is still okay, I need at least 1. So, this is just the expression for this left-hand side definition, written out using the logical operations and an order and interpreting the entries in the original adjacency matrix A as true and false. But notice that this is the expression if we write it using this algebra here, we replace every AND by × and we replace every OR by +. So, this is an algebraic way of writing out this ANDs and ORs.

(Refer Slide Time: 21:04)



So, once we have written this out using ANDs and ORs, it is very clear that we have written out a matrix multiplication entry in terms of this new interpretation of ANDs and ORs. So, A^2 ij is a summation of all k from 0 to n-1, $A_{ik} \times A_{kj}$. This is exactly the equation we wrote for an arbitrary matrix product, we said C IJ is a summation from K =0 to n-1 of $A_{ik} \times B_{kj}$.

So, that is the reason why A square is just in matrix terms, A × A provided we think of a as having Boolean values 0 1. And we think of OR as +, and, AND as multiplication. So, using that interpretation, and those rules for+and ×, we get precisely A × A. Now proceed, next step in transitive closure, we look at paths of length 3, and we said a path of length 3, for instance, could break up as a path of length 2, followed by a path of length 1. So, I need to go from I to k in 2 steps, and then from k to j in 1 step.

So, then, in the same logic, we can compute that this is the all over all k of this. So, it is a summation over all k of A^2 ik × A k j, because A^2 captures the length 2 path, A captures the length 1 paths. Now of course, we could also decompose this length 3 paths as a length 1 path from I to k and a length 2 parts from k to j is the same thing, I just have to reduce that path of length 3 to 2 things.

I already know either 2+1, or 1+2. So, it could be A^2× A, it could also be A × A^2, the intermediate k could be first finding an entry in A and then finding an entry from k to j in a square, it does not matter. So, A^3 is either A^2× A or a 10 A^2. Notice that here, when I did A square, I did not have an option, because both are A. So, it does not, there is only 1 way to decompose, a 2-length path is 1+1, because that is the only information I have. But for a 3-length path, I can do 1+2, or I can do 2+1.

Now, in general, we said if I want to, if I already knew paths of length l, then I can extend to l+1 by again multiplying by A. So, I can take all paths, which are Connect pairs, which are connected by a path of length l, extended by 1 more thing, and then I get all paths of length l+1. So, each successive length path can be got by matrix multiplication. Again, we could also invert it, we can first take 1 step and then take l steps. And of course, you could do other things also, supposing l is 7, you could first take 3 steps, and then take 4 steps. But it is convenient to write it in a uniform way. So, that we know how to do it systematically from 1, 2, 3, onwards.

So, this is how we get paths of different lengths by consistently multiplying 1 more time by A. So, we take the matrix, we already have for L steps, multiply 1 more time by A, So, each time we multiply by A, we take the matrix we have computed in the previous step and multiply 1 more time base. So, it is A × A × A × A. So, implicitly, this thing contains a lot of A × A × A in it L ×. So, there are l × that, and then 1 more time.

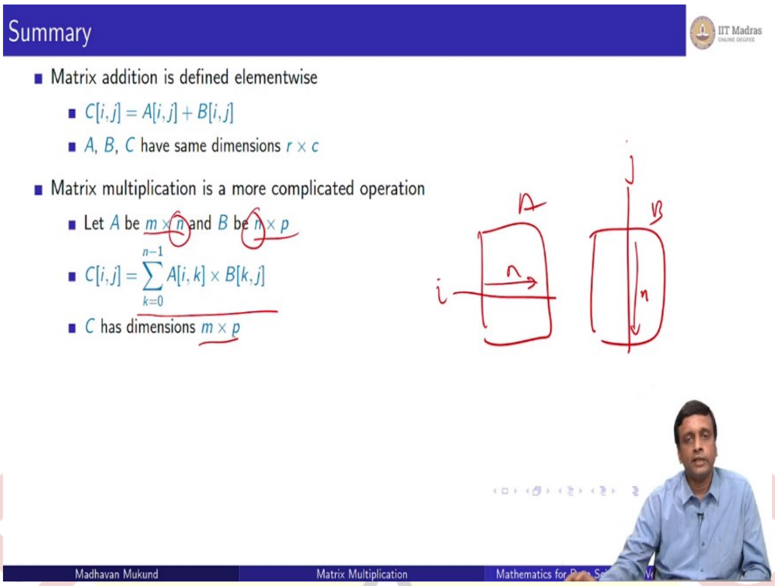
So, this now gives us paths of specifically of length 1, length 2, and 3 and all that but what we said is we wanted to know whether there was a path of any length. And then we observed that of course, because it is a graph. If there is a path at all, there must be a part of at most length n-1 because if it is longer than that, then there will be a loop which you can remove, because there are only n vertices in the whole thing.

So, if I take n-1 steps, starting from a vertex, and each time I go to a new vertex, there are only n-1 new vertices I can go to, if I take 1 more step, then the next vertex must be 1 of the n vertices I have already seen before, and that would be useless. So, therefore, it is sufficient to look for this matrix of length, path lengths up to N-1. So, what we want is, is there a path of length 1 or link 2 or length 3 or this but each of them is represented by a matrix and OR is +.

And so, now we can go back to matrix addition, we want for each ij, we want to check whether A_{ij} is 1, or A^2_{ij} is 1, or A^3_{ij}s 1, or A^4_{ij} is 1 dot dot dot upto A^{n-1}_{ij} is 1, we want to check if any of them is 1. And that is the same as doing++++. So, this is why we get the transitive closure, which we wrote as A to the power +, the transitive closure is just the sum the matrix sum of these n-1 matrix is the A, A, which is this is really A to the power 1 implicitly.

So, A which is the 1 length path, A^2 which is the 2 length path, and so, on. And, that is how we got from transitive closure to matrix multiplication, by using the entries in A as Boolean values and interpreting the 2 operations+and × as logical OR, and logical AND.

(Refer Slide Time: 25:49)

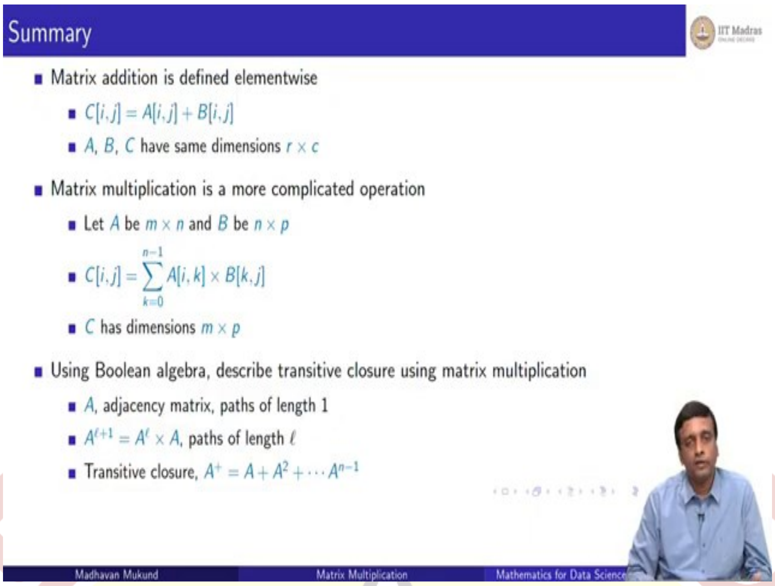


So, to summarize, what we have seen is that we can take matrices and operate them as a whole. And depending on what operation we are doing, the operation is defined differently. So, matrix addition is very simple. It is just defined element wise. And it tells us that the ij'th entry of the final sum is the sum of the 2 ij'th entries in the initial matrices. And for this, we need that A and B are compatible in the sense that they have the same number of rows and the same number of columns. So, then we get C=A+B.

On the other hand, if we want to do a matrix multiplication, then we have to take for each entry ij. In the final thing, we have to go through row i and column j of A and B, and then pairwise, multiply all the terms and then add them up. So, that is the summation. And for this, we need that the number of rows a number of columns in a must be =the number of rows.

So, if I walk right, some n ×, then I must be able to walk down some n ×. So, n is the number of columns in A, it must be =the number of rows and B. So, we have this constraint that A and B must agree on this component, the number of columns in A and the number of rows and B. And finally, we end up with a matrix, which has as many rows as A and as many columns as B.

(Refer Slide Time: 27:05)



And then we saw that, if we interpret the adjacency matrix as a Boolean matrix, the values 0 and 1 as being representing true and false, and we think of the operations AND and OR as being represented by multiplication and addition, then we can think of our transitive closure computation as initially doing a bunch of matrix products. So, we get A to the power l+1 as $A^l \times A$.

And finally, we add up all the matrices we have computed from A^1 the original 1 up to $A^{(n-1)}$, where n is the number of vertices in our graph.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

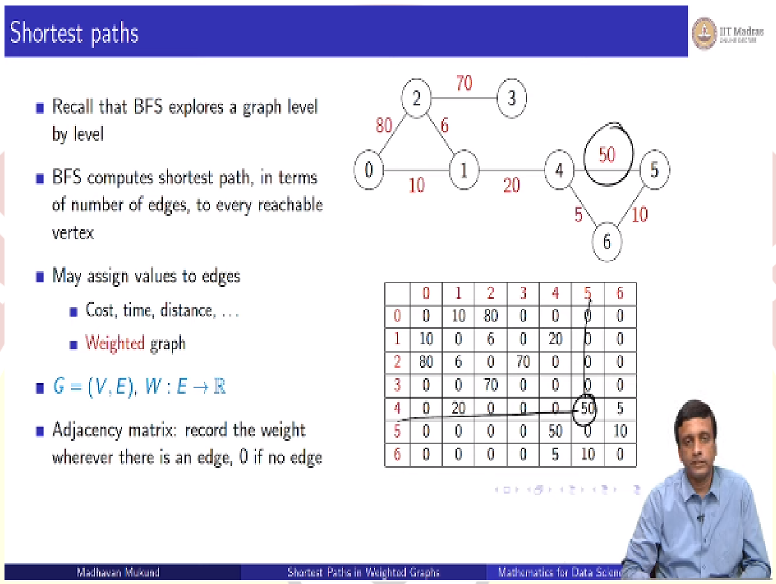
सिद्धिर्भवति कर्मजा

IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Chennai Mathematical Institute
Lecture No. 12.1
Shortest Paths in Weighted Graphs

(Refer Slide Time: 0:18)



So now, let us look at a new type of graph called a weighted graph. So, remember that in a graph like this we have seen that a systematic way to explore this graph is breadth first search and breadth first search explores this graph level by level and therefore because it does it level by level, it discovers the vertices reachable from the starting vertex at successively longer distances and therefore, the BFS computes the shortest path in terms of number of edges to every reachable vertex.

Now in practice we often assign some values to the edges, so if you look at for example, a road map then you might see some numbers against each section of road representing the length of the road. Similarly, if you look at say a railway map or a airline map you might see the time it takes to do a segment of a journey or it could be the distance or it could be even the cost you know how much does this ticket cost.

So, these numbers indicate some more abstract information about the length of an edge than just the fact that I take one edge, so it is not enough for me to say that I took two flights, I need to know how long these flights are. So, if I take a hopping flight in a short distance say I go from Chennai to Bangalore to Mangalore, right it is not the same as taking a hopping flight which takes me from Chennai to Delhi to San Francisco. So, the two are still two length travels but one is enormously longer than the other in time.

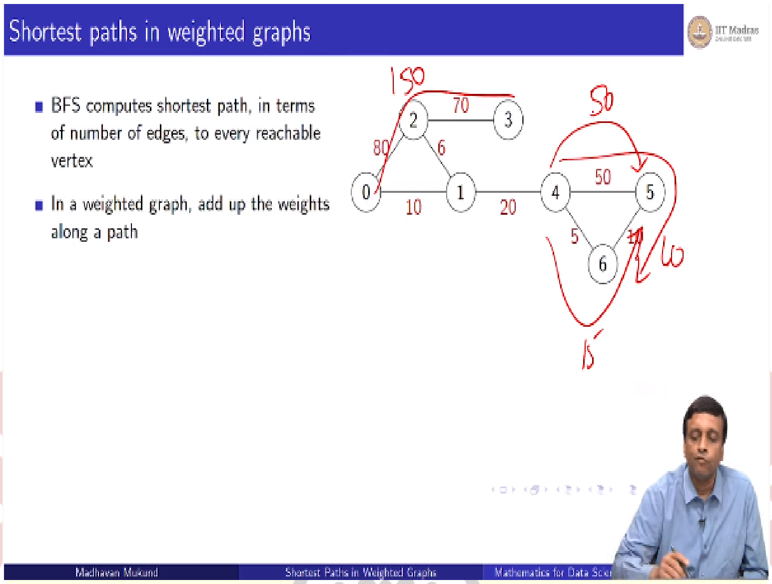
So, this is the kind of information that we want, so we want to take this kind of a weighted graph, so what is a weighted graph formally? A weighted graph is just a graph, so we have a set of vertices, we have a set of edges but we have weights, so weight formula is a function, right, a function which takes every edge that is present in the graph and assigns it some real number.

So, notice that we are not claiming at this moment that the real number is positive or negative, we will discuss what negative weights will mean but most of the interesting things that we can think of the weights will be 0 or more. So, we can think of a 0 cost edge sometimes the cost an edge which is not there has a 0 cost edge or sometimes it may be there but typically in any reasonable scenario weights are positive but we will see a situation where weights could actually be negative and make sense.

So, the first thing is we have to, we are going to work with these graphs and we have an adjacency matrix way of working with a graph in general which records the presence of an edge. So, how do we record a weighted graph, so what is our representation of weighted graphs? So, in adjacency matrix what we can do is that whenever we normally put a 1 saying there is an edge, instead of the 1 we can put the weight, so assuming that there are no 0 weights, then wherever there is a 0 there is no edge, wherever there is a 1 there is a weight.

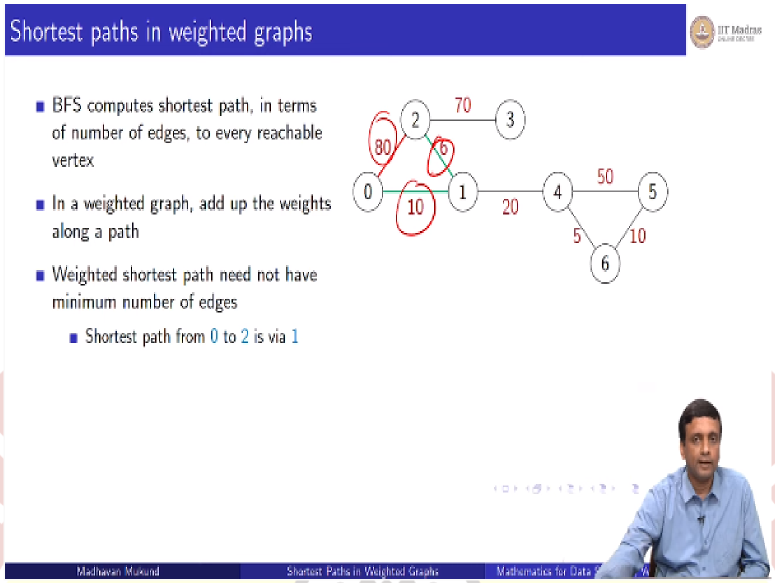
So, if you look at this graph here for instance, if I look at the edge for instance 4 to 5, then I look at the entry 4, 5 and now I have a 50 there rather than just a 1. So, this is a very simple way to represent weighted graphs just take the adjacency matrix and at each entry i comma j put the weight of the edge i comma j and if there is no edge or if the weight is 0 put a 0.

(Refer Slide Time: 3:18)



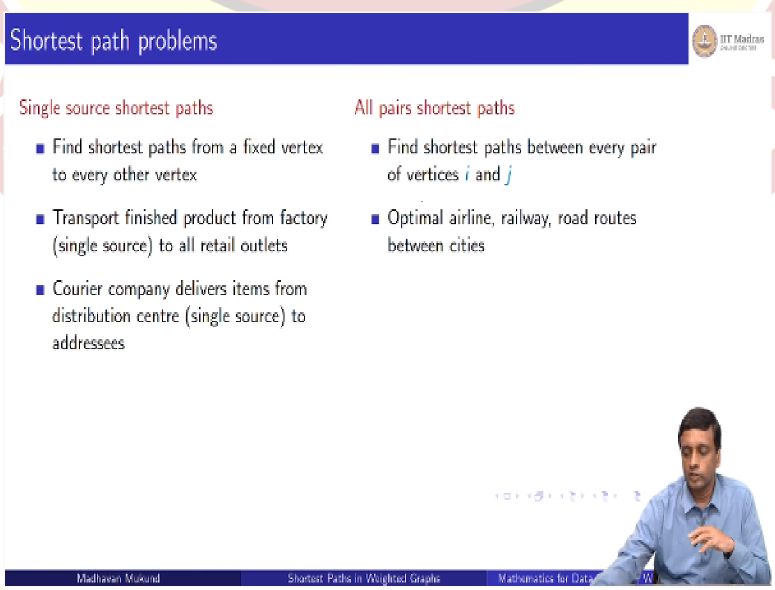
So, our interest is initially to compute shortest paths in such graphs, so we have these weighted graphs, where we have some edge weights between edges and we want to find the shortest path, so what is the shortest path now? For us the shortest path will be the sum of the weight, so for example, if I take this path, then the weight of that path, the length of that path is 80 plus 70 is 150. If, I take this path for instance, the length of that path is 60. More interestingly if I take this path the length of that part is 15 whereas the direct path from here to here is just is 50. So, going from 4 to 5 via 6 is actually shorter than going from 4 to 5.

(Refer Slide Time: 4:02)



So, in general this is the situation that in a weighted graph, the weighted shortest path does not need to have the minimum number of edges in the unweighted sense. So, if I look at the very beginning of this graph from 0 I can go from 0 to 2 in one step but then I pay a cost of 80, whereas if I go from 0 to 1 and 1 to 2, so I take two steps then actually I get a cost which is 10 plus 6 which is much less 16. So, the shortest path from 0 to 2 is actually via 1, an indirect path.

(Refer Slide Time: 4:30)



So now, what are we going to try and solve with these weighted graphs? Well there are more than one type of shortest path problem, so the first type of shortest path problem is one where we start at a fixed vertex and we want to find out the distance to every other vertex, so this is called the single source shortest path problem. So, we would start a fixed vertex and find out how long it is from this vertex to every other vertex, why is this an interesting problem?

Well, there are many applications where this is interesting for instance, suppose you are a manufacturer, so you have a factory and you make things and now you have to take things from your factory, your finished products and distribute them to the shops where they are sold, the retail outlets. So, you have a single source your factory and then you have to find the most efficient way, so the shortest path in terms of whatever you are measuring the cost of travelling or the time it takes to travel or the distance it takes to travel, whatever is the cost that you want to count towards the transportation cost, you would like to find the shortest transportation cost from your factory to every one of your retail outlets. So, this is a single source shortest path to every other vertex.

Alternatively for instance, you could be a courier company, so what happens in a courier company is that they have these flights between cities, so all the packages which go to say Delhi come from different destinations and they land in Delhi and they go to a centralized clearing facility in Delhi. So, overnight you might have flights coming from Calcutta, from Bombay, from Bangalore, from Chennai and all that all this information, all these packages come to Delhi and now they have to deliver them out.

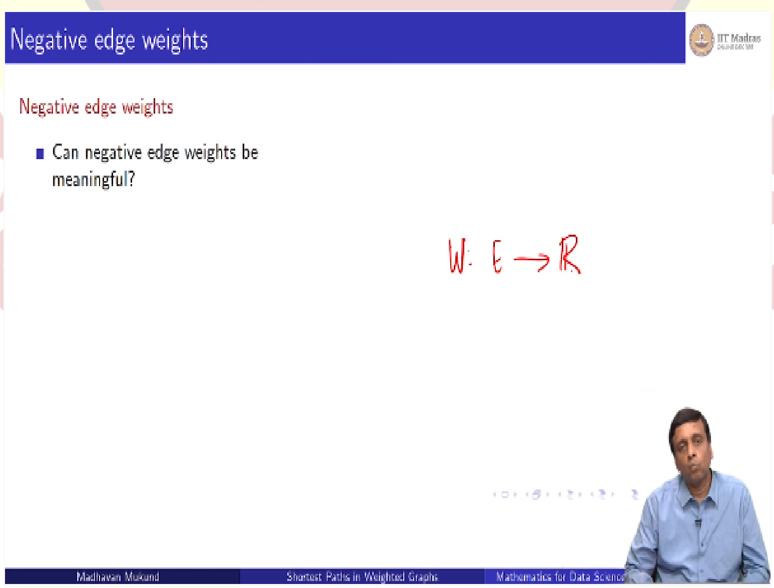
So, the starting point, the distribution center where all these things are initially brought in from the airport or air cargo wherever they come, that is a single source and now they have to now find the most efficient way to distribute it to all the destinations where these packages have to be. So, the single source shortest path probably has a number of applications and therefore it is an interesting version to solve.

On the other hand, sometimes we want to know something about every pair. Now, of course, you could take the single source thing and start from every vertex and find every other vertice distance and then you will get all pairs but generally there may be a better way to find the distance between all pairs. So, for every i and every j we want to find out the shortest distance from i to j. So, the single source is fixed and starting point, say fix vertex 7, and from 7 what is the shortest distance to everything?

Now, this is for every i and j, from not only from 7 to j, I want to know from 9 to j, from 11 to j, from 7 to 11, from 9 to 11 and so on and this is the kind of thing typically that say if you are managing a booking site and somebody says I want to go from city A to city B, then you have to be able to provide in terms of the cost or time or some metric the cheapest way to go from A to B.

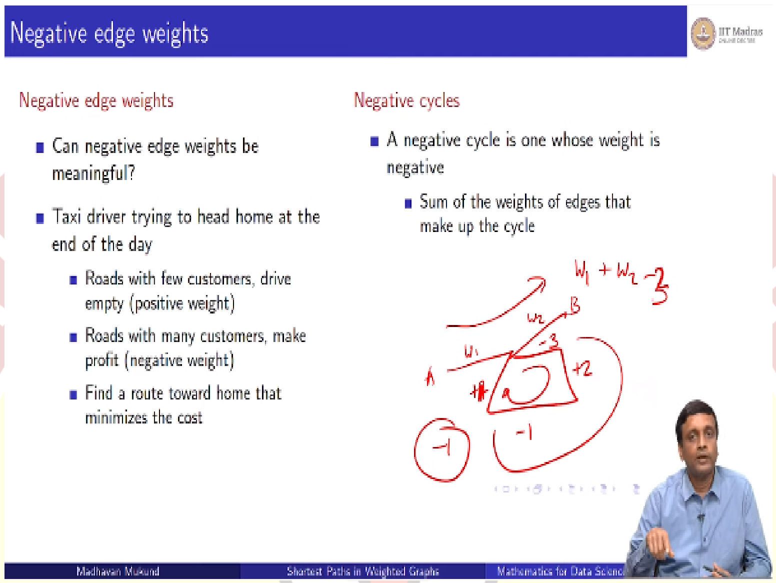
So, somebody might say that I want to reach there as fast as possible, so what is the shortest flight? Some people might say I want the cheapest ticket. So, based on whatever is the cost that you are associating, then you will have to find and then you need to be able to do this for any pair because you have customers who can be going from anywhere to anywhere. So, this is the all pair shortest path problem, so these are the two problems that we will initially look at in the context of weighted graphs, single source shortest paths and all pair shortest paths.

(Refer Slide Time: 7:45)



Now at the beginning I alluded to the fact that we have this function, so we said that we have this function which takes every edge and gives us an arbitrary number and in principle there is nothing to prevent this number, this weight of an edge from being negative, so what if I am thinking of this cost, what would be a reasonable scenario where I could have negative edge weights?

(Refer Slide Time: 8:12)



So now, imagine that you are an Uber driver or an Ola driver or any one of these cab companies, so you have a certain amount of hours when you drive your cab and at a certain point you want to start heading home and reach home roughly when you finish your work rather than eventually traveling across the entire city empty, so this is all, I mean many of us in days when we have taken Ubers, I found that towards the end of the day for example it is little difficult sometimes to get long distance things if the driver is not living in that direction and they would say 'No, no sorry I am going in the opposite direction because I am heading home'.

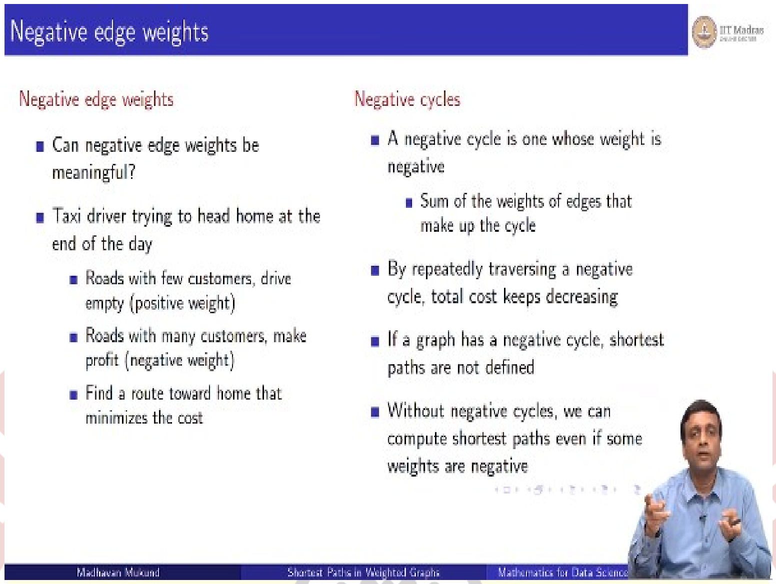
So, here is a taxi driver trying to head home at the end of the day, so maybe he has an hour of service left and he wants to try and optimize where he lands up at the end of this hour, so he has a minimum amount of time to travel home. So, now he has to take a choice, so he has to start looking for customers maybe right, so if he takes a road which has very few customers then he will be losing money, so there will be a cost that he is paying, so that is a positive cost.

On the other hand, if you travel on a road with many customers then you are likely to find somebody who will hail you for a ride, you might get a call so therefore you will earn money, so you have a negative cost, so where you are not taking customers, you are paying for driving the car, you are paying of petrol cost and other costs and therefore, you are losing money so that is a positive cost, where you are gaining money, where you are earning money it is a negative cost and you want to obviously get more money, so you want to reduce the cost you want to make the negatives more than the positive. So, you want to find actually a route towards home which minimizes the cost, so this is one example where negative edge weights make sense.

Now what happens to our shortest path problem in the presence of negative edge weights? So, the problem is not so much with negative edge weights but negative cycles. So, supposing I have somewhere in my graph, a cycle which has something like minus 3, plus 2, minus 1 and plus 2 for example, if I go around the cycle then what do I do? I add up the weight, so I get minus 3 plus 2 is minus 1, right, minus 1 plus, so maybe I should maybe make this also plus 1, so minus 3 plus 2 is minus 1, minus 1 minus 1 minus 2, minus 2 plus 1 is minus 1, so the total weight of the cycle is minus 1. In other words if I go around the cycle once I reduce my cost by 1.

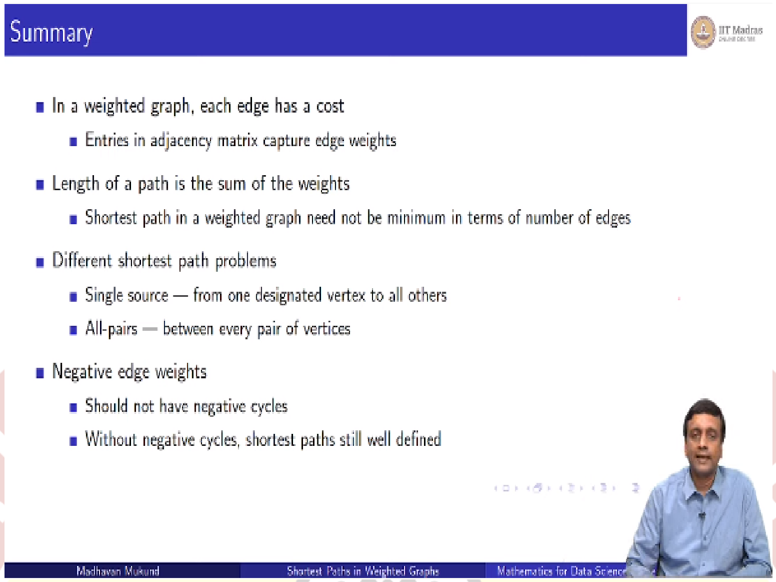
So, if I am supposing I am going from A to B, so there could be some arbitrary weights here W1 and W2, so I do W1 plus W2 to go from here to there but if I want to reduce it I go around the cycle uselessly once and I get minus 1, I go around it again and I get minus 2, I go around it again and I get minus 3, so I can go around the cycle as many times as I want, and keep reducing my cost, I can make it as negative as I want and it does not make sense because I mean we are actually taking longer and longer paths but because the cycle has a negative cost we are able to do this, so this is clearly something which is, which makes the problem undefined.

(Refer Slide Time: 11:10)



So, if we have a negative cycle we can just go round and round that cycle and then there are no shortest paths anymore on anything which goes past that cycle because every time I want to reduce the cost just go around cycle once. But if I do not have negative cycles then it is fine. So, when a graph has negative cycles, shortest paths are not defined but if you have negative edges you might have some edges which are negative but you do not have negative cycles, then it is fine. So, if you have negative edges but no negative cycles you can still do shortest paths but you have to then be careful that your algorithm does not depend on the edges being positive.

(Refer Slide Time: 11:45)



To, summarize we have looked at what is called a weighted graph, so in a weighted graph we attach to every edge a cost, a number, which we call the weight and these can be described for us in our adjacency matrix by entering the cost or the weight of every edge in place of a 1. So, then once we have these edge weights, then we can measure the length of a path in terms of the weight. So, not just how many steps I take in terms of edges but what is the total sum of the weights across these edges.

And so now I get a new notion of shortest path which is probably more natural from the way we think about graphs representing sort of spatial things at least, so we get the sum of all the edges that we traverse but the weights of the sums, not just the number of edges and we saw that this now will give us something which is not necessarily the same as the shortest path in terms of number of edges. We could have a shorter path which has a longer higher cost as compared to a longer path.

So, we said that there are two types of shortest path problems with at least two types which we will find interesting, one is the single source path where we start at a fixed vertex and we want to find out where we can, how fast we can go to every other vertex, so this is for example the delivery problem for a courier company or we have the all pairs problem which is typically the type of problem that you need to solve if you run a travel agency, you need to be able to tell somebody from anywhere to anywhere what is the best way to go.

And finally, we looked at this peculiar problem of negative edge weight, so we gave a justification that there can be reasonable situations which are modeled by negative edge weights and if we still want to be able to compute shortest paths in the presence of negative edge weights, what we need to ensure is that there are no negative cycles because if we have negative cycles, then the shortest path is not defined but we do not have negative cycles even if we have negative edge weights, we can hope to find shortest paths.

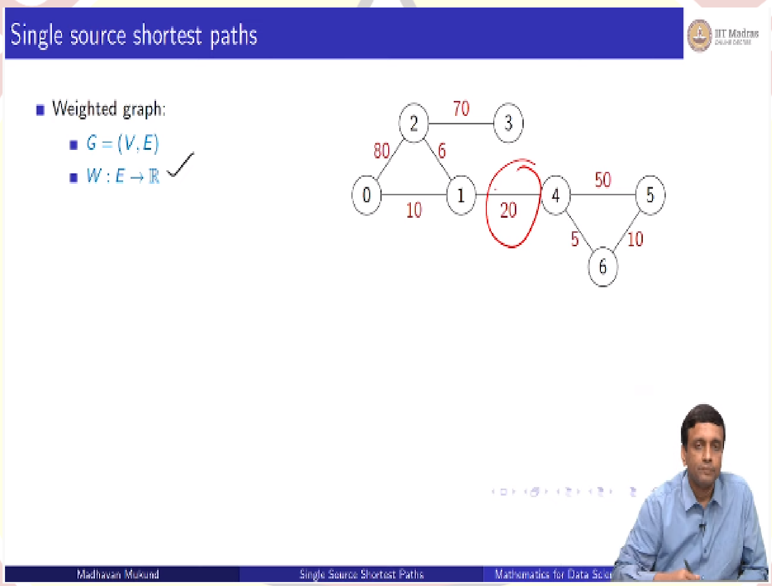
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Chennai Mathematical Institute
Lecture No. 12.2
Single Source Shortest Paths

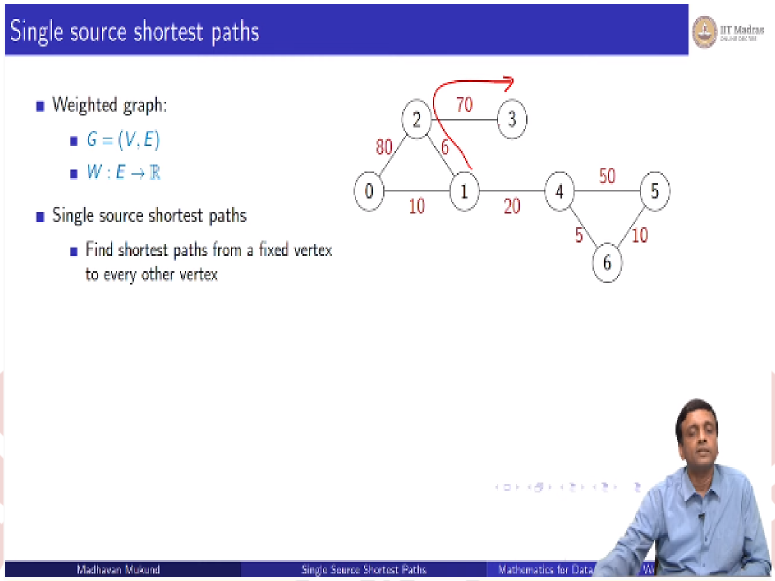
So, we are looking at weighted graphs, and we are looking at shortest path problems on weighted graphs and we said that there are two types of shortest path problem that we will focus on, the single source shortest path and the all pair shortest path. So, let us look at how we can compute single source shortest paths in weighted graphs.

(Refer Slide Time: 0:32)



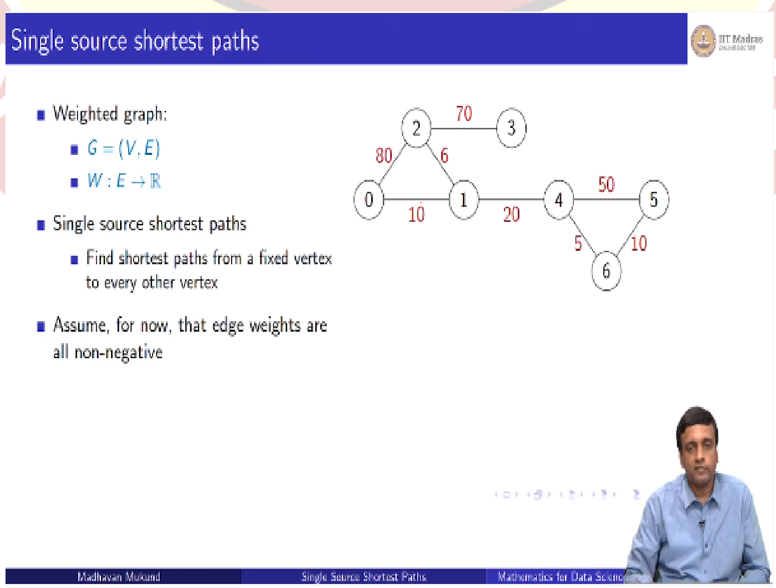
So, remember that a weighted graph is just a graph, which has in addition, this weight function, this weight function assigns to every edge some real number. So, here on the right we have a graph where at every edge we have a number written against it, saying, for example that the weight of the edge from 1 to 4 for instance, is 20.

(Refer Slide Time: 0:53)



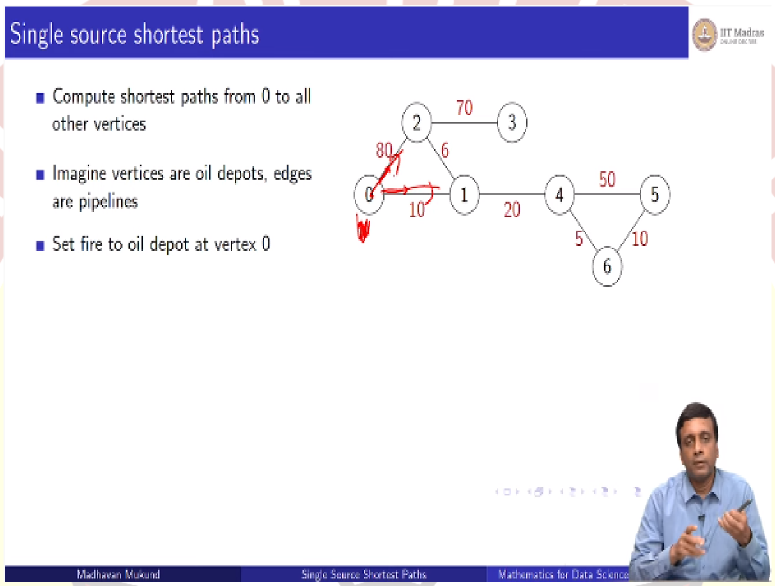
So, in the single source shortest path problem, we want to look at these edge weights, so now we are taking paths as representing the edges underneath as representing the weights underneath, so if I take a part, for instance, if I take this path, then the total length of this path the 6 plus 70, it is not two the edges but the weights attached to it that I have to add up. So, we want to find some source vertex, it could be any vertex, I start from there and I want to find the shortest path to every other vertex in the graph.

(Refer Slide Time: 1:23)



So, in order to solve this we will first assume that edge weights are all non negative, that is we could have 0 weights, but we certainly do not have negative weights, otherwise algorithms that we are going to look at will not work and we will also explain why it will not work. So, for now we are looking at graphs like the one here, where all the edge weights are either 0 or positive, not negative.

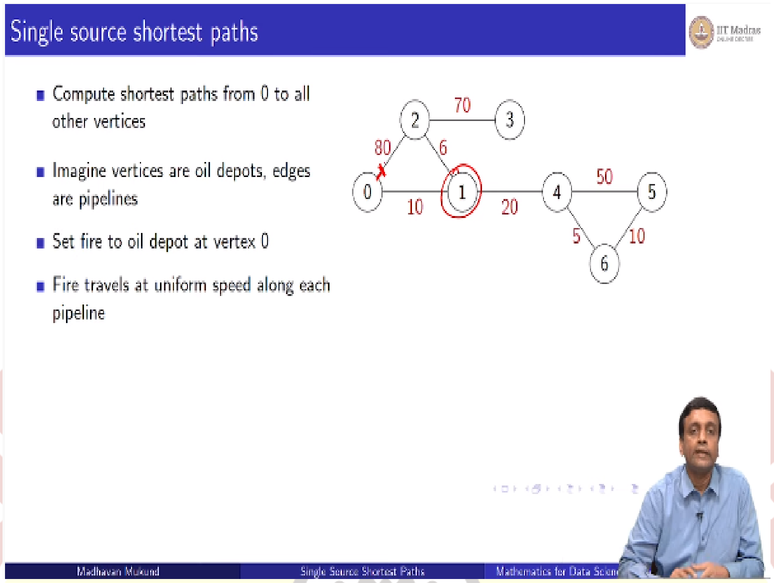
(Refer Slide Time: 1:48)



So, let us say that our single source in this particular example is 0. So, we want to find the shortest path from 0 to every other vertex in this graph. So, how would we do this? So, one way to imagine this or to visualize how the algorithm works, is to imagine that these are all some kind of pipelines carrying oil and all these nodes are actually oil depots which are storage tanks full of oil.

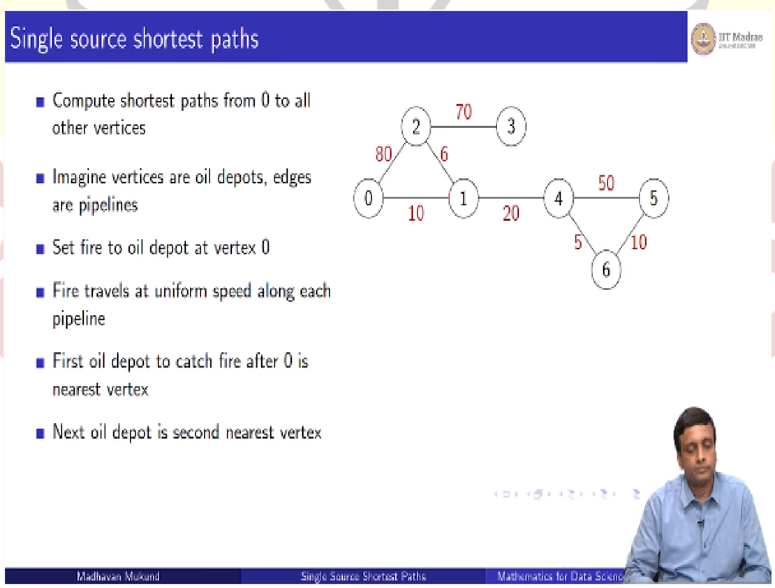
So, now imagine what happens if we set fire to this thing, so we set fire here. So, what will happen is that the fire will start traveling away from 0 along the pipeline because the pipeline also has oil, so you burn the tank and the pipeline burns. But of course, if you have ever burned firecrackers you know that it takes time for the fire to go along the thread, the wick. So, it takes time for the fire to travel, but it travels in all directions at the same speed. So, the fire, fire is traveling in this direction towards 1, it is also traveling in this direction towards 2, but the length of the edge determines which one will burn next.

(Refer Slide Time: 2:59)



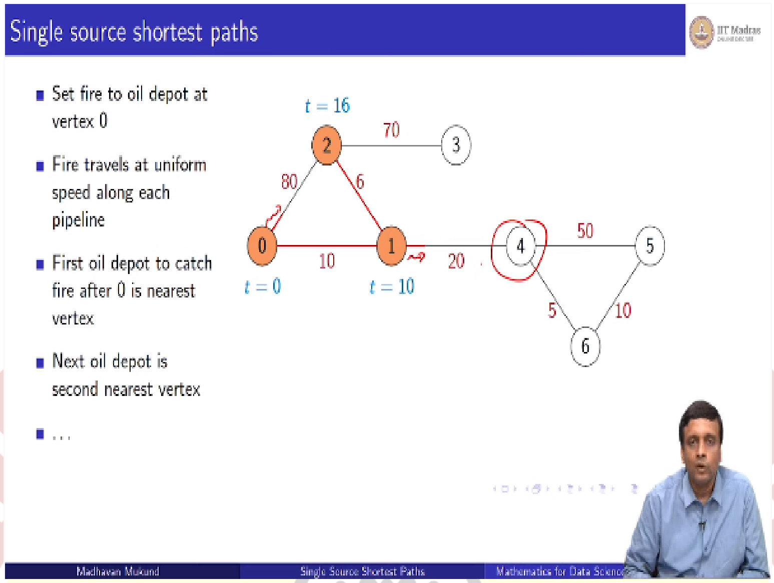
So now, if the fire is traveling at a uniform speed, then after a certain amount of time, right, after some amount of time this one will burn. At that point, because this is 10 and that is 80 the fire would have only reached up to about here. About one eighth of the way.

(Refer Slide Time: 3:16)



And now, once the next vertex burns, it will start fires in these two directions, so now I have a fire running in that direction, this thing is fully burned and so on. So, this is the intuitive idea that we want to capture. So, let us see how it works.

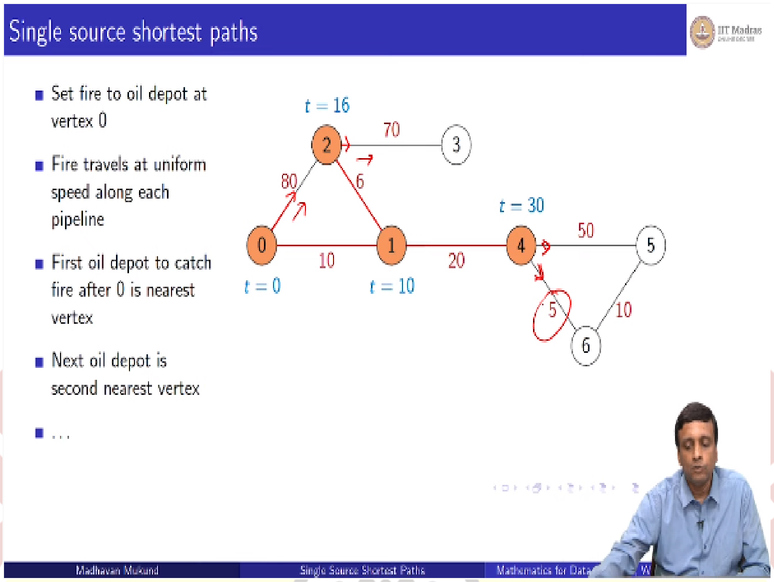
(Refer Slide Time: 3:34)



So, we start with this thing, saying that initially when time is 0, we burn the zeroth vertex and now the fire we said starts moving in these two directions. So, if we imagine the fire burning then after some, let us assume that fire moves at one unit of this, one unit length per one unit time. So, here it has to move 10 units of length, so in 10 units of time it reaches vertex 1. Now in 10 units of time as we have indicated here, it has traveled some short distance along the edge from 0 to 2 as well, but it is not yet anywhere near 2.

Now that this is burned, now something else is going to happen because I am going to end up having fire going in these two directions. So, after 6 units of time from this, vertex 2 is going to catch fire. Notice that this fire which started from 1 is still growing but it is nowhere near 2 and meanwhile, this thing has also started traveling towards 4. So, what is going to happen next is that this fire is going to reach this at time 10 plus 20, remember it is assuming the fire is moving at 1 unit distance per unit time.

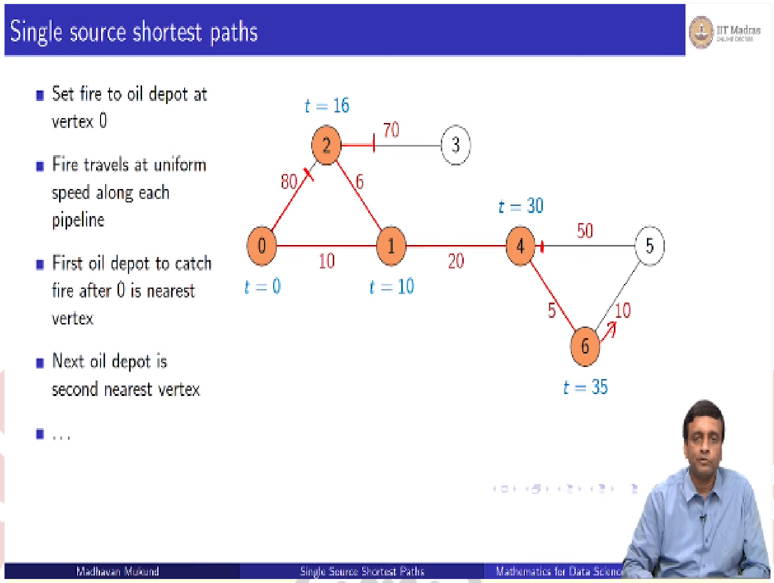
(Refer Slide Time: 4:48)



So, at t equal to 30 vertex 4 catches fire and meanwhile, this fire which is at 2 has started moving there and this fire now is kind of reached a little bit less than halfway from 0 to 2 because it has (())(5:02) units so we have already spent 30 times units so far, so suppose to be about three eighths of the way from 0 to 2.

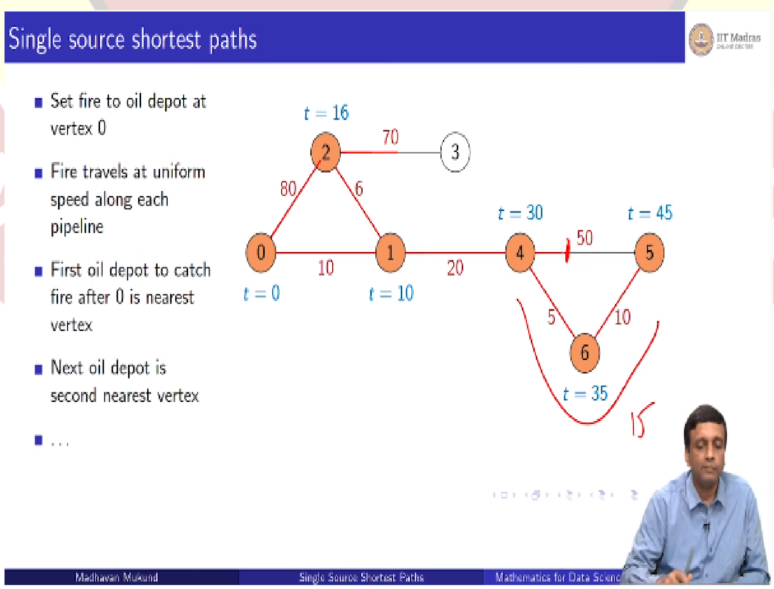
And now what happens here is that this fire now starts going in these two directions, so this fire is going here, this fire is going here, these 3 edges have all burned and now two new fire start, so which one is going to reach the next one? Well, this is only 5 units of time, so in 5 more units of time the fire from 4 will reach 6.

(Refer Slide Time: 5:30)



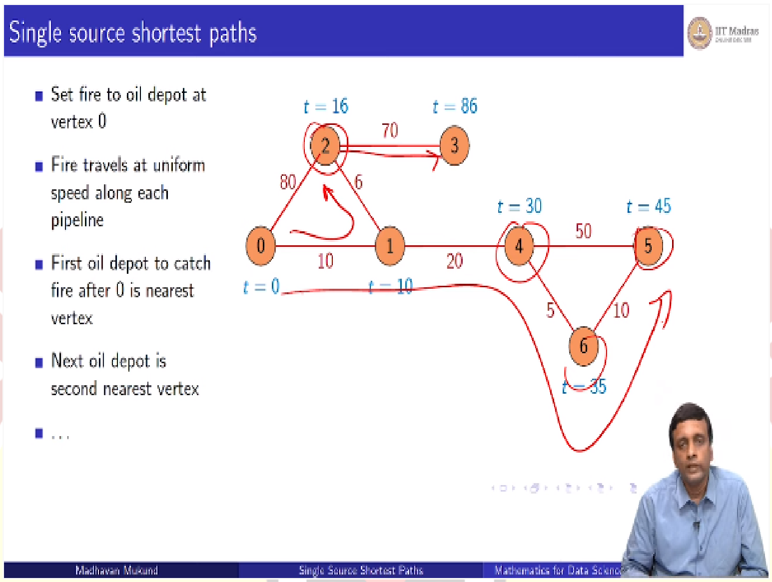
So, at t equal to 35 vertex 6 is going to burn. Meanwhile, this thing is started, so in 5 units of time this fire has reached a little distance, this has made more progress, this has made more progress and then from here, a separate fire starts going towards vertex 5 in addition to the old fire which is coming from 4.

(Refer Slide Time: 5:52)



So, now the separate fire is going to reach in 10 units of time so at t equal to 45 our vertex 5 is going to burn at this point, we have gone down 15 units of time have passed since 4 was burned, so this fire has only moved 15 by 50 of the way, about one third of the way there and meanwhile, these things are still going.

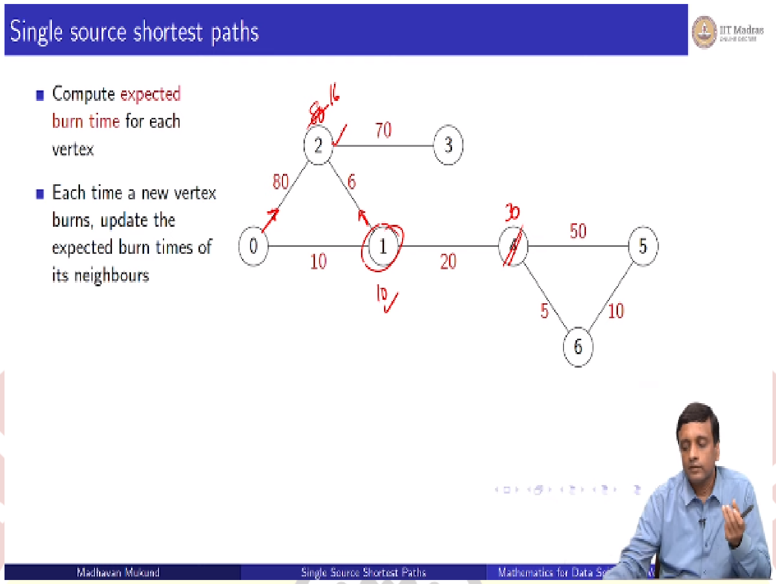
(Refer Slide Time: 6:10)



And then finally, what happens is that at time 86 this fire which started from here reaches here, at this point everything has burnt, so all our vertices have burned but all the edges have also burned it so happens, it does not have to happen always, but now we have discovered the earliest time at which every vertex burns and now we can check that these earliest times are actually the shortest paths.

So, we have discovered some interesting shortest paths for instance, we have discovered this shortest path, we have also discovered this shortest path, so the shortest path to 5 is not 0, 1, 4, 5 but 0, 1, 4, 6, 5. So, this is how this algorithm works. So, let us try to understand how we would actually do this calculation that we said by drawing these burning pipelines, so how do you keep track of when the pipeline is going to burn next.

(Refer Slide Time: 7:00)

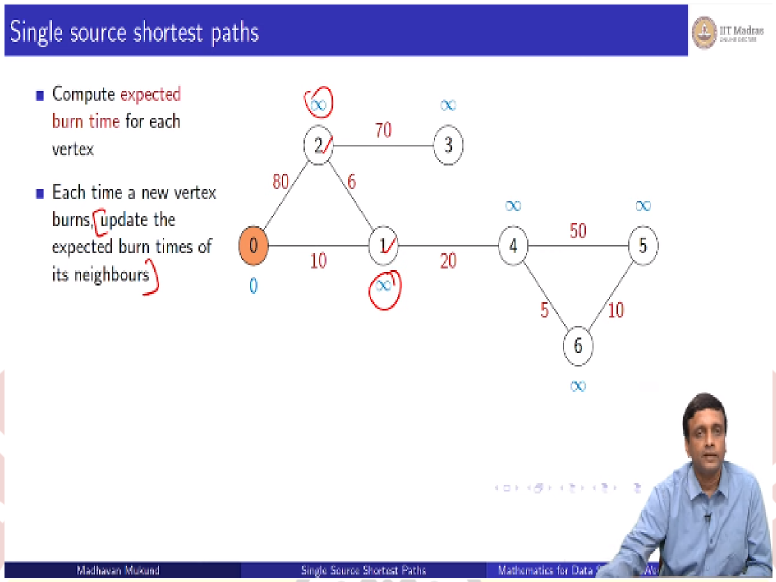


So, what we do is we compute at every given point when each vertex is expected to burn. Based on the information that we have so far about which pipelines are burning and which oil depots have burned, we compute the expected time to burn, now among all the vertices, the one which is expected to burn next is going to burn next. So, at that point we burn it, so we update the vertex which is burned. And once a vertex burns, it starts. Remember, new fires along its neighbors towards, so its neighbors.

So, we have to check whether any of its neighbors will burn now faster because of this, so this is what we saw, so we said that when 0 burned, we started a fire here, so we could expect that this will burn at time 80. But after time 10 when this burn, a new fire starts so we said if this burns at 10, then this is actually not going to burn 80, this is going to burn at 16.

So, each time a vertex burns, we figure out something about the neighbors, this is going to burn at 30, at least cannot burn beyond 30 because I already know that a fire is started at time 10 and it is going to reach there in 20 units, similarly a fire has started here at 10 and it is going to reach here in 6 units. So, 2 cannot burn any later than 16, maybe it will burn even earlier but not later than 16. So, each time we burn a vertex, we update the burn time.

(Refer Slide Time: 8:19)

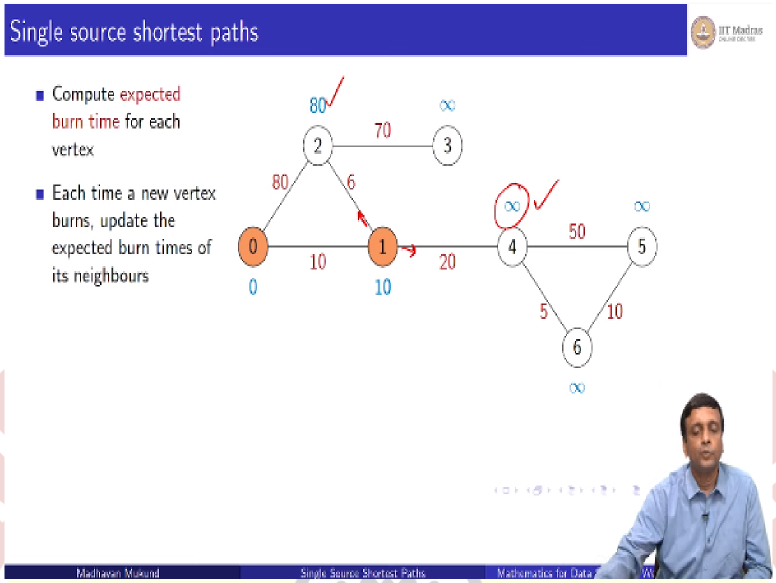


So, let us see how this goes. So, initially we know nothing, so initially when we start this whole process every graph, every vertex is going to burn at some indefinite time in the future, so as far as we are concerned it is never going to burn, at time infinity. Maybe the fire is never going to reach there, we have no idea. Now we said, we will start at a single source, so for the single source we know that the burn time is 0, because that is when we start ticking, the clock starts ticking at 0 so, at time 0 we burn our source vertex which in our case is vertex 0.

Now, this is where we start doing this, so we update the burn time of the neighbors. So, once we have burned 0 at times 0, we have to look at its neighbors namely 2 and 1 and say, do I know something better, what do I know about 2? Well, as far as I know 2 is never going to burn but now I know that 2 is going to burn at least by time 80, after time 80 it is there is no chance that 2 is not burnt.

Similarly, 1 will burn at least by time 10, because 0 is burned and the fire has started moving in that direction. So, I update these two entries, these two entries can now be updated with better information I can reduce it to 80 and 10.

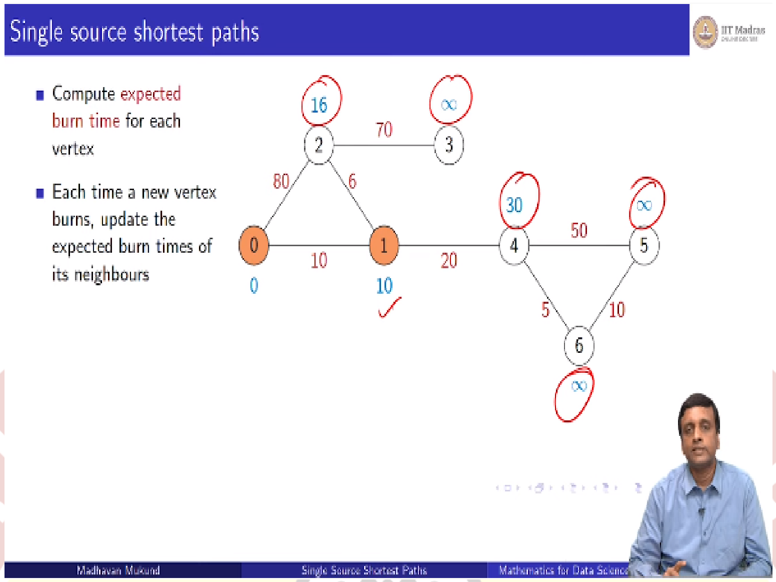
(Refer Slide Time: 9:24)



Now, I look at all the vertices that I know about and I see which one is going to burn next. So, I look for the smallest unburned vertex, I look for one with the smallest time to burn. So, the smallest time to burn at this point is 10. So, I go to 10 and I say okay you are going burn next. Now, the same thing happens when I burn 10, it starts two new fires if you remember, so it will start a fire back of course there is no fire going back because 0 is already burned but in the directions where there is no, where the vertices are not burnt I get new information.

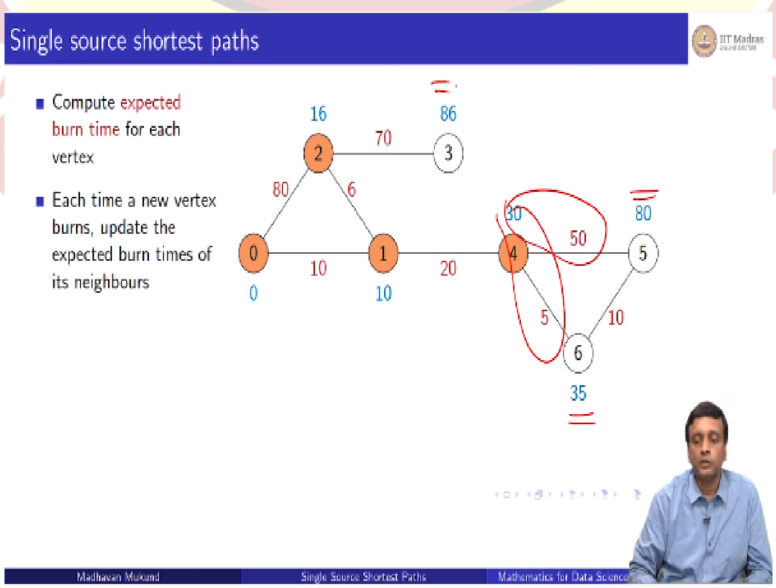
I already believe that 2 will burn at time 80 but now I can tell that it cannot stay more than time 16. And similarly, 4 which was previously not known to burn at any time at all. Now I know it definitely burned by 30. So, I can update this entry as neighbors of 1, this entry and this entry to be now 16, and 30.

(Refer Slide Time: 10:19)



Now, again I look among the unburned vertices, so these are my unburned vertices, I look among my unburnt vertices for the one which has the minimum time to burn. Now this is the absolute time from time equal to 0. It is not 16 more units of time but at t equal to 16, which is 6 units of time from now, because right now I am at unit, I am at time 10, 10 is when this happens. So, in 6 more units something is going to happen and that is going to be 2.

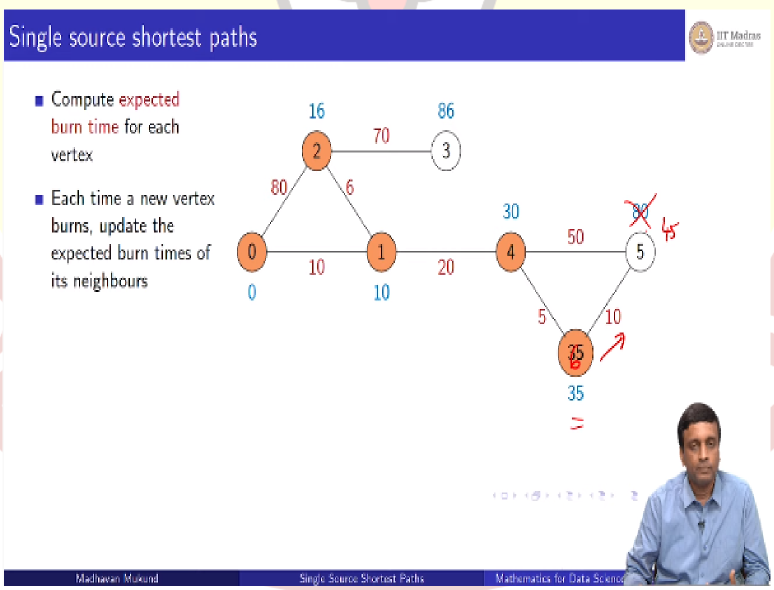
(Refer Slide Time: 10:48)



So, at times 16 2 burn, when 2 burn, I look at its neighbors, which are not burn, namely 3 and I say 3 which I never knew was going to burn at all, now I know it is going to burn at 86. But I get no new information on this side because 2 is not connected to anybody else, so I have no further information yet about 5 and 6. As far as I know, 5 and 6 are never going to burn.

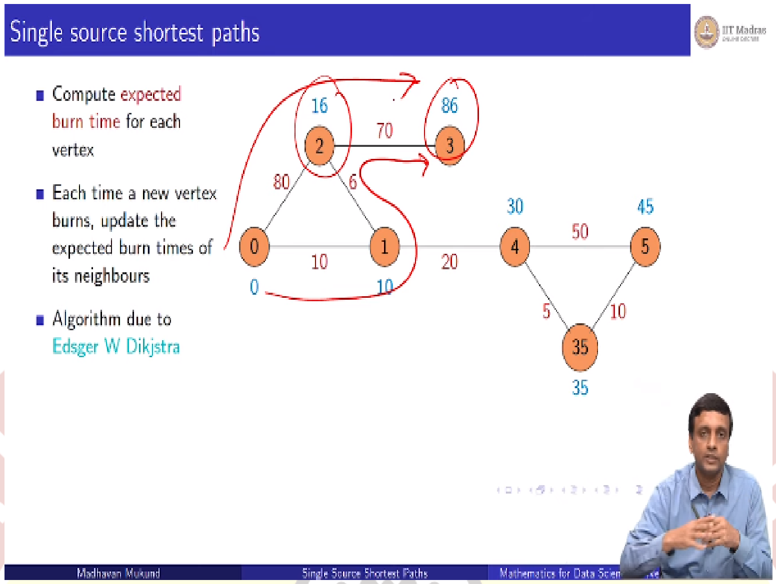
Now, again I look at those vertices which are not burned and I look for the minimum. And I find that vertex 4 is going to burn at time 30. So, I burn it, so now time is 30. So, now I have information about 5 and 6 because once I have burned 4 I have started these two new fires, which are going here. So, this tells me that vertex 6 is going to burn at time 30 plus 5 and vertex 5 is definitely going to burn by 30 plus 50, so I can update those to 80 and 35. So, this is 30 plus 50 is 80 and this is 30 plus 5 is 35. So, now again I have the same thing which are the ones who did not burn, 35, 80 and 86. So, which is the next one to burn clearly vertex 6.

(Refer Slide Time: 11:52)



So, I burned vertex 6, now having burn vertex 6 again, I have, this should been a 6 here, I do not know how it became 35, so this 6 now passes a new fire here. So, since this was a 35, I know that this is going to burn at 45, my previous information was 80, so I have to remove the 80 and make it 45, with each time I improve my estimate if I can based on new information that I get.

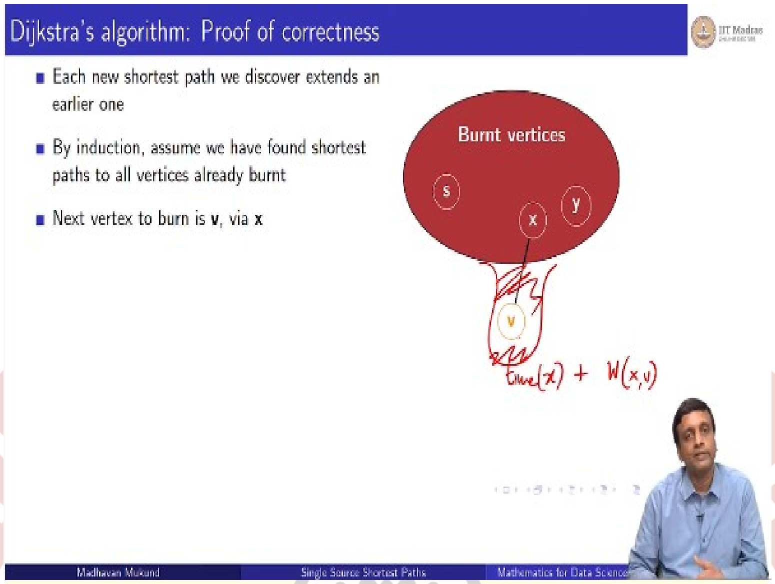
(Refer Slide Time: 12:19)



So, now I improve that 80 to 45 and having improved that 80 to 45. Now I burn it because among the things which are not burned is 60, 86 and 45. So, when I burn 5, I get no new information about 3 or anybody else, because it is not connected. So, finally, 86, 3 burns, so this is my algorithm, this algorithm although we have described it picturesquely in terms of burning vertices and all that, you can keep track of all this information in a matrix and keep track of what is burned we can keep updating this time to burn and so on by looking at the edges in my graph and finding out what are the neighbors, what is the current burning time and what is the expected new burning time and so on.

So, it is a fairly straightforward thing to do if you have the matrix and if you have this information about these burning times, we are not going to describe this algorithm in precise detail now, later on we will study it in the course on algorithms. So, this algorithm is very well known it is due to a very well known computer scientists called Edsger Dikjstra, so this is called Dikjstra's algorithm. This is the Dikjstra's algorithm for the single source shortest path.

(Refer Slide Time: 13:25)



So, why does this work, I mean why is this kind of update reasonable? So, the idea is that every time the new shortest path that we find extends an earlier shortest path, right, so if we look back, we go back to this. So, the shortest path to 86 is an extension of the shortest path to 2. So, basically the shortest path to 2 is this way. So, I get the shortest pathway here, I cannot get a shortest path to 86 which goes the other way by a longer path, so every prefix, every beginning segment of a shortest path is also a shortest path.

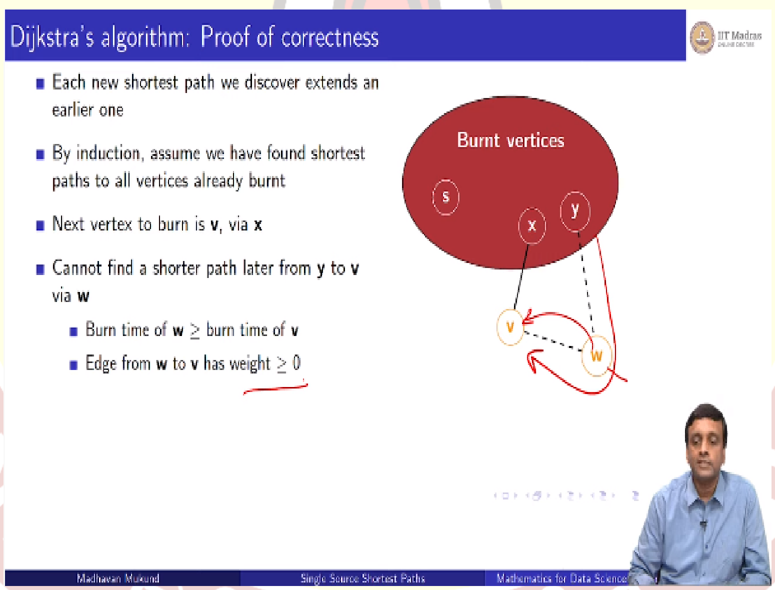
It cannot be that I find the shortest path to a vertex and then to go to its neighbor I find the shortest path with bypasses I mean find the longer path to this vertex and goes. So, every shortest path, actually extends, an earlier shortest path. So, what we are going to show as a kind of correctness proof is to assume that at every point when we have burned some vertices the numbers we have associated, the burning times we have associated are the shortest time, so by induction this is correct, because the starting point is a source vertex which we label with 0.

So, when we label the starting point with time 0, obviously that is the shortest time at which it burns because that is the shortest path from a vertex to itself is length 0. So now, at any given point we have some vertices which are burned and we have some vertices which are yet to burn and with each of them we have an estimate. So, we have some vertices in this case s, x and y, which are burnt and now suppose our next vertex to burn with the shortest burning time is v and it is connected to x. So, what we are assuming therefore is that the shortest path to x now extends by x to v to shortest path to v.

So, basically the time to v is going to be the time to x plus the weight of the edge xv, I mean that is basically what we are saying, there is no faster way to get to v than to first get to x, which is the time it takes for x to burn plus the cost of going from x to v. So, now the question is, is this going to be the final value for v, can we discover, because now we are going to burn it, so this is going to get burned. So, this burn thing is going to include and by our induction assumption once we burn a vertex, its value is correct.

So, is this value correct, that is what we want to know. So, if we choose to burn it now, are we being hasty, are we going to discover a better path to V later on, can it happen?

(Refer Slide Time: 16:03)

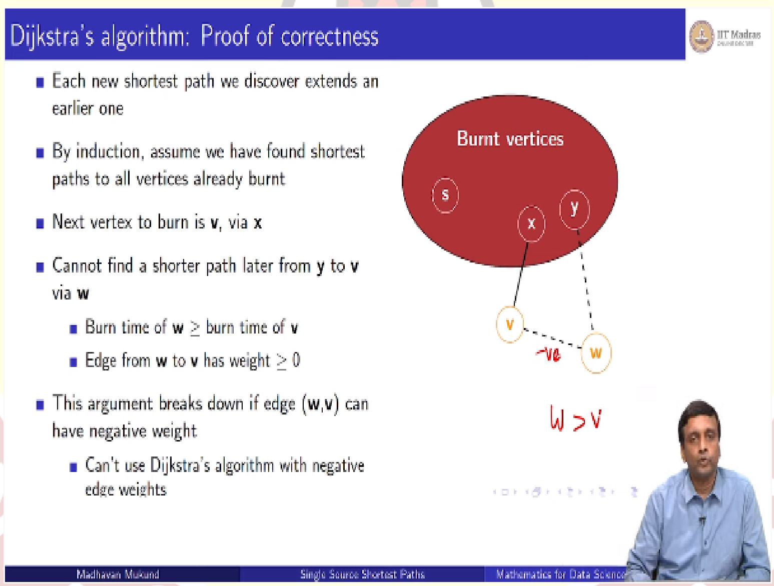


So, supposing there is a better path later on so where will that better pat come? It will come from some other vertex which is not yet burnt because among the burnt vertices we already found the best path. So, at a later stage supposing we find a path from y to v via some w, we burn a new vertex w and through that new burning vertex we somehow magically discovered that there is a shorter path to v.

Now, can this happen? So, first of all, why did we burn w after v? We burned w after v because w has a burning time which is not smaller than v, it could be exactly equal but it is greater than or equal to it. So, the time at this point is the time is already equal to or greater than v and then I have to come from here. So, I have to spend some time, because there is some cost involve with that edge, it could be 0. So, I certainly I could get the same value, maybe by some magic, the burning time of w is exactly equal to the burning time of v, it cannot be smaller because otherwise we burned w before v. So, it is at least as big, maybe it is not bigger but maybe equal to, but even if it is equal to, at best, then I can come in 0 cost from w to v because edges are not negative.

So, the edge from w to v has a weight which is bigger than equal to 0. So, this is crucial, so if we did not have this, then the decision to burn v based on its distance from x could have been premature.

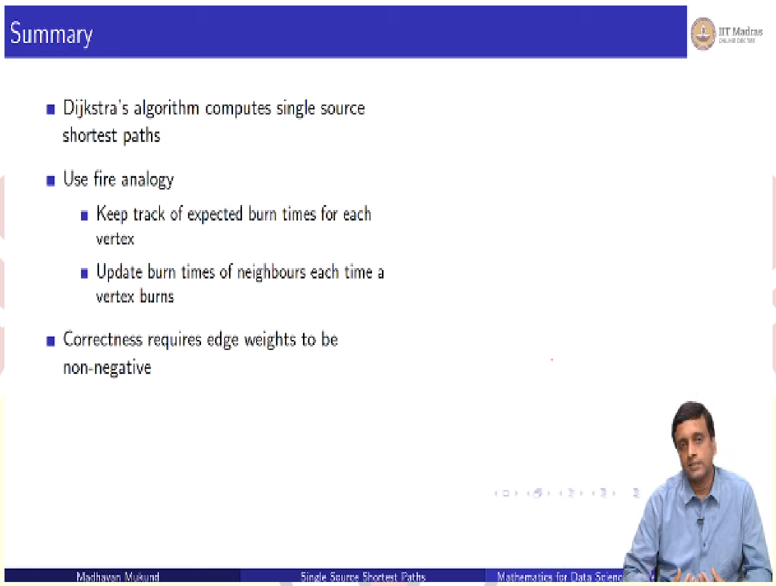
(Refer Slide Time: 17:21)



So, this tells us that we cannot use Dikjstra's algorithm, if we have negative edge weights because this could happen, I might discover a long path after I burnt a vertex which becomes shorter, because there are some negative edge weights which cancel out the initial thing, so it could be that the, this is, so w is bigger than v, but because this is negative, the cost of coming back from w to v actually drops the cost of v below the cost that I had got when I burnt it.

So, for this reason we have to assume, when we run Dikjstra algorithm that the edge weights are bigger than equal to 0 otherwise this strategy for updating is not going to always work.

(Refer Slide Time: 18:00)



So, to summarize, we have found an algorithm to compute single source shortest paths provided the edges have non negative weights. And the way to think about this algorithm is to use this fire analogy, so we set fire to every, we set fire to the initial vertex, think of it as an oil tank, and then think of the edges as pipelines. So, this oil, fire that starts at source vertex spreads at a uniform rate through these pipelines, and then we can calculate through this pipeline uniform burning the rate at which, the time at which every vertex burns, and when we can systematically keep track of this, you will get the shortest path to every vertex. So, this is Dikjstra's algorithm, which works single source shortest paths for weighted graphs with non negative edge weights.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

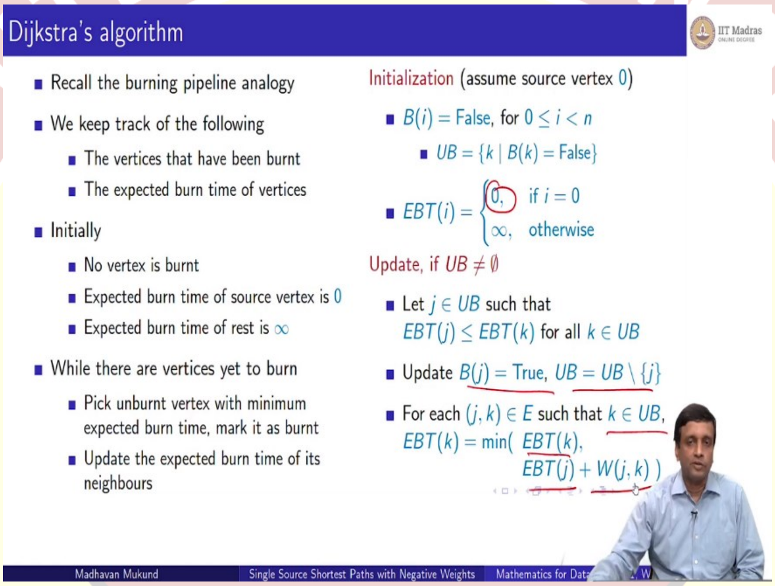
IIT Madras

ONLINE DEGREE

Mathematics for Data Sciences 1
Professor. Madhavan Mukund
Chennai Mathematical Institute
Lecture No. 12.3
Single Source Shortest Paths with Negative weights

So, we are studying weighted graphs and we started looking at the shortest path problem, in particular, the single source shortest path problem and we saw Dijkstra's algorithm, and we said that it will work if we do not have negative weights. So, let us see what happens when we have negative weights.

(Refer Slide Time: 00:28)



So, first, let us recall Dijkstra's algorithm and look at it a little more formally than we did last time. So, remember that we thought of Dijkstra's as I build them operating in this burning pipeline story. So, we had these vertices as oil depots, and we had the edges as pipelines. And if we set fire to a source vertex, then the fire spreads at a uniform pace along all the pipelines, and then we try to calculate the order in which each of the vertices will catch fire and propagate the fire further along new edges.

So, in the process of doing this as an algorithm, what we do is we keep track of which vertices have already burned, that is the vertices for which we have already computed the shortest distance. And we have an estimate about how long it is going to take to reach the others so that we have an, what we call the expected burn time, which we keep updating as we go along.

So formally, we keep track of these two things, the burn status and the expected burn time. So initially, we set the burn status to be false. So, let us call it B of j. So, for every B of i for every vertex i, B of i is initially false. And just to keep track of an auxiliary quantity, we will let UB.

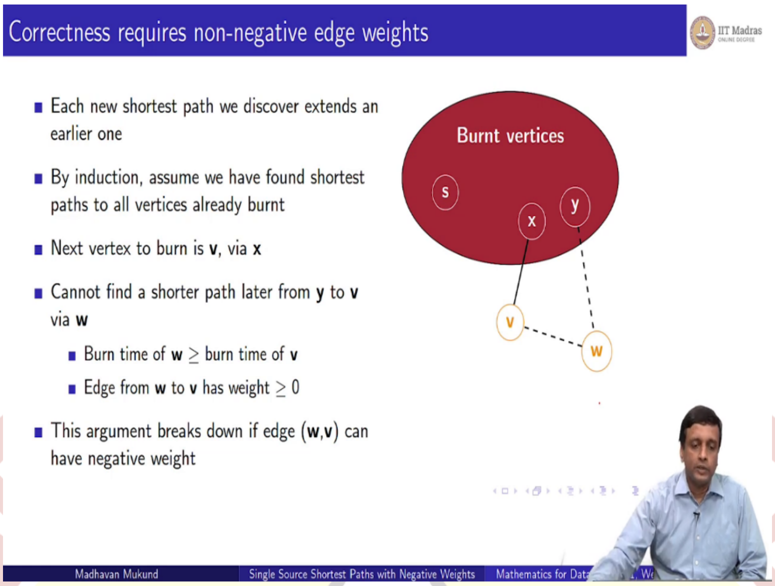
So, UB is just a set of unburned vertices here. So, UB is just stands for unburned. So UB is those k for which Bk is false.

So initially, UB is a set of all vertices. And the way we start this algorithm is we set the burning time of the source vertex which we are assuming to be 0 in this case. We assume that the burning time with the source vertex is 0 and everything else has vertex burning time infinity because we have no information. And then what we do is as long as there are unburned vertices, in this case, everything is unburned. So of course, there are unburned vertices, you pick up one of the unburned vertices, which has a minimum expected burn time.

So initially, there is only one that is the source. So, you pick up the source vertex, and you update its status, you update the status of the vertex, you pick up saying it is now burned. So, therefore, in some sense, it is expected burn time is frozen as whatever it was now, and correspondingly, the unburned vertices will reduce because this particular vertex that you have chosen to burn now is no longer unburned.

Now, more importantly, what you do is you look at every outgoing edge from this new vertex, so you just burn j. So, you look at every j k edge and check if k is unburned, that is, we still have not fixed its distance, then you update it is distance to be the minimum the distance you already know. So, this is what you already know about k plus the new information that you get, if you process this edge j k, which may or may not have been taken into account before obviously, because you have not reached this day before this. So, you look at what is the time that you burnt j and how much time it will take from j to k and this might well turn out to be smaller, so this is Dijkstra's algorithm.

(Refer Slide Time: 03:14)



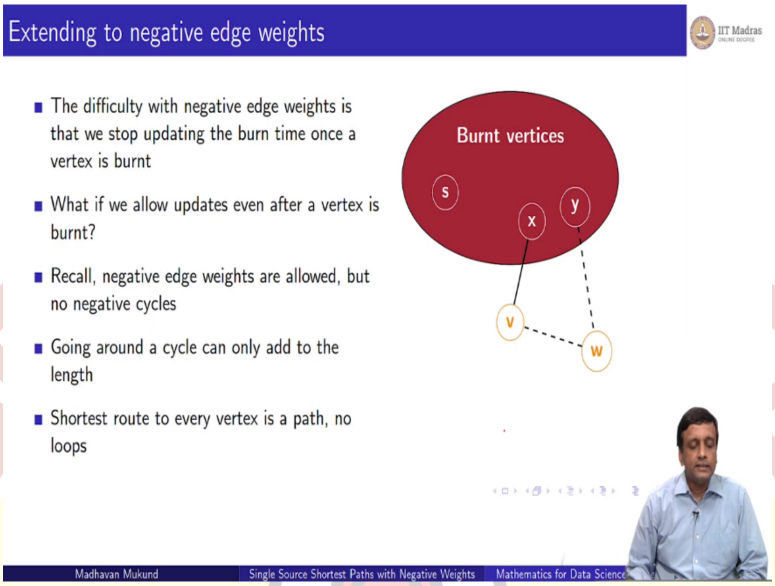
So now, we look at this correctness proof and we argued why we need this edge weights to be non-negative. So, we said that we are incrementally discovering shortest paths. So, every new shortest path extends an earlier shortest path. And inductively, the burnt vertices are those for which that distance has been computed and frozen, we are never going to update those things again. So, at last, at some point in the algorithm, we have a big set like this, so we have this big set of burnt vertices.

So, all the vertices in this set, so we have the starting vertex, which now I have called s just to denote it and then we have various vertices x, y, and all that which we have learned so far. And what Dijkstra's algorithm now says is among the remaining, look for the one with the smallest expected burn time, and it will turn out that that will be connected to some vertex in the set, so it will be connected by an edge.

So, it will be the minimum of something plus the edge from x to v. And the argument was that if we now choose to burn v and add it to the set and freeze it is time at the current time, we will not be making a mistake and that is because if we find a new path to be later, it will come through another w, but that w must also start from inside the burnt vertices. So, when I burnt v, w had a higher expected burn time. So, otherwise, I would have been w.

So, when I get here, if I look at the cost of going to w plus the cost of going from w to v it cannot be less than what I already have for v and this crucially depends on the fact that this edge from w to v is not negative, because if it was a large negative edge, then going by a w, I could suddenly save a lot of cost when coming to v, so that is the crucial thing. So, the argument does not work, so we cannot freeze the cost of vertex the first time we burn it, if we are allowed to revisit it by a negative edge later on. So, that is why Dijkstra's algorithm requires non-negative edge weights.

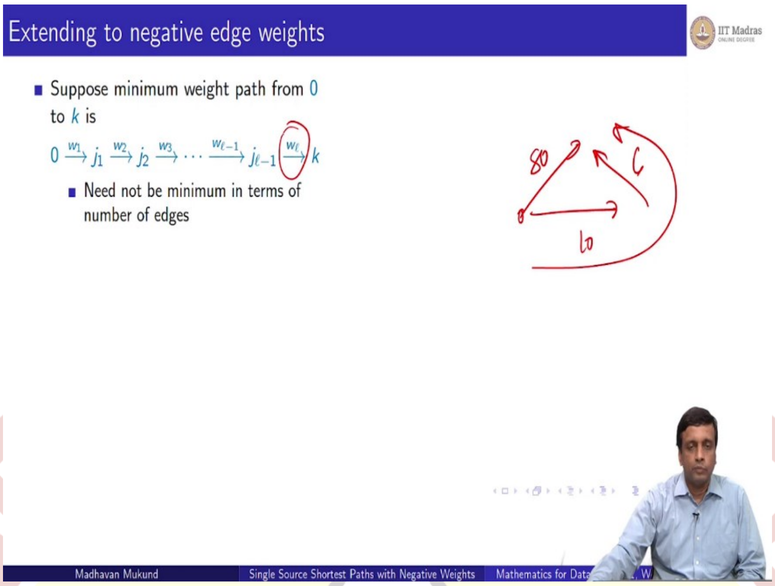
(Refer Slide Time: 05:11)



So, the difficulty is precisely this, that we stop considering updates once we burn a vertex. So, what happens if we start allowing updates even after we have burnt a vertex? So, then the notion of burning does not really make sense. So, this analogy that we have does not really make sense, but it is a plausible strategy and why is it a plausible strategy? This plausible strategy because though we are allowing negative edge weights, we do not have any negative cycles.

So, this means that if I am going from say, the starting vertex to some vertex x, it cannot help me to go through a loop because every loop is guaranteed to have a non-negative weight. Because there are no negative cycles. So, if I want the shortest path from any starting vertex, from the starting vertex to any other vertex x, I may as well assume that that path is really a path in the sense of, we have defined that is it has no loops, it has no repeated vertices.

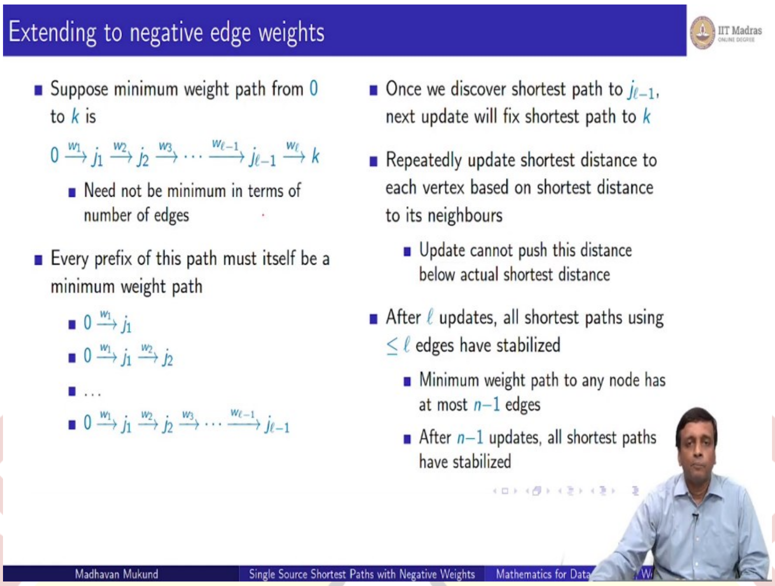
(Refer Slide Time: 06:15)



So, if it has no repeated vertices, then what we can consider is in terms of the length of the path, not in terms of the weights of the path alone, but we have the shortest path, which takes us from say 0 (())(0:27) source vertex to a vertex k. So, not only is the weight minimum, but the reason that this is the path I chose is because there are no shorter paths which have the same weight or less, so it need not be the shortest path overall, we saw examples where you could have one edge, which takes me with 80. But then if I take two edges, I might go with 10 plus 6 16.

So, the shortest path might well be a roundabout path. But what we are saying is that if there are, there is a path of length 2 which is shortest there is no path of length 1, that is what this means. So, this is the, so l, if I have to do this, I have to take these l steps in order to get to k, and there is no better way of getting to k to achieve this cost. So, this is not the minimum number of edges going from 0 to k without considering weights, but it is the minimum number of edges and the shortest weight if you consider weights.

(Refer Slide Time: 07:14)



Now here, we again go back to our old argument, which says that, okay if this is my path, then what happens when I come to j1, could I have come to j1 any better than using the weight w1? Well, if I could have come by a different route and come less than w1, then I could come to j1 and then continue with the same path to k, so if I could come to j1 earlier than I have now, then I could use that path plus the path from j1 to k and get a shorter path to k.

So, if this is the shortest path to k, then this must also be the shortest path to j1, the shortest path to j2, the shortest path to j3, and all that. And so, every prefix of this path has to itself be a shortest path. So, this gives us a starting point to think of an algorithm to deal with negative edge weights. So, in some sense, once we have updated this one, once we have found the shortest path to jl minus 1 by some algorithm, then we know that the update that we get for k is going to be a final update, there is not going to be a better one than that, because, if I keep decreasing, I can only decrease up to the shortest path, I cannot go below the shortest path.

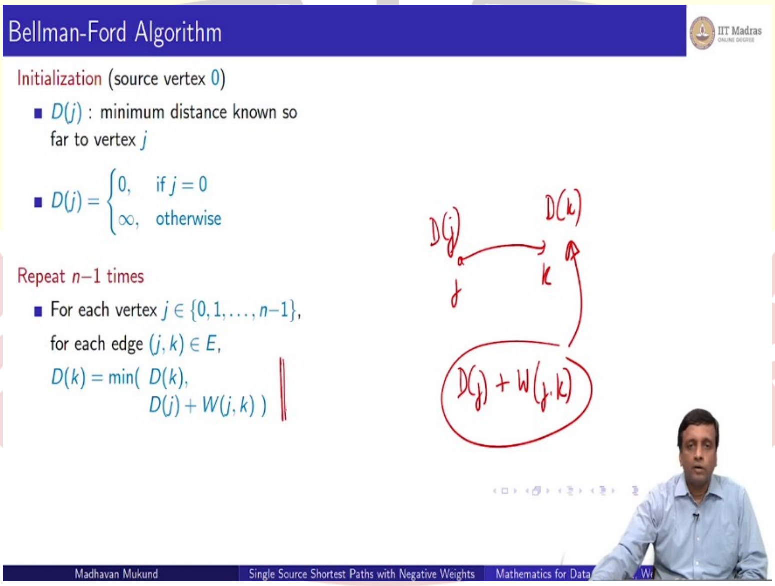
And when will I hit the shortest path when my nearest neighbour which feeds that shortest path is also frozen. So, in some sense, if my neighbour's shortest path is known, then my shortest path will be known in the next step. So, therefore, for that neighbour, that neighbours, neighbour should have been known in the previous step. So, we can try to see if we can fix these shortest paths one at a time, if I can fix j1 then I can fix j2, we can fix j2, then I can fix j3, and so on.

And once I fix jl minus 1, then I can fix k. So, what we want really is an algorithm, which tells us that after we have done l updates, so an update for us is what we did when we burnt a vertex, when we burnt a vertex we reset the burning time to be the minimum of what we already had plus the new burning time we discovered through the recently discovered vertex. So, if we can guarantee that after we have done this l times, we have made sure that there are all paths of length l are at a minimum, there are no shorter weight paths of length l, then since there are at most n minus 1 edges, then if we do this n minus 1 times there are no paths without repeating vertices and no strong notion of a path there are no paths which have more than n minus 1 edges in them because once you take the nth step, you have to repeat a vertex.

So, if we have this property that we can update and make sure that after l updates, all the paths of length l in terms of number of edges have achieved their minimum that is there are no shorter ways to go l edges or less, you have to take maybe one more edge and take a negative edge that is different, but you cannot get that l edges. This is then we can guarantee that we can see sort of first find all shortest paths.

So, this is like a combination of breadth-first search which finds shortest paths by length of path, and our weighted thing which finds it by length of weight. So, what we are saying is that up to this path length in terms of edges, there are no shorter paths in terms of weight. And this is the kind of property that we want.

(Refer Slide Time: 10:14)

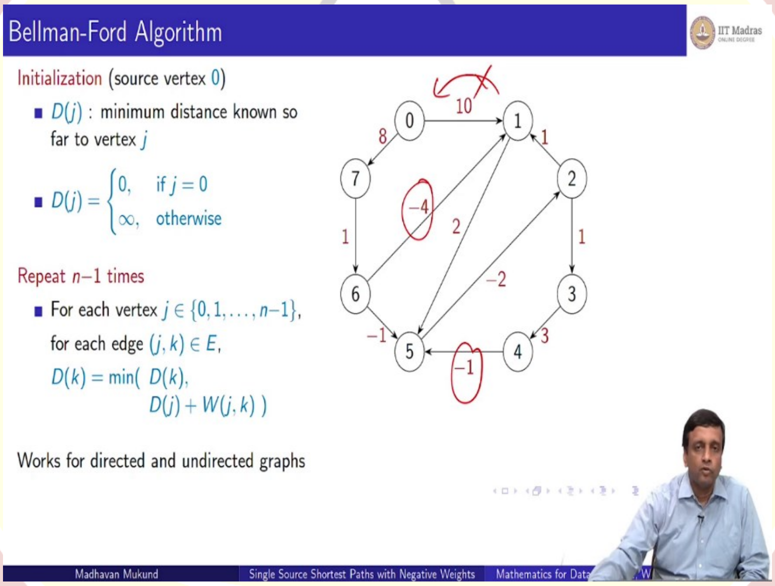


So, this algorithm is called the Bellman Ford Algorithm. So, it is a much simpler algorithm, in some sense, than Dijkstra's to think of, although it requires a little bit of understanding like we just did to see why it works. So, all we do is we like Dijkstra, this expected burn time, so we keep track of the distance to every vertex as far as we know so far. So initially, the distance to the source vertex, which again, we assume is 0, the source vertex is 0, and its distance is 0, and everybody else distance is assumed to be infinity.

And now comes the update. What Bellman Ford does is just n minus 1 times it just blindly updates everything. So, it takes every edge and it looks at the starting point at the edge and the ending point of the edge. So, I have an edge, which goes from say, some j to some k. So, there is currently at this iteration, there is some distance that I have associated with j and then I have associated some distance with k.

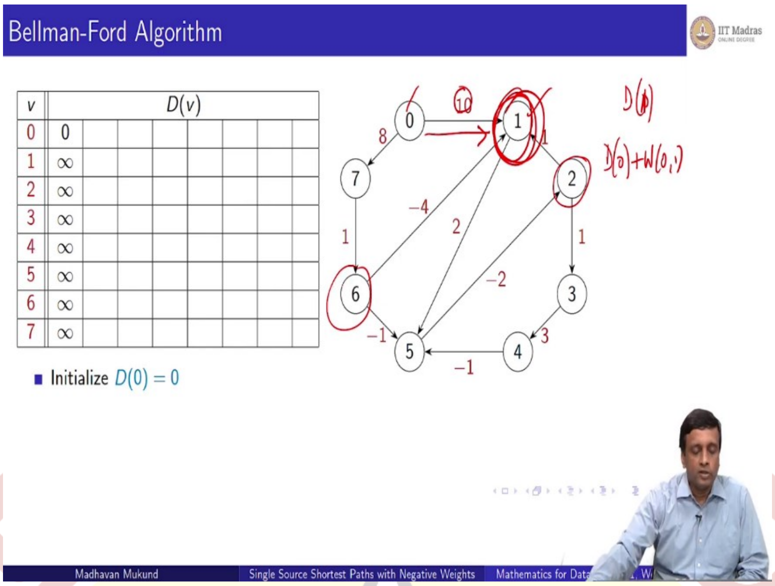
And I have what I would get if I take this edge and append it to j. So, this is a candidate to replace D of k. So that is what we do, we just check for every k, whether the weight that the distance that we are currently assumed for k, is it smaller than the distance that I would get if I take one of my neighbours and add that edge weight from that neighbour to that distance. So, I just blindly do this n minus 1 times and the claim is that this will give you the shortest path.

(Refer Slide Time: 11:48)



So, this works for both directed and undirected graphs. So, the example we did for Dijkstra's algorithm was for undirected graphs, but you could as well do it for directed graphs. Because anyway, we are following edges in one direction only when we compute the shortest thing. So, let us look at this example, so this is a directed graph, it has some negative edge weights, like minus 4, minus 1, and there are arrows in this, so there are some so you can go for example, from 0 to 1, but you cannot come back from 1 to 0, and so on. So, let us see how this Bellman Ford algorithm would work on this.

(Refer Slide Time: 12:18)

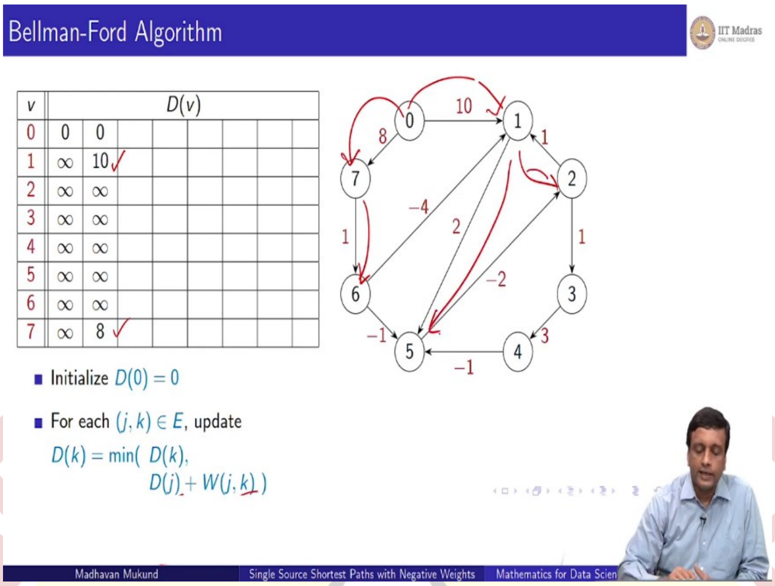


So, what we do is we keep recomputing this distance D of v. So, remember that D of v is the best distance I know of right now for vertex v. So, in this case, there are 8 vertices from 0 to 7. And we are going to iterate this thing after initialization, n minus 1 times. So, there are 8 vertices, so n is 8. So, we are going to run this algorithm, this iteration 7 times, and then 7 times, it should hopefully give us the shortest path from 0 to every other vertex.

So, we initialize it by setting the distance of 0 to the vertex 0 to 0, and everything else to infinity and this is our initialization for Bellman Ford. So initially, we know nothing about how to get to any other vertex. And now we do this update. So, now we look at every vertex or we look at every edge is how the Bellman Ford algorithm says, we look at this edge and we say, what do I know about the starting point plus this weight versus the ending point?

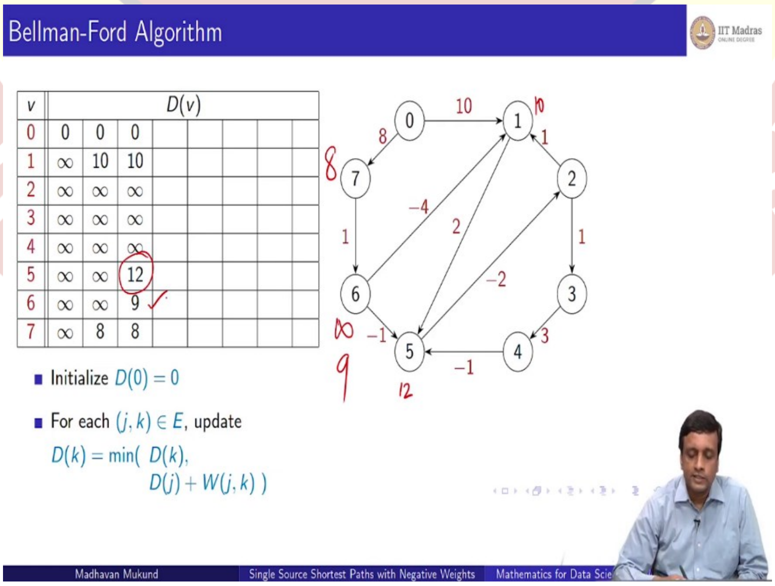
So, the update is compare D of 0 to D of 0, D of 1 to D of 0 plus the weight of 0,1. So, what should I put here, is a question. So, should I leave it as what it is, or should I update it by some new information that I have got about the edge coming into it. Of course, I could do it for D of 2 also, but then I know that for D of 2 and D of 6, nothing has happened because everything is infinity, so is really D of 0, which carries some importance at this stage.

(Refer Slide Time: 13:42)



So, if we do this, we find that from 0 I can get to 1, and from 0 I can get to 7. And therefore, the entries for 1 and 7, get updated from infinity, which is what I knew before to 0 plus 10, in the case of 1, and 0 plus 8 in the case of 7. So, I updated to the time to reach 0 plus the weight of the edge from 0 to that vertex and everything else is infinity because I cannot reach it from 0 at this point. But now, I have some information about 0, 1, and 7 so, in the next step, I can look for any vertex, which is either connected to 7 or connected to 1. So, not that one, but say this one. So, 1 is connected to 5.

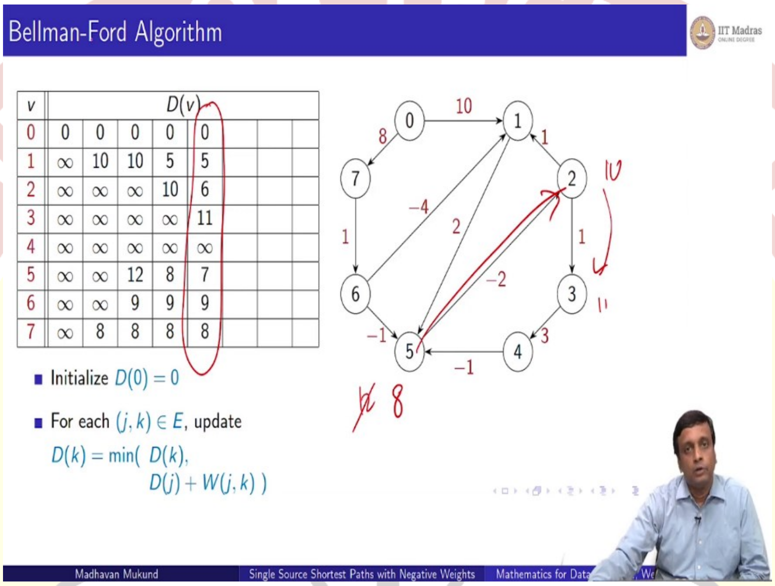
(Refer Slide Time: 14:25)



So, in the next step, what happens is that because 1 is connected to 5 and I know that I can reach 1 at time 10, then in 10 plus 2 I can reach 5 in time 12 earlier I believe it was infinity.

So, I can replace it by 12. In the same way, if I look at 6 for instance, earlier I thought it was infinity but now I know that I can reach 7 and time 8, and 7 plus 1 is 9, 8 plus 1 is 9. So, therefore, I can reach vertex 6 in time 9. So, these two things which are connected to the vertices are recently burned, get updated. So, we keep doing this. So, now we have burned, 6, and 5, not burned but we have updated 6 and 5 in addition to 0, 1, and 7. So now, we will find new paths because 6 also has outgoing edges.

(Refer Slide Time: 15:13)



So, in particular now, what we will find is that there is a strange phenomenon that we have discovered, which is that if I come from 6, so remember that 6 was, had been assigned 9 before, so, right now these are the numbers that we have everything else is infinity. So, now, if I am at 9, and if I take this negative edge, then I can come from 0 to 1 at cost 5. So, instead of going directly in cost 10, which is what I had earlier assumed, I could take this roundabout route, and I could do 8 plus 1 minus 4, and come there and 5.

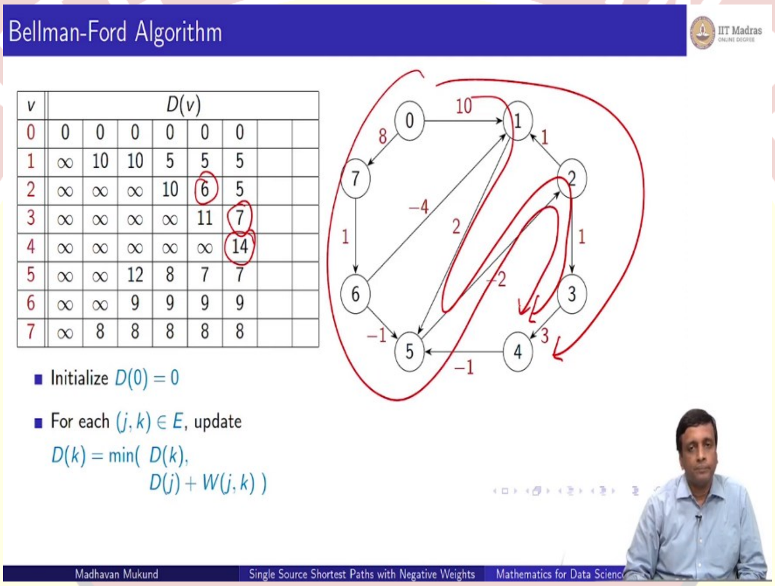
So, this is the kind of update that Dijkstra's algorithm would not have discovered it because it would have really frozen that thing at some point, saying that it is already given, and therefore, I will not update. So, this 10 becomes 5, what about 2? Well, now 2, I can reach from 5. So, it is 12 minus 2, so this becomes 10. And 5 itself, now is interesting, because earlier, I had to come this way. And that was costing me 12. But now because I can come from 6 directly, I can come to 6 and 9, and then 6 to 5 will give me so this has become 8.

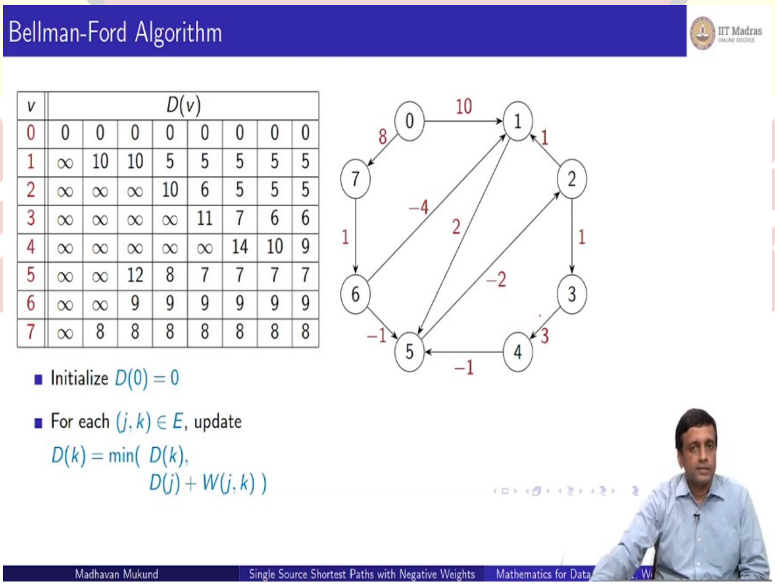
So, in this way, we keep updating, so now after this, now that I know 2, I can even update 3 because 3 is reachable from 2. So, if I can reach 2 in time 10, I can reach 3 in time 11. But of course, 2 itself has now got a better route. Because having come to 5 in time, 8 now.

Remember, it was 12 and then it became 8. Now I can go from 5 to 2 in times 6. So, each time I am looking at the previous row, so the fact that 2 gets updated from 10 to 6 now does not yet reflect in the fact that 3 should be updated, so 3 is updated from infinity to 11 because I knew that 2 was 10 before I did.

So, all these updates are happening at this time based on the previous times information. So, I am not doing it in sequence in that sense, so though I calculate that the distance to 2 is 6, I do not use it to calculate the distance to 3 is 7 yet, I will do it in the next round.

(Refer Slide Time: 17:28)

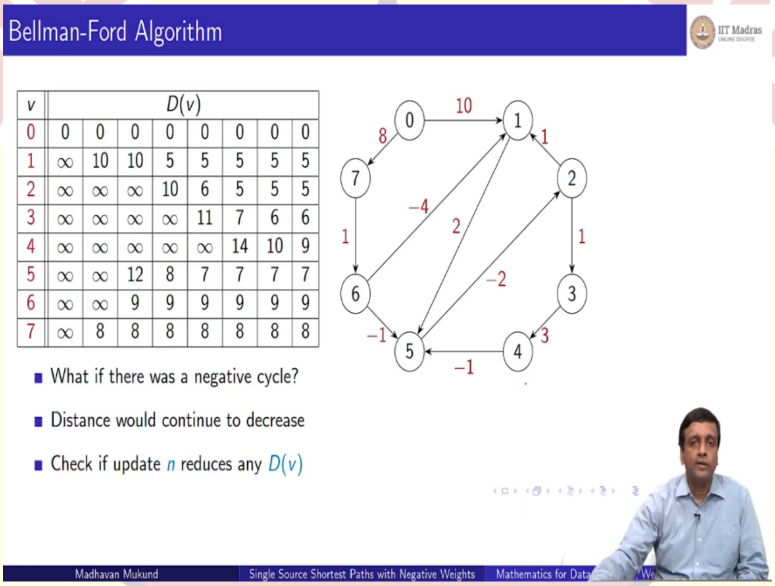




So, in the next round, I discovered that 2 was frozen at 6, and therefore now D has become 7. And finally, I have also found something that reaches vertex 4 because now I have got this path which goes this way or there are other paths also. So, we have this path also, which goes this way, and so on, so we have many paths which come to 4, but I only now reach that and I calculate 14. So, I keep iterating this, and I get slightly better paths everywhere.

And finally, after I have done this 7 times, I have discovered a stable thing and you can calculate that if you do this one more time you should get no updates. There should be no shorter paths if there are no negative cycles. So, this is how the Bellman Ford algorithm works. It just keeps updating every vertex every time and you do it a fixed number of times, which is the number of vertices in your graph, minus 1. And once you have done that, you are guaranteed that all the paths have stabilized.

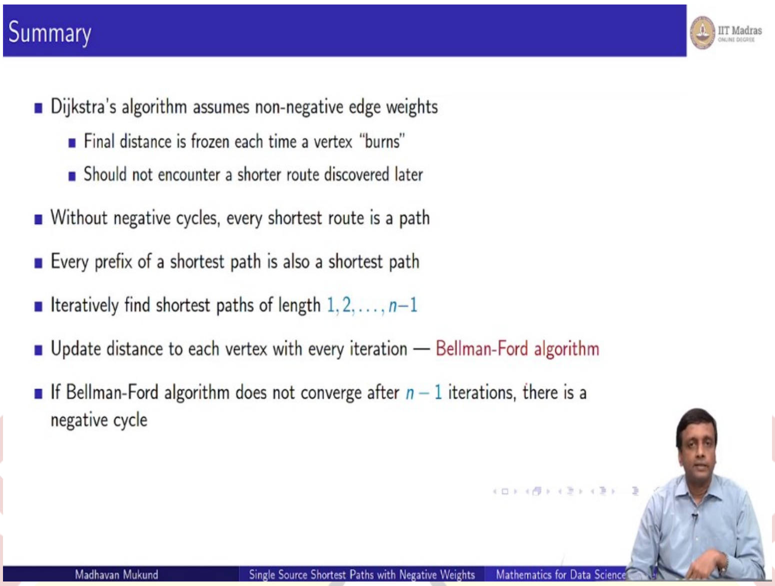
(Refer Slide Time: 18:22)



So, what would happen if there was a negative cycle? Well, the path would not stabilize, there would be a way to take a path longer than n, n minus 1 edges go beyond that and go around the cycle and get still shorter. So, if I iterate this one more time, and the distances decrease, then I know that there is something wrong. So, this is one way. So, either you can assume there are negative cycles and keep running it or you can run it.

And then when you come to nth iteration, which you normally should not need, you need to stop with n minus 1, but you can run it one extra iteration. And see if you get a decrease in the distance. And if you get a decrease in the distance, that means there was a negative cycle. So, check that the nth update should not reduce any D of v. If it does not reduce once, then after that, once the D of v is stabilized, there is going to be no update because the previous update is just going to propagate. So, once I get the column repeating, there is going to be no change further on. So, if the column changes at the nth step, then you know that you had a negative cycle.

(Refer Slide Time: 19:19)



So, we saw Dijkstra's algorithm and Dijkstra's algorithm assumed non-negative weights and the reason that we needed that property was because of this strategy we used to freeze the distance to burn vertices once they were burned. So, we never looked for updates to that. So, we should not have found any new updates through negative edges. Otherwise, that strategy is not correct.

But assuming that we do not have negative cycles, remember if we have negative cycles, the notion of a shortest path is not defined, because you can go round and round the cycle and you can make a shortest path as short as you want. So, the shortest path is defined only when there are no negative cycles, even if there are negative edge weights. And in such a case, what we just said is that the shortest path is a path, it cannot involve a loop and every prefix of that shortest path is also a shortest path.

And you can use this in this Bellman Ford algorithm to iteratively find the longest, the shortest paths of length 1, of length 2, length 3, and so on. And after n minus 1 iteration, you have automatically found shortest paths to everything. So, in a way, Bellman Ford is a much simpler algorithm to think of, it is just a blind iteration, which is n minus 1 times you keep updating it, you do not have to keep track of anything that was burned and you do not have to keep track of expected burn time separately and all these things. You just keep updating every time you make an update, you look at all the neighbours and update them again.

And the property of this, this fact that the shortest paths are monotonic in the sense that every path is an extension of a shortest path guarantees that after n minus 1 steps all these updates will converge unless you have a negative cycle. So, you can also use this algorithm to find negative cycles in that sense, if you go through this whole process and you do it one more time and you find a decrease, then you have a negative cycle.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

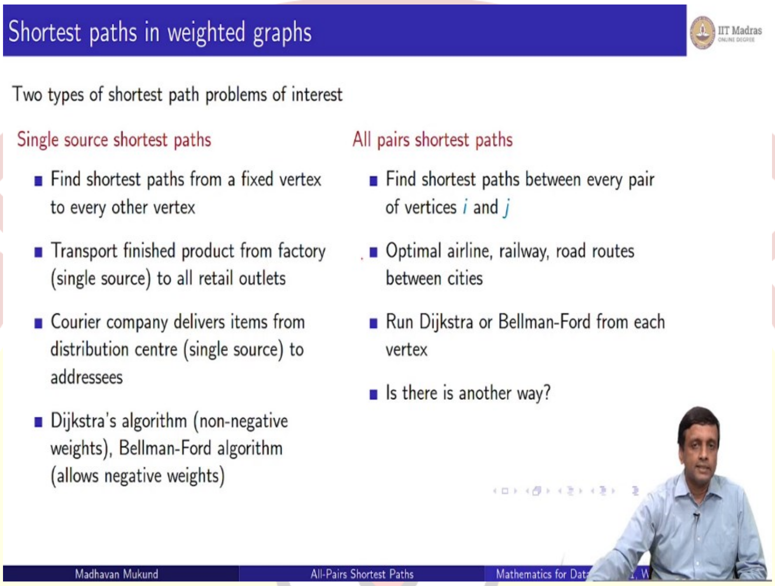
IIT Madras

ONLINE DEGREE

Mathematics for Data Sciences 1
Professor. Madhavan Mukund
Chennai Mathematical Institute
Lecture No. 12.4
All-Pairs Shortest Paths

So, we are looking at shortest paths and weighted graphs.

(Refer Slide Time: 0:17)



And we said that there are two types of problems that we might look at in terms of shortest paths, the one that we have looked at already is the one which is called the single-source shortest path problem, which asks you to find the shortest path from this some designated starting vertex, source vertex to every other vertex and we imagined two scenarios where this might be useful.

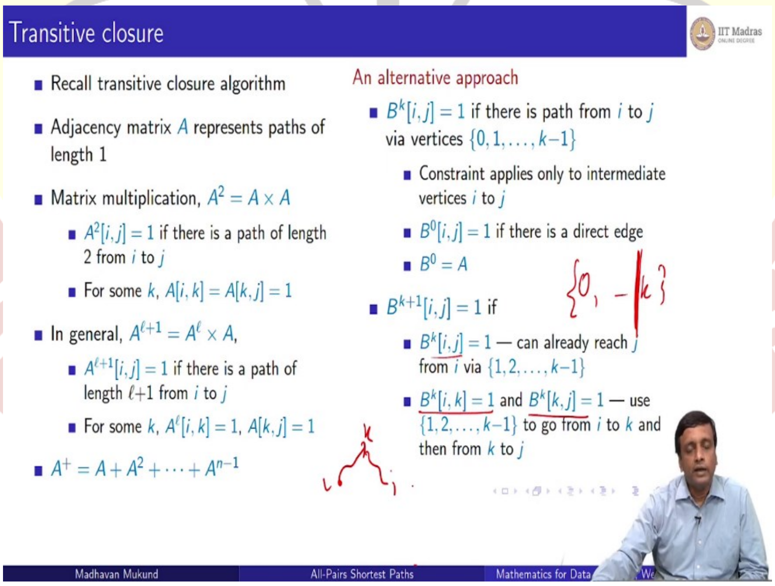
So, if you have a kind of factory, which is shipping out finished products to retail outlets, or if you have a courier company, which has got a centralized delivery facility from where they have to ship out their things for delivery, both of these would l[i,k]e to know the shortest route from wherever they are shipping out things, either the finished product or the courier deliveries to all the other locations in their transportation graph.

On the other hand, another natural problem is to find the shortest distance between every pair of vertices, and this we said would be very natural if you are running something l[i,k]e a travel site, somebody wants to book a ticket from somewhere to somewhere else, you should have the information about the shortest route from this starting point to the ending point, regardless of which two points they are.

So, you cannot say I am only booking tickets from this city and not from another city. So, that does not make any sense. So, if you have a travel booking site, it should be able to give you optimum routes in terms of cost, or time, or distance, or whatever, from anywhere to anywhere. So, we have seen two algorithms with a single source shortest path problem. Dijkstra's algorithm which works if there are no negative weights, and the Bellman-Ford algorithm, which works even if there are negative weights. But of course, remember that negative cycles are always forbidden because with negative cycles, the whole thing does not make sense at all, the idea of a shortest path is meaningless.

So, how do we find all pair shortest paths? So, we can find the shortest path from a single vertex to every other vertex. So, we can just iterate this by starting it from every vertex. So, I started from 0, I get all paths from 0, I go to 1, and I started from 1 and I get all paths from 1. If I do this n - 1 times, I will have the shortest path from every starting vertex to every ending vertex. So, this is one way we could do this is just to repeatedly run Dijkstra's or Bellman-Ford, across all starting points. But what we are trying to see is whether there is another way to do this, which does not involve this kind of going back and restarting the calculation. So, that is what we will look at in this lecture.

(Refer Slide Time: 02:28)



So, we will begin with something that we saw earlier, namely, the transitive closure. So, remember that the transitive closure, we discussed in the context of unweighted graphs and we said the transitive closure, is the reachability calculation. So, we wanted to find out just l[i,k]e all-pairs shortest paths, we wanted to find out all pairs reachability given i and j, can I get from i to j in this graph? Is there a path from i to j. So, again, there, we said that we could have run

BFS or DFS from every starting vertex and got it. But instead, we ran this other algorithm, which uses this matrix approach.

So, we started with this adjacency matrix and we said that an adjacency matrix represents implicitly edges, which are paths of length 1. And then we wanted to use this matrix and the operation of matrix multiplication to bootstrap this from length of 1, paths of length 1 to paths of length 2, and so on. So, we said a path of length 2 decomposes are two paths of length 1. So, if I want to find out the path of length 2 from i to j, I just have to check for some intermediate k, with that I can go from i to k in one step, and then k to j and another step.

And we can generalize this and we discussed why this is matrix multiplication. We can generalize this to l steps. If I know how to get an l steps from i to j, then I know how to get an l + 1 steps from i to j, because I go to an intermediate k and l steps, and then from k to j in one step. So, by decomposing l + 1 step as l steps followed by one step, I can again do a matrix multiplication and do A^l times A is equal to A^{l+1}.

And finally, in the transitive closure, we want to look for a path of any length. But paths remember, are always bounded by length n - 1 because there is no need to go through a loop to read something. So, it is sufficient to calculate path length 1 to n - 1, and then take the sum, the sum, if you remember, was the OR, we said that this is the OR of so if there is a 1, if A[i,j] is 1 in any of these matrices, then there is a path of length 1,2,3, or up to n - 1 and therefore A^+[i,j] the transitive closure reports one.

So, we are going to slightly reformulate this calculation of transitive closure, so we are going to do so. This is a kind of inductive definition, so, we first find transitive closure by first finding paths of length 1, paths of length 2, and so on and then we take the cumulative effect of all this by doing the + at the end. Now, we are going to do another inductive thing but we are going to use a slightly different way of calculating simpler paths and complicated paths not in terms of path links, but which vertices we will go through.

So, we will just to distinguish it from the earlier one, let us call it B. So, earlier we had A sub A to the K. If I said A^k[i,j]. So, this was referring to paths of length k, it says that there is a path of length k from i to j. So, here I have something else, B, k[i,j], does not refer to the length, it refers to which vertices I am allowed to visit on the way from i to j, and not obliged to visit all of them. But I cannot visit anything outside the set. So, if I say k as my superscript, it says, I am only allowed to go through vertices, 0 to k - 1, I am not allowed to visit any other vertex going from i to j.

So, if I have this constraint, can I reach i to j. So, it could be a direct thing in which I do not go to anything, it could be 3 of these vertices, it could be 2 of these vertices, it could be any number, but it cannot be outside the set. Now, remember, it is going to be a path. So, I never want to visit a vertex in the set more than once. But that does not matter so much, instead of telling me that I can get from i to j without going outside the set 0 to k - 1.

And this constraint, of course, applies only to the vertices a visit in between, so i and j need not be between 0 and k - 1, i could be anything j could be anything. So, I am just saying that when I leave i and before I reach j in between I should not see anything, which is outside the set 0 to k - 1. So, in this setting, what would B of 0 mean? So, B of 0 by this definition, would be some 0 up to - 1, because k is 0. So, this is saying that I am not allowed to visit any vertex in between because the set of vertices I visit in between can have at most number - 1, but I know that my vertices start with 0, so there is no vertex in the set, which starts between 0 and ends before - 1.

So, therefore, B^0[i,j] says that I cannot have any intermediate vertex because no intermediate vertex can satisfy this constraint, that its index is - 1 or smaller. So, therefore, B 0 of[i,j] precisely captures the fact that I can go from i to j without visiting an intermediate vertex means there is an edge from i to j. So, B 0 is just my adjacency matrix A. So, just l[i,k]e in the earlier case, we started off with paths of length 1, and we said paths of length 1 are precisely our adjacency matrix A. Here, we are saying paths which do not pass through any intermediate vertices are exactly captured by an adjacency matrix A. So, the starting point of this induction is A in both cases.

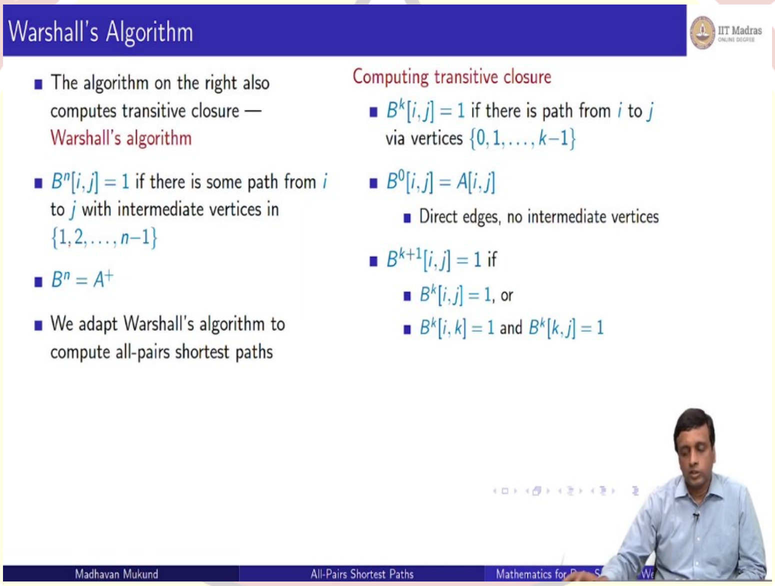
So, how would we proceed? Well, I want l[i,k]e before, I want to calculate what would be B k + 1. So, B k + 1 says that I am allowed to use 0 to k. So, if I am allowed to use 0 to k from i to j, then there are two possibilities. One is that I do not need this k at all, I could already do it without 0 to k. So, that means that B^k[i,j] is already 1, it means that without going to k, just staying within 0 to k - 1, I have a possibility of going from i to j or I need to use k.

But if I need to use k, then remember now we use this property that we implicit, that we are going to visit k only once there is no point in visiting k multiple times. So, if I need to use k, I need to get from i to k. And I need to get from k to j. But if I am using up k and going from i to k, then I cannot be using k in between, I am using k only once overall. So, going from i to k, I do not see k, going from k to j again, I do not see k. So, I can find out whether I can go from i to k using only the vertices 0 to k - 1.

And similarly, with that, I can go from k to j using only the vertices 0 to k - 1. And then I can pick these two paths. So, I have a path from i to k and I have another path from k to j. And now together, this gives me a new path from i to j, which visits k, I am forced to go to k. But in between, I do not do anything outside 0 to k - 1 except when I hit K. So, this is my condition now. So, I say that B $^{k+1}$[i,j] is 1 either if it is already 1 in Bk.

So, BK[i,j] is 1 meaning I do not need K, or I go to k explicitly, in which case I go from i to k in Bk - in k - 1, using k - 1 vertices, vertices up to k - 1. And then I go from k to j using vertices up to k - 1. So, this gives me the inductive calculation and this is a different way of calculating the transitive closure.

(Refer Slide Time: 09:23)



So, this particular algorithm is the more standard algorithm that you see in books and this is called Warshall's algorithm. So, this is Warshall's algorithm. So, we use the other algorithm, because it is nicer in terms of describing why it is matrix multiplication. So, this requires a little more work to put it in that framework of matrix multiplication. But essentially, it is capturing this inductive property of finding more and more complicated paths.

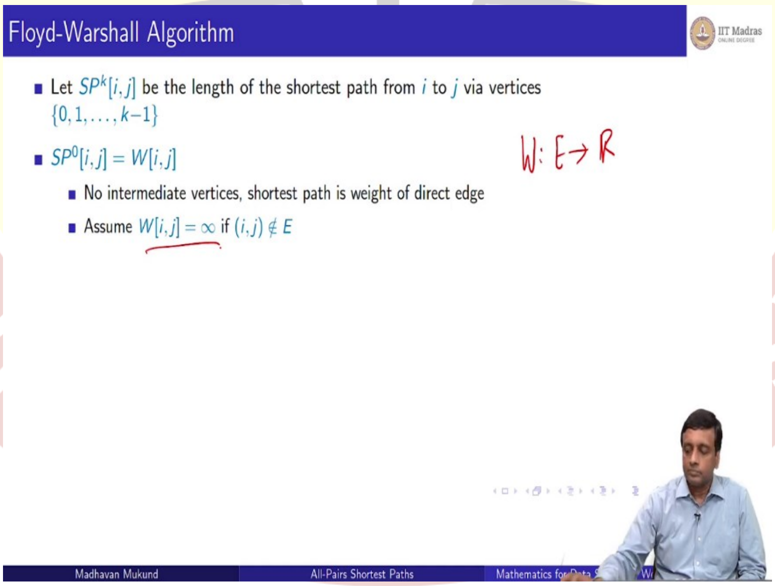
So, here the complication is a number of which are the different vertices you have to see in between rather than the total number. The total number is what the other one is talking about the length of the path. Here, is restricting you to some subset of vertices and that subset keeps growing. Eventually, you can use all the vertices. So, this is the thing around the formula. Now, so B 0 of[i,j] is just A[i,j]. And B $^{k+1}$[i,j] is 1 if either B^k[i,j] is 1, so we already can get from i to j without using the K'th vertex, or Bk [i,k] is 1 and B^k [k,j] is 1.

So, this should look very familiar because it looks very similar to the other one. This is the interpretation of this superscript attached to the matrix is different. So, in this now, when do we stop? Well, when the superscript becomes n, it says I am allowed to use any vertex from 0 to n - 1, it should be I am allowed to use any vertex from 0 to n - 1 that means I am allowed to use any vertex at all. That means there is no constraint so, I have some path of the other.

So, if I do this up to B^n, so in the earlier case also added it up to A^1 to A^{n-1}. So, here if I do it up to Bn, starting from B 0, then I am done. So, Bn is the same as A +. And now what we are going to do is we are going to see that this shortest path algorithm that we have not the shortest path this transitive closure algorithm that we have due to Warshall's is very easy to extend to the shortest path problem for all pairs.

So, remember, the transitive closure is the equivalent of all pairs reachability. Dijkstra's algorithm is the equivalent of BFS, except that we have weights. So, weighted reachability is Dijkstra's. Now, weighted all pairs reachability is all pair shortest path. So, we are going from transitive closure, which is unweighted all pair reachability to weighted all pairs reachability. So, with weights, we are asking what are the shortest paths.

(Refer Slide Time: 11:39)

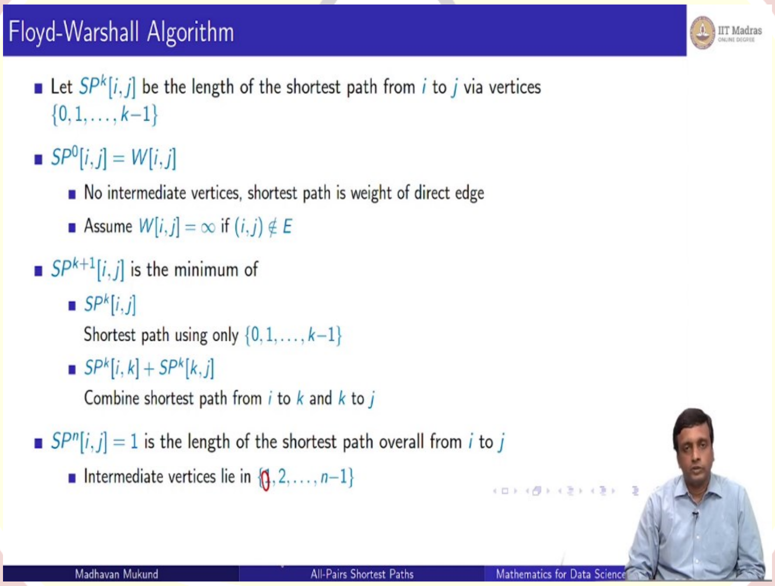


So, we will use a similar notation. So, let us call SP for shortest path. So, shortest path with superscript k between i and j is the length of the shortest path provided I stay within vertices 0 to k - 1 when going from i to j. So, that is what this means. So, earlier, it was saying that there is a path. Now, I am saying among the paths that are there, what is the shortest length, I am keeping track of the shortest length, not just the fact that there is a path but there is a path of minimum length and keeping track of that minimum length.

So, what is the base case? Well, if there is no possibility of going through an intermediate vertex, then we saw that $B^0[i,j]$ was just A[i,j]. In this case, I do not want the edge or not edge, I want to know the weight of the edge. So, SP^0 [i,j] is W[i,j]. Remember that W is our weight matrix. So, W tells us for each edge, what is the value? So, W[i,j] is the weight function. So, W[i,j] tells us how much cost is there with this edge.

Now, of course, we have to take care of the fact that there is no cost and no edge carefully. So, what we will do is we will assume that when we have no edge the weight is somehow assigned to a very large number in particular, we can treat it mathematically we can treat it to be infinity. So, if there is no edge, SP ^{0}[i,j] will report infinity, if there is an edge, it will report the weight of the edge.

(Refer Slide Time: 13:06)

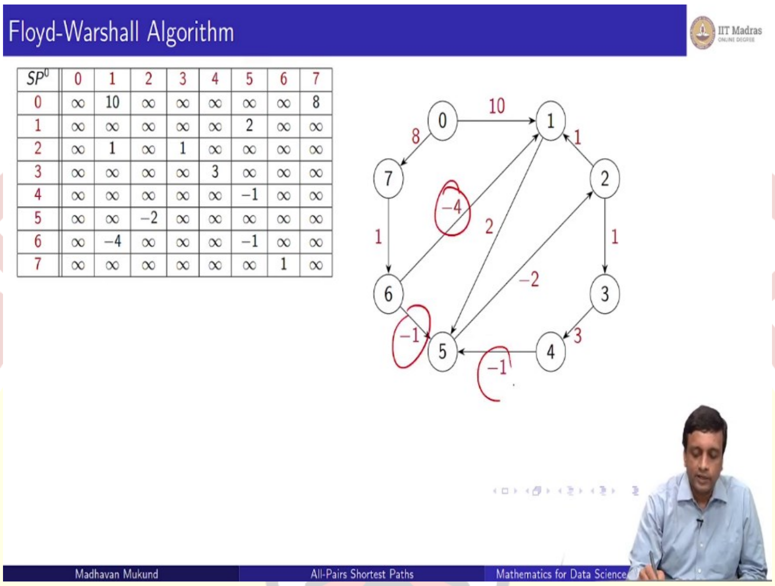


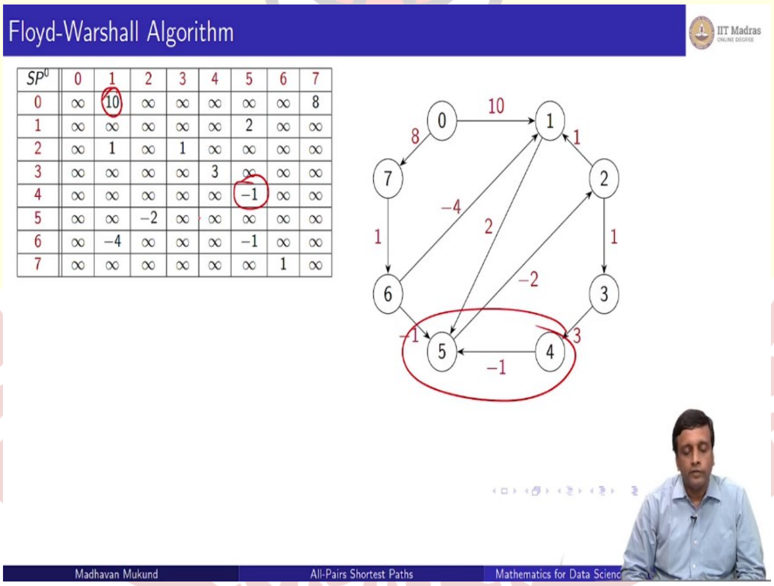
Now, the update rule is slightly different from the transitive closure rule. So, this is the slide that you need. So, either I can go from i to j without using k. So, SP $^{k+1}$[i,j] is either SP ^{k}[i,j], or I go using k, and then I have to check the cost. Now earlier, I just want to check whether this was there, or that was there, I just needed to do the logical or either I can go without k or I can go with k and I take the OR.

Here, I have to say how much does it cost me to go without k? And how much can I gain by going with k? So, I either take the cost of going with k which is the shortest path to k from i followed by the shortest path from k to j, or I take the shortest path without going through vertex k at all and I will take the minimum of these two. So, this is my update rule for the shortest path matrix.

And as usual, so this again, this same typo, but the shortest path matrix, if I now calculate up to the nth step, then it will tell me the shortest path among all paths, the length of the shortest path among all paths that go through any intermediate vertex from 0 to n - 1 and that covers all possible intermediate vertices. So, this will be the overall shortest path from i to j.

(Refer Slide Time: 14:21)

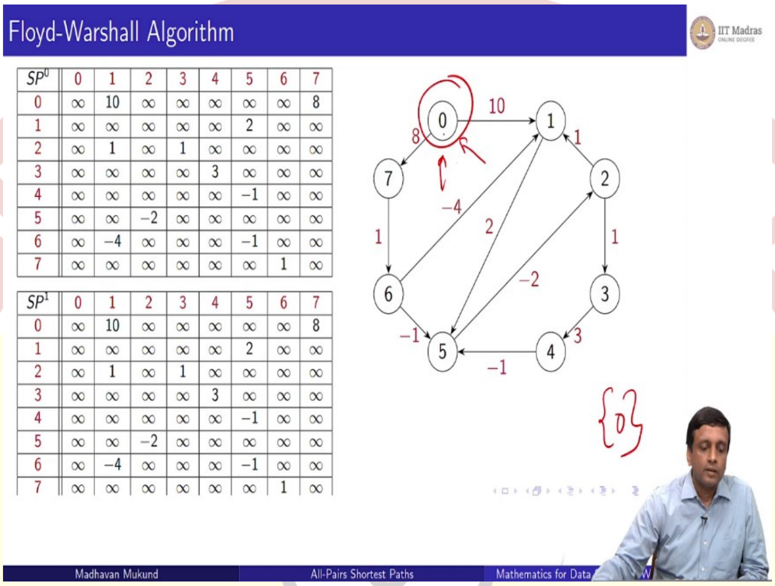




So, this particular algorithm is called the Floyd-Warshall algorithm. So, the original algorithm for transitive closure is due to Warshall and Floyd is the person who adopted this algorithm for shortest paths, all pair shortest paths, so jointly it is called the Floyd-Warshall algorithm. So, let us go back to this graph, which we have seen before. So, this is a graph with directions and with negative vertices. So, I have these negative edges. So, I have some negative edges here and there.

So, now, SP 0 is the adjacency matrix of this graph, is A[i,j] where I replace every entry the weight. So, if I look at 0 to 10, 0 to 1, it has weighed 10. For instance, if I look at 4 to 5 for instance, it has weight - 1, and so on. So, this is just the adjacency matrix for this graph, where every entry[i,j] represents the weight and if there is no edge, then it is infinity. So, we have a lot of infinity entries, we have very few entries, which are not infinity.

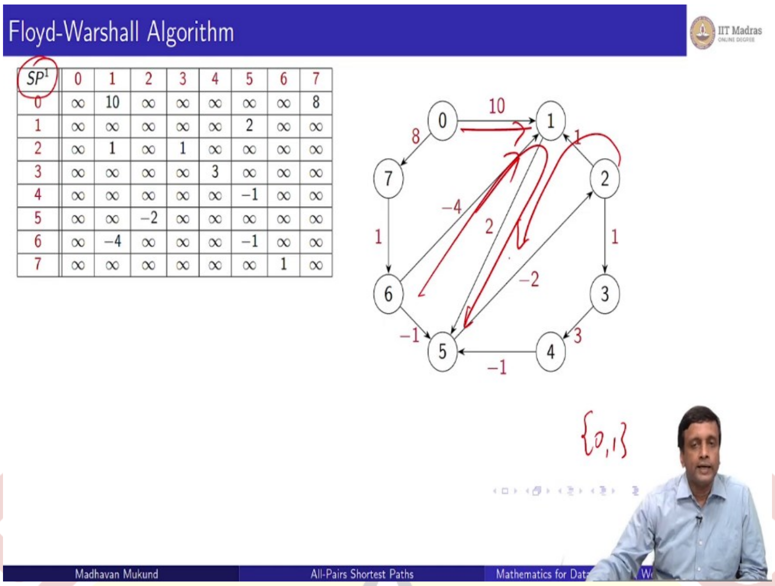
(Refer Slide Time: 15:20)



So, now what is SP 1, so SP 1 represents all things that you can reach going through the set 0. But look at 0, see 0 has no incoming edges, I cannot go to 0 and then go from 0 anywhere. Because 0 is such that you can only go out from 0, so I cannot go from i to 0 and then 0 to anywhere. So, intermediate, 0 as an intermediate vertex is not useful in this graph, there is nothing I can go to via 0. So, therefore, SP ^{1} is the same as SP ^{0}, that is if I am allowed to go through 0, again, nothing. If I am allowed to go through 0, again, nothing.

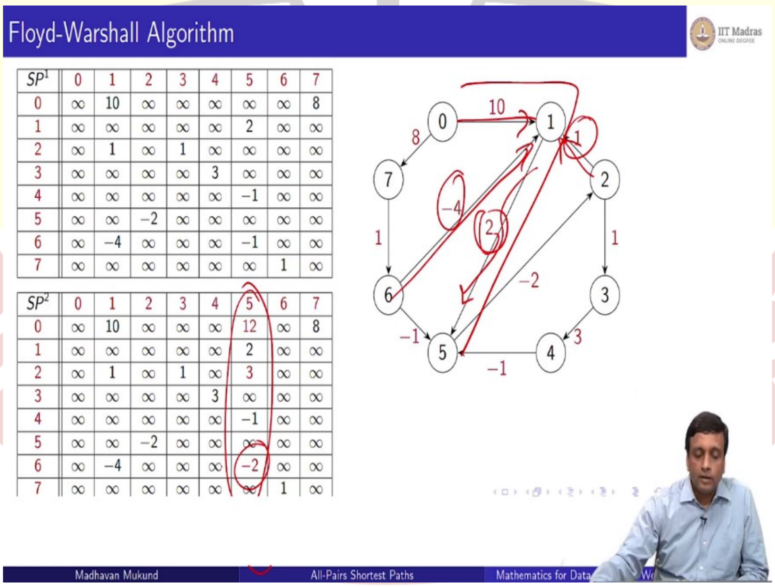
So, this is a certain difference between this and the transitive closure thing. So, there are paths of length 1, and then there are paths of length 2, but the paths of length 2 cannot go through 0, because if I go through paths of length 2 through 0, I have to enter 0, and I cannot enter 0, because the directions are all pointing out of 0. So, in this case, the first iteration produces no change.

(Refer Slide Time: 16:19)



Now, this was my first iteration. So, this is SP 1, which I just calculated is the same as SP 0. Now I want to calculate SP^2, SP^2tells me I can go through 0 and 1. So, now I can go through 1. So, there are edges which come into 1, so where can I go by going into 1 and then out of 1 because I cannot go back to 0 anyway. So, I can go, for instance, from 6 to 5 via 1, or I can go from say 2 to 5 via 1. So, these are two of the things that I should get new.

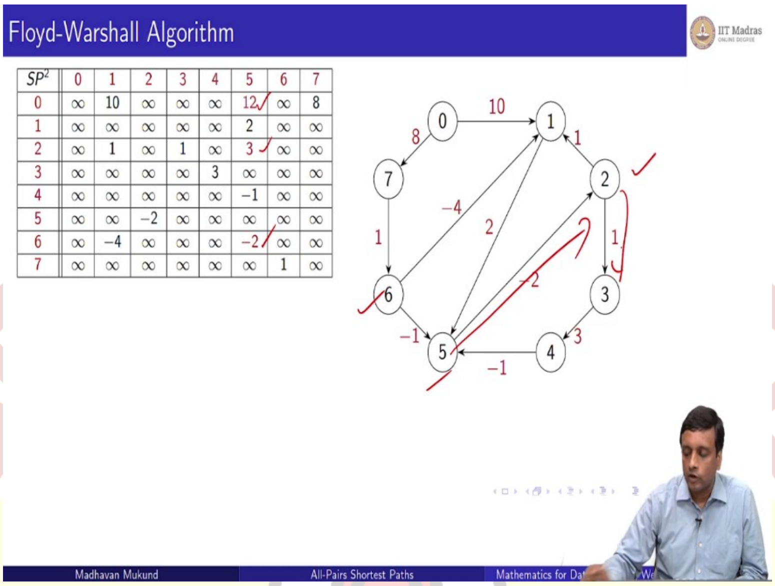
(Refer Slide Time: 16:51)



And indeed, when you see that, you will see that I get from, so the 3 incoming edges are 0, 6, and 2, these are the three incoming edges to 1. So, these give me new possibilities of going from 0 to somewhere from 2 somewhere and 6 to somewhere, but the only thing I can reach from 1 is 5. So, the all the updates happen in column 5 because that is the target. So, from 0 to 5, I now find that there is a shortest path, which is 10 + 2, so I go 10 and then 2. Similarly, from

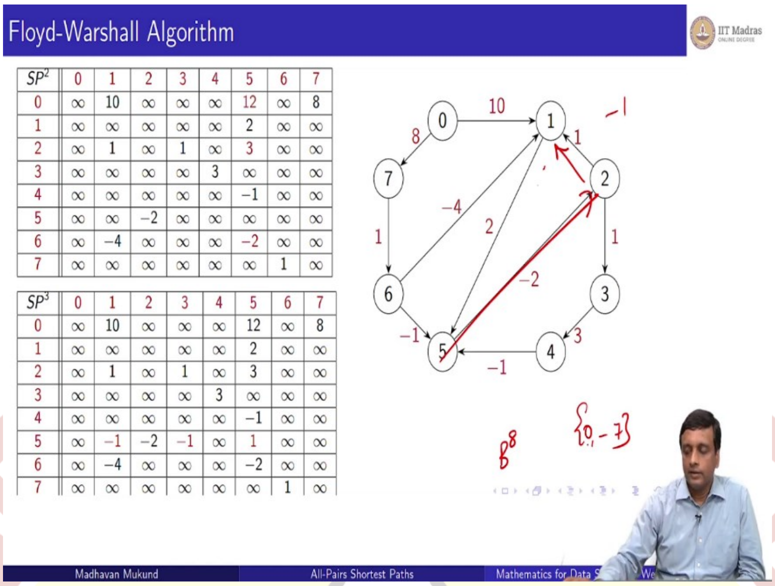
2 to 5, I find the shortest path, which is 1 + 2 is 3. And similarly, from 6 to 5, I find the shortest path, which is - 4 + 2 is - 2. So, this is my first update.

(Refer Slide Time: 17:37)



Now, let us do one more update. So, I have now shortest path. So, these were the updates, which I got when I did the first update, and now I can now I have discovered that I can reach 2, I can reach 6, and I can reach 5, no, the new thing I have discovered is I can reach 5. So, now I obviously want all I can reach from 5. So, from 5, for example, I can go back to 2, but from 2, I can go to 3 and so on.

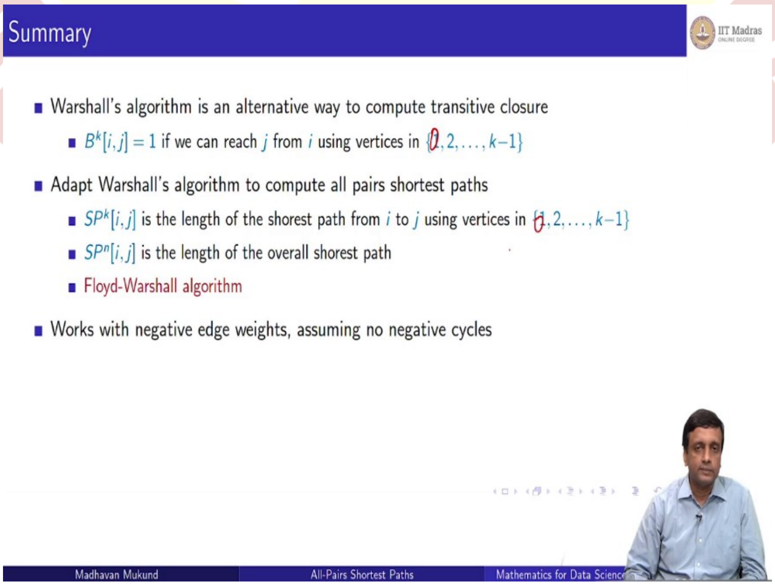
(Refer Slide Time: 18:04)



So, in the next round, I get a lot of updates, so I get updates from 5 to 1 because now I can come to 5. And then I can go to 1 using - 1. And so why is that the case because I can go around this way. So, I can go up and go this way. And this gives me - two + 1, which is not the one I want, I think, 5 to 1, - 2 + 1 is - 1, that is the 1.

So, in this way, I can now update everything one more time. And I can keep doing this. And if I do this, now there are 8 vertices in this. So, if I do this up to B to the power 8, then I will have all paths which go to 0 to 7. And you can work it out, I am not going to work it out, you will get a matrix which gives you all the shortest paths from everywhere to everywhere. So, this is the Floyd-Warshall algorithm.

(Refer Slide Time: 18:59)



So, to summarize, we started with Warshall's algorithm. So, Warshall's algorithm is an alternative way to compute transitive closure. So, it is an iterative transitive closure algorithm. Earlier, we did it in terms of lengths of paths and now we have done it in terms of which intermediate vertices are I am allowed to visit while traveling from i to j. So, that is the difference between Warshall's algorithm and the way we had formulated in terms of paths of length 1, 2, 3, and so on.

So, $B^k[i,j]$ is 1, if I can only reach i to j using only vertices 0 to k - 1. So, when we adapt Warshall's algorithm for shortest paths, what we do is we replace that calculation of OR, I can either reach it without using k or I can use it reach it with using k, I replace that OR by minimum, what is the minimum I can do without using k, what is the minimum I can do with using k and I use that to update this thing from i to j. So, this is called the Floyd-Warshall algorithm and it should work as we saw with negative weights provided there are no negative cycles.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

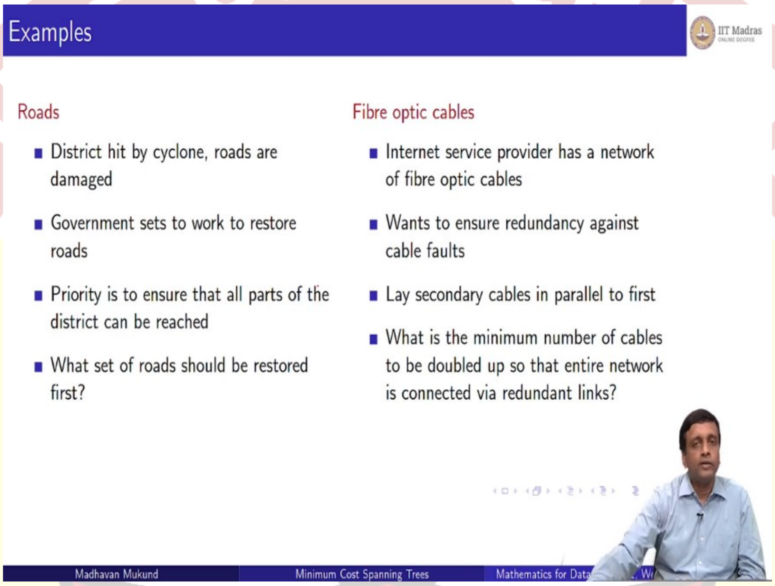
IIT Madras

ONLINE DEGREE

Mathematics for Data Sciences 1
Professor. Madhavan Mukund
Chennai Mathematical Institute
Lecture No. 12.5
Minimum Cost Spanning Tress

So, we have looked at shortest paths, both the single source version and the all-pairs version with and without negative weights. And now in the context of weighted graphs, we move to a different problem, which is the problem of computing minimum cost spanning trees.

(Refer Slide Time: 00:31)



So, to motivate this problem, let us look at a couple of examples. So, here is the first example. So, supposing you are in a district, which has been hit by a cyclone, and many of the roads are damaged. So, immediately after the cyclone, of course, the first priority is to restore the roads. But you also want to restore the roads in such a way that everybody can move around as quickly as possible.

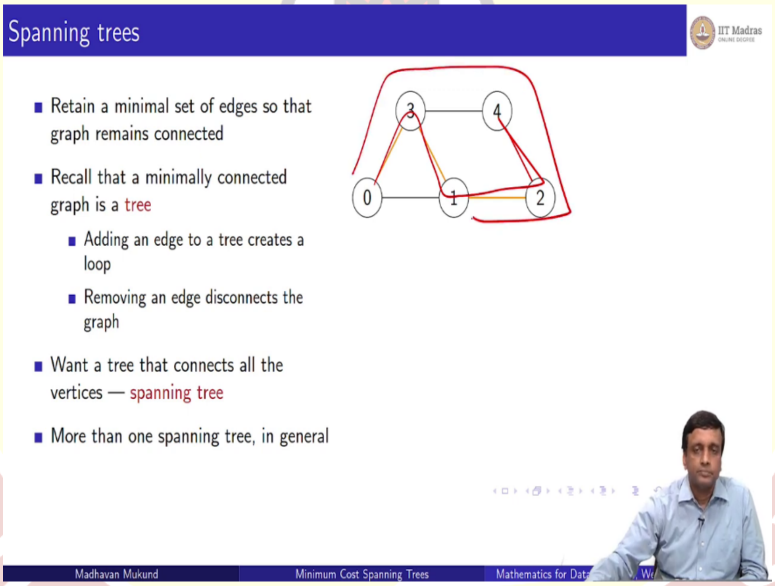
So, you do not want to start at one end of the district and move sequentially to the other end of the district, what you want to do is prioritize the roads to be repaired, so that everybody is connected to everybody as fast as possible. So, which set of roads should you restore, so that connectivity across the district is maximally restored, rather than individual parts being connected and other parts being disconnected?

Here is another context. So, suppose you are an internet service provider. So, you provide internet connectivity to a large number of customers in different cities, and then your customers are demanding reliability. They are saying that in some cases, because of some damage, either due to an accident or due to some construction or something, if a cable between two cities gets cut, then their services cut.

So, you want to lay a parallel cable to ensure that if one cable is cut, the other cable still works. But at the same time, you want to do this in such a way that you do not spend too much laying parallel cables everywhere, you do not want to double up every cable in your network, you want to double up sufficient number of cables, such that between any two locations on your network, there is a redundant route.

So, you are not obliged to put a double cable between every pair of nodes or every pair of cities on a network, only enough of them so that everyone is guaranteed to be connected to everyone else even if one link fails. So, this is a related problem. So, these are both problems will feed into this problem of finding a spanning tree.

(Refer Slide Time: 02:15)



So, a spanning tree essentially asks us how do we take a graph which is connected and retain a minimum set of edges so that it remains connected. So, a minimum set of edges that is connected is a tree. So, we said that a tree is a connected acyclic graph, and we will talk about trees in more detail in this lecture. But the intuition is that if you want to connect n nodes in a minimal way, what you end up doing is connecting them in such a way that there are no redundant paths, there are no cycles, so this is a tree.

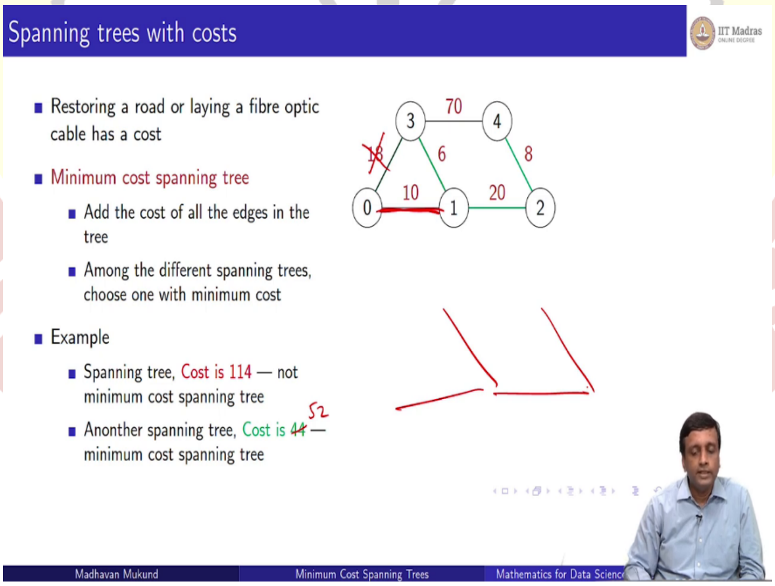
So, if you add an edge to a tree, you add redundancy, so you get a loop. If you remove an edge from a tree because it is kind of minimal, if you remove an edge from a tree, the tree will fall apart, it is no longer going to be connected that is why it is a minimal acyclic connected graph.

So, what we want in this situation both in the road situation and in the telecom situation, that ISP situation is that we want to connect a subset of the nodes. Now, we want to say, we want to restore a subset of the roads, or we want to double up a subset of the links such that everything is connected to everything.

So, we want to find a subset of the edges in the original graph, which if I deal with them, either by repairing the roads or by upgrading them to a double link, I will end up connecting everything to everything. So, here on the right for instance is a graph and this red thing is a spanning tree. So, it is a spanning tree. So, spanning tree is something that connects all the vertices, so it spans the graph, it touches every vertex in the graph and it is a tree so it is a subset of the edges it touches every vertex in the graph, it is a tree so, the red edges here form a spanning tree.

Now, this spanning tree is not going to be unique. So, here is another spanning tree. So, this orange where it is now also this one is also a spanning tree. So, the earlier one was one which went this way and now we have one which goes this way. So, we have two different spanning trees you can have multiple spanning trees. So, you could also have a spanning tree which goes like this for it this is also spanning tree.

(Refer Slide Time: 04:22)



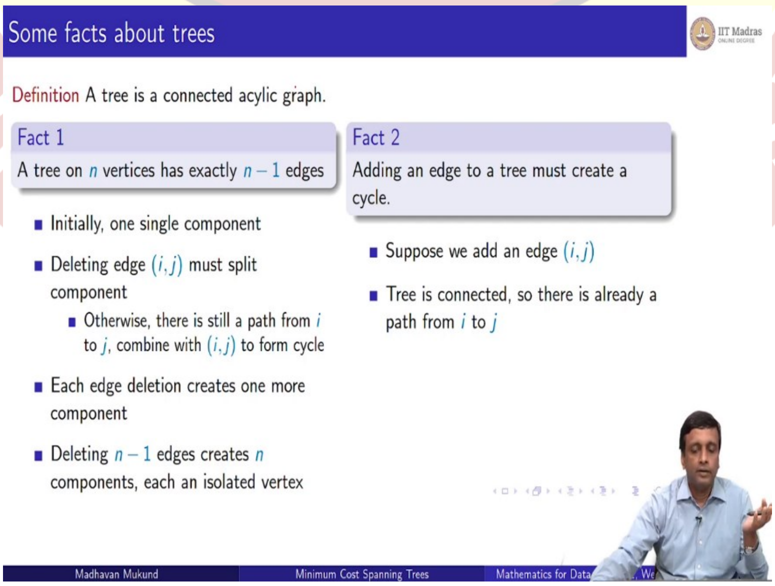
So, our interest is weighted graphs. So, supposing our goal is not just that we want to find a subset of roads to fix or a subset of edges telecom links to double up, but there is a cost associated with this. So, laying a road depending on the location and various other features, laying road may not be the same cost all over the place. Similarly, there may be difficulty in laying cables in some places, not in other places.

So, now if we have a difficulty or a cost or some kind of measure associated with every edge that we want to deal with, can I find a shortest or minimum cost way of doing this? So, I will want to find a minimum cost spanning tree? So, it is not. So, we saw that there could be many different spanning trees. So, it is not just any old spanning tree, but a spanning tree, who if I look at the cost of the all the edges, which I am adding to that tree, so that is the way I am going to define.

So, remember, when we had a shortest path and a weighted graph, we added the cost of all the edges in the path. So, here we are constructing a tree in a graph and we are going to take all the edges that fall into the tree and say that is the total cost I am going to spend if I am going to build this tree, if I am going to repair these roads, or if I am going to develop these cables, this is going to be my total cost. So, I want to find the minimum spanning tree and this is called a minimum cost spanning tree.

So, if I look at this example, for instance, so here is one spanning tree. So, this spanning tree has cost 18+ 6 24+ 17 94+ 20 14. Now, we can easily check that this is not a minimum cost spanning tree, in this case, is small graph, because we can construct this green tree for instance, which has a shorter cost. So, this is 18+ 6 24, 44, this is not this is 52. So, this is actually 28+ 6 44 52. But if I take out this and I put this instead, so this is also a spanning tree. So, this is a spanning tree also for this graph and this you will check has cost 44. It is 28+ 16 is 44. So, among all the trees that I can draw on this particular graph, it turns out that 44 is the best one.

(Refer Slide Time: 06:28)



So, in order to come up with algorithms or strategies to discover minimum costs spanning trees, we will do some basic facts about trees and these will be useful in general. So, it is very good to write them down once and for all so that you know them and you remember them so that you are aware of what you are doing when you are dealing with trees. So, as far as we are concerned, the basic definition of a tree is that it is a connected graph and it is acyclic this is all we are told you are given n vertices, the graph on n vertices is connected, and it has no cycle. So, we are assuming this is an undirected graph. So, it has no undirected cycles. What can you conclude from this?

So, this is a tree, a tree is just a connected graph, which is acyclic. So, the first thing we will conclude is that if the graph had n vertices, then the tree must have exactly n-1 edges, not more, not less, it has exactly n-1 edges. So, here is one argument why that is the case. So, we know that this graph is connected. So, remember that we talked about connected components. So, as an undirected graph, this whole graph, that I am given initially, is a single component because it is connected.

But now I also know that it is acyclic. So, it is acyclic, I claim that if I delete an edge, then it must disconnect the graph because if it did not disconnect the graph, then if I delete an edge, and I can still go across that edge from i to j via some other route, so I have deleted an edge ij, in the tree that is given to me, before I deleted the edges connected after deleted if it is still connected, it means there is still a way to go from it.

But if there is still a way from a go from i to j, does not involve this edge ij. So, if I add back this edge, ij, then I can go from i go to j by the other path and then come back on the site. So, there is like, but I also know that the tree is acyclic. So, therefore, it must be the case that when I remove an edge from a tree, the tree will fall apart into two components. It cannot fall by more than two components, because there is only one edge only connects two parts. So, the whole thing was one component, it is like a cut one thread, and the whole thing falls apart and two pieces.

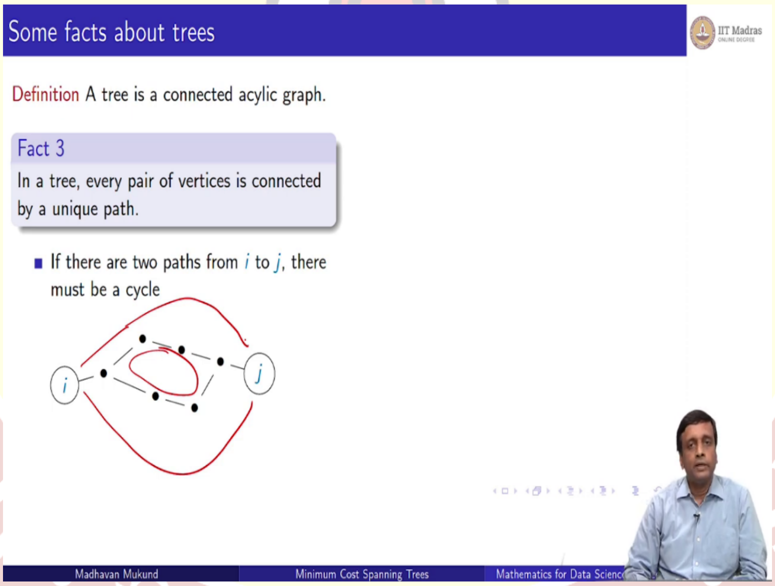
Now I have two connected things. I cut one more edge, what will happen, one of these two will fall into two more things. So, every time I cut an edge, I create an extra component, though I make one component or two components, the other components are unchanged. So, every time I delete an edge, I am going to create one more component. But how many components can I create?

Well, I claim that at most, I can create n components because there are only n vertices. Finally, the minimum component is disconnected vertices isolated by itself with no connections. So, I started with one component, and I ended up with n components. And every time I did+ 1, so how many times can I go from 1 to n to n+ 1, n-1 times, so I could only delete n-1 edges. So, this is one argument saying that every tree on n vertices must have exactly n-1 edges.

So, this says there are no more than n-1. And you can obviously argue that if I had fewer than n-1, then at some point, this thing would have got disconnected earlier. And that is also a contradiction. The other flip side to this is that if I add something to this tree, then it will create a cycle. So, in some sense, this is a minimum connected graph, it is a minimal connected graph, that is adding more edges will only complicate the situation in terms of connectivity.

So, adding an edge is essentially symmetric to what we said before. So, we said before that if you delete an edge, you must split the graph into two otherwise it would have been a cycle. Now if I add an edge, I know that i and j are already connected in the tree. So, if I add an edge by the same logic, I have created a cycle. So, therefore, whenever I add an edge to a tree, it creates a cycle.

(Refer Slide Time: 10:13)

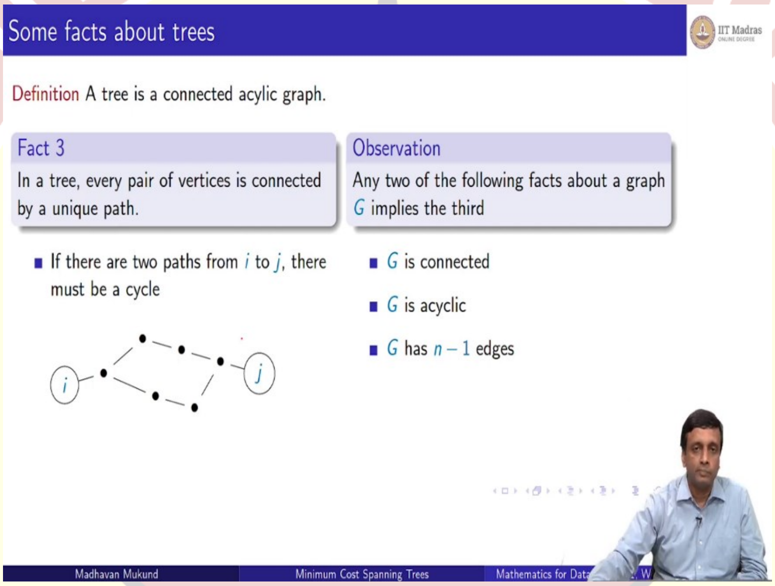


The third fact is that between any two points in a tree, there is only one way to go. There is only one path between any two vertices in a tree. This is not true, in general, as we have seen in many graphs, you can go many ways. For instance, when you are calculating shortest paths, we found alternative paths, which got us shorter weights, and so on. But in a tree, this is not possible in a tree, I can only go from i to j in one way it is connected, guaranteed, but it is connected by only one way.

So, we will just look at a pictorial thing. So, supposing there are two ways to go from i to j. So, the argument is that if there are two ways to go from i to j, then somewhere in between, there must be something like this, a structure like this, where the two paths diverged, and the two, so the two paths might divergent at i and j itself, it might be that I have to completely separate paths i to j.

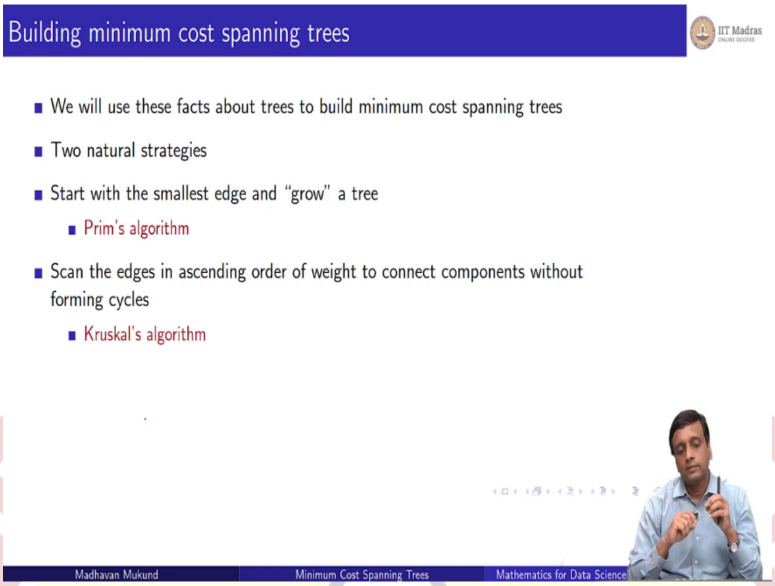
But whichever way if I can go to i one way and come back the other way, either on the entire full circuit or somewhere in between, there must be the cycle where I can go around the cycle. So, if I have multiple paths from i to j, there must be a cycle somewhere. So, these are these 3 facts about trees. So, it has exactly n-1 vertices. If I add an edge, it creates a cycle, and there is a unique path between any two vertices.

(Refer Slide Time: 11:21)



So, to combine this, we can say that, if I give you any two of these conditions, then the graph is a tree. So, if I tell you that the graph is connected, and acyclic, what is the definition of tree, I showed you that it has n-1 edges, it is connected as n-1 edges, then it must be acyclic, if it is acyclic and n-1 edges must be connected. So, if I tell you any two of these three facts, you can conclude that the graph you are looking at is a tree. So, this is a very useful thing to remember when you are going forward.

(Refer Slide Time: 11:49)



So, we are going to use some of these facts in order to design algorithms for this problem that we are considering, which is to build a minimum cost spanning tree. So, remember, a minimum cost spanning tree is a tree which touches every vertex of the given graph by taking a subset of edges, which covers all the vertices. And among those, you want a tree in which the sum of the edge costs that you have used to build this tree is minimum.

So, there are two strategies that one can think of to do this. And we will look at two algorithms follow these strategies. The first strategy is to start from a single vertex or a smallest single edge and grow a tree. So, you try to build a tree incrementally, you start, and then you keep building a tree. So, you start at an edge make another tree, add an edge, make a bigger tree, to add an edge and it does not make a tree you do not consider it. So, you just grow a tree, so we will look at it is called Prim's algorithm.

The other way is to take a disconnected thing and connect it into a tree. So, initially, you can say that all the vertices are apart and you say, let me take a small edge and connect to things. So, now I have got the starting point. Let me take two other edges, two other vertices, connect them and then let me connect this to that. So, you build a tree by kind of grouping together the components rather than growing one tree. So, this is called Kruskal's algorithms. We will see both of these in detail. So, you will understand the difference between these two strategies and see how they work.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS
सिद्धिर्भवति कर्मजा

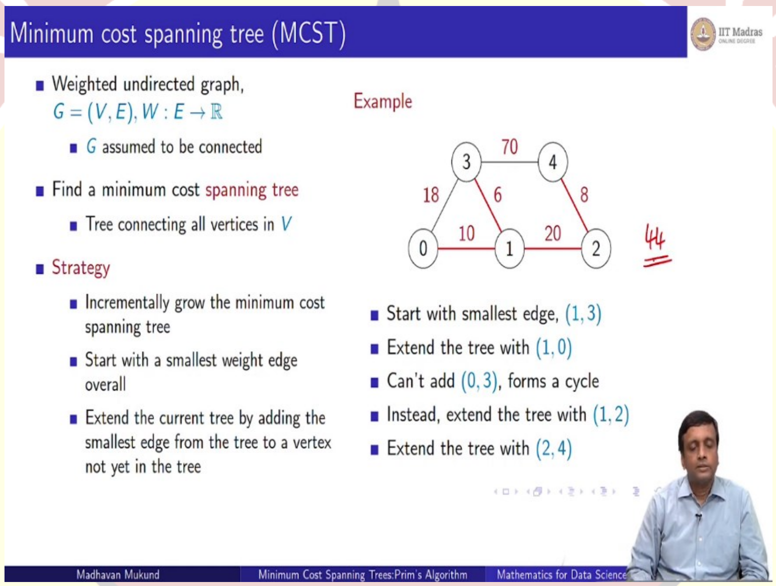
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Chennai Mathematical Institute
Lecture No. 12.6
Minimum Cost Spanning Tress: Prim's Algorithm

So, we are looking at minimum cost spanning tress and weighted graphs and we said there are two natural strategies which are in simplified by two standard algorithms. So, the first one is Prim's Algorithm.

(Refer Slide Time: 0:25)



So, we have a weighted graph, so a graph with a weight function which assigns a number to every edge. And let us assume that the graph is connected because otherwise, if is not connected to begin with, we cannot super impose a tree on it. What we want is a spanning tree, a tree which is the subset of the edges which connects the graph and if there are not enough edges to connect the graph to begin with, we do not even have a starting point.

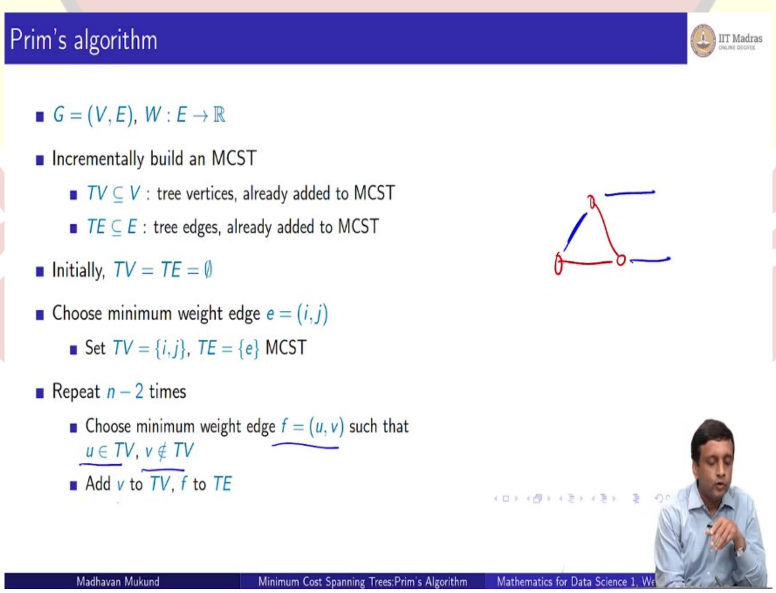
So, assuming the graph is connected and it has weights. We want to find a minimum cost spanning tree which connects all the vertices in V. So, the strategy that we are going to adopt is to incrementally grow it. So, we start with the smallest edge and we will keep growing the tree by adding the smallest edge that we can to the current tree while retaining a tree. So, let us look at this example that we saw before.

So, supposing we start with the smallest edge. So, the smallest edge is the edge between 1 and 3. Now we want to grow this tree. So, the smallest edge with which so we need to grow it meaning we have to choose an edge which will extent this tree. So, we cannot choose for instance this edge right now because this edge is not connected to the tree. That is the smallest edge over all but it is not connected to the tree. I want to grow the tree. So, I can take any one of the edges coming out of here. But not that one that one is out I cannot do that one.

So, I can take any one of the edges which is leaving the tree and extend it. So, I take the smallest among those which is 0 to 1. And so now I have a tree which has 2 edges 1 to 3 and 0 to 1. Now I can take any edge which is leaving this tree. So, the smallest among them is this edge with weight 18. But unfortunately when I add this edge with 18 I create a cycle. And this is not a tree. So, I cannot do this, so I have to throw this away. And I have to go to the next one.

So, I go to the next one which is the 20. And now I have tree which connects four of the five vertices. And now finally because I have reach 2 and now allowed to add this edge. Because now this edge is connected to the tree that I have constructed so far. So, now add that edge and I will get this tree which if you remember we have said last time that this tree has weight 44. And you can check that we have actually found that one, so this is Prim's Algorithm.

(Refer Slide Time: 2:34)



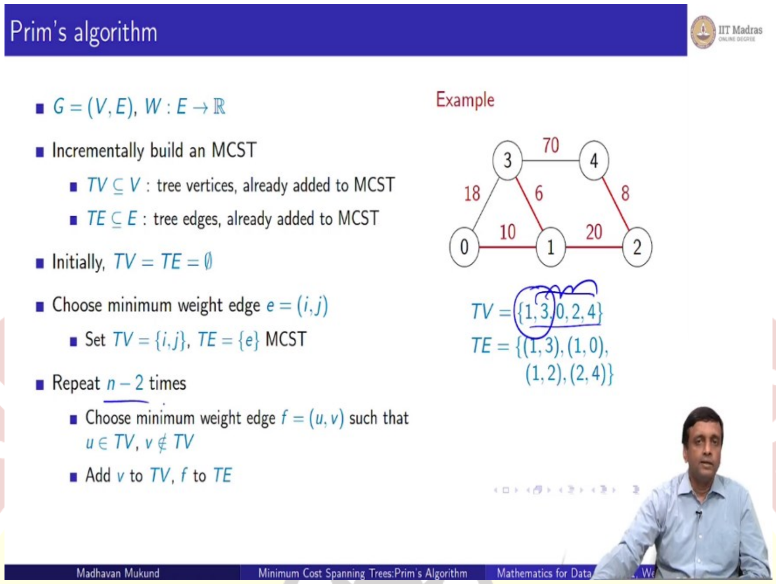
So, formally what Prim's Algorithm does is it incrementally builds an minimum cost spanning tree. So, it keeps track over set of vertices which have been added and keeps track of set of edges which have been added because just because of vertex has been added does not mean that all the edges between them have been added. So, if you look at previous one for instance when we have added 3 and 0, it does not mean that the edge 0, 3 is added. So, we have to separately note which edge is belong to the tree.

It is not just enough to say, it is not like a Dijkstra's algorithm where we said we have burned these vertices so we are done with them. We need to know which are the edges were used in order to construct the tree. So, we keep track of these two things separately TV, the tree vertices. And TE, the tree edges. So, initially everything is empty. The way we describe it we will start with the smallest edge. So, overall we look at all the edges in the graph and we pick the smallest edge that edge let it be called, let it be from i to j.

So, once I do that, then I am started with the minimum tree. Let a tree consisting of just one edge. So, the TV, the vertices in the tree are i and j and the edge is this edge e which I just added which is from i to j. And now what I will do at each step is I will take an edge which starts in T and goes out of T. So, that is the reason we could not introduce when we had this graph already drawn the reason we could not choose this edge is because it both end points are already in the tree that we have constructed. So, we need an edge like this or like this which starts in the tree and end outside the tree.

So, we choose a minimum weight edge such that u belongs to tree and the outside the other edge end point of the edge belongs to the other side. It is not in the tree. Among all these edges which has starting from the tree and going out you take the smallest one and add it. Once you add it you have added a new vertex to the tree. So, TV gets expanded to add v in it and TE gets expanded to add this edge. So, this is the algorithm for Prim.

(Refer Slide Time: 4:28)



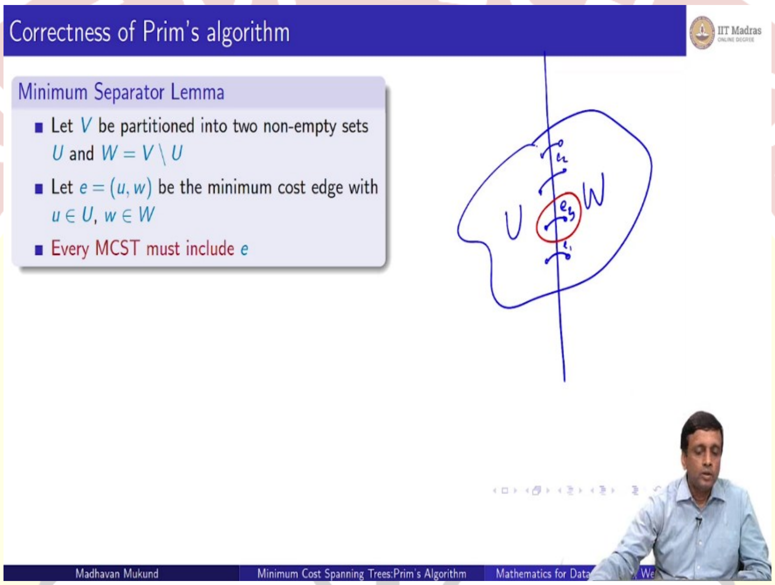
So, again going back to that example so this is how we start. We start with both sets empty and now we take the smallest edge overall which is this one. This is how we start the algorithm. So, when we start the algorithm we say that 1 and 3 form a tree. So, the tree vertex at TV has 1 and 3 and the edge set consist of the edge 1, 3. Now I look at the smallest vertex which has one end point and TV and the other end point outside and it turns out to be the edge 0, 1. So,, I add 0 to TV and I add 0, 1 or 1, 0 because I am drawing it from 1 to 0 to this edge set.

Now I cannot do 18 because it is not inside. So, the next stage that I can do is 1, 2. Because I need to find an edge which is inside to outside. So, this edge is not allowed because I have this condition that you say for example 0 must be in the tree vertex set and 3 must be outside. But both 0 and 3 are inside. So, I have to drop that edge. So, I can only look at edges which leave the tree and go out. So, among those 1, 2 is the next one. So, I get 2 now in my tree set. And my tree edge set has 1, 2.

And finally, from here I can get 8. You are use that edge with weight 8 and get 4 into my edge set in a vertex set. And 2, 4 in the edge sets. So, now I have got all, so I did this n-3 times as the important thing to do. So, I basically started with 1, 3 and I did 1, 2, 3 times. So, did this n-2 times. So, I had five vertices I already started with two vertices. So, I have n-two vertices to go. Every time I do add an edge I had one more vertex into my set.

So, after I do 10-2 times my vertex set has covered all the vertices remember to assume that the graph is connected otherwise it is not going to do work. They not be into able to connect it. So, this seems a little bit, I mean at one side seems reasonable another side it one might ask why is this going to work. I mean why does this particular strategy actually give me guaranteed the shortest smallest tree overall.

(Refer Slide Time: 6:27)

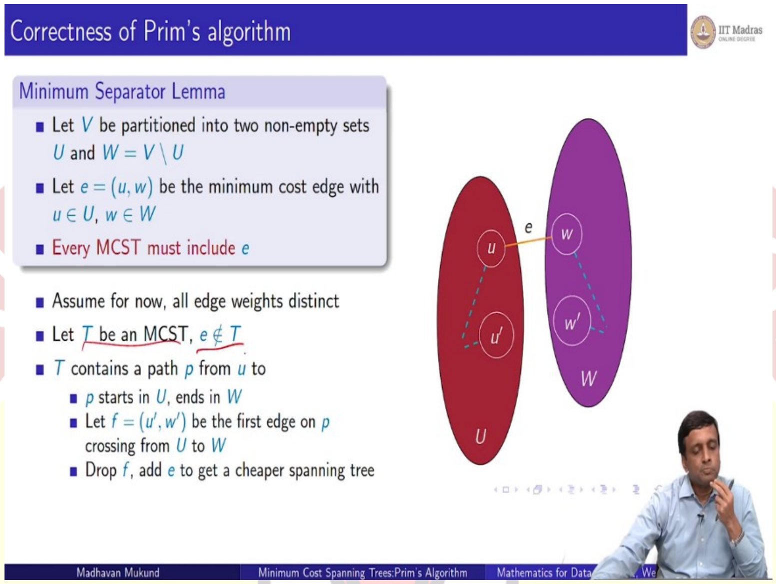


So, to do this let us prove some small graph theory fact call the minimum separator lemma. So, what this says that supposing I take my graph. And I partition it partition it means I divide the vertices in two sets U and W. Now I look at all the edges which have end points on opposite side of this. So, there will be some e_1 there might be some other edge e_2 and so on. There might be multiple edges which go across this partition.

So, they have one end point in U the other end point in W. So, among them let us assume that we pick one which is smallest. So, let us just for moment assume all of them have different weight. So, we will see what to do with different with they have equal weight later on. But let us assume that they have smallest one supposing one of these has actually smallest. So, maybe this is the smallest one.

Then what is lemma says is that no matter how I construct an MCST on this graph that particular edge e3 has to be in MCST. So, the intuition is that somewhere in my graph I will be separating U from in my tree I would be separating U from W and the best way to connect U to W is via e3. So I must use e3 in my MCST. So, this is the claim. So, let us see why this is true.

(Refer Slide Time: 7:49)



So, as I said let us assume for now that all the edge weight in our graph are actually different from each other. So, it just simplifies the argument a little bit. So, this is the situation we are looking at. So, we have a tree and then this says for every partition. So, this is a universal property it says no matter how I partition may think as U and W. This property must hold so let me just assume this is some arbitrary partition on my graph. So, I have split my vertices into two sets which are disjoint which together cover all set.

So, partition means exactly that partition, if I partition my toys with my assistor then I have to take some and she has to take some and both of us get one or the other toy. And nobody leave gets left no toy gets left out. So, partition just means that the two sets are disjoint and the union is the full set. So, that is what partition is. So, U and W together cover all the vertices and there is no overlap.

So, now the question is that supposing I have a tree and I look at this graph this edge which I am promised must be in the tree. It says that the smallest edge which connects a vertex in U to a vertex in W. The smallest edge connecting this partitions must be in my tree. So, supposing it is not in my tree. So, I am assuming the for the sake of contradiction that I have built a tree which excludes this vertex which a lemma promises me should be in my tree.

Now other hand I have built a tree. So, this hypothetically this capital T is an MCST. It is a tree, it is a connected graph on the underline vertices. So, there is no problem going from u to w. the small u to small w, there must be path because it is connected. So, this path starts in the left partition. And it ends in the right partition. So, imagine that this is river separating two sides of the city you cannot go from this side to that side without crossing the river.

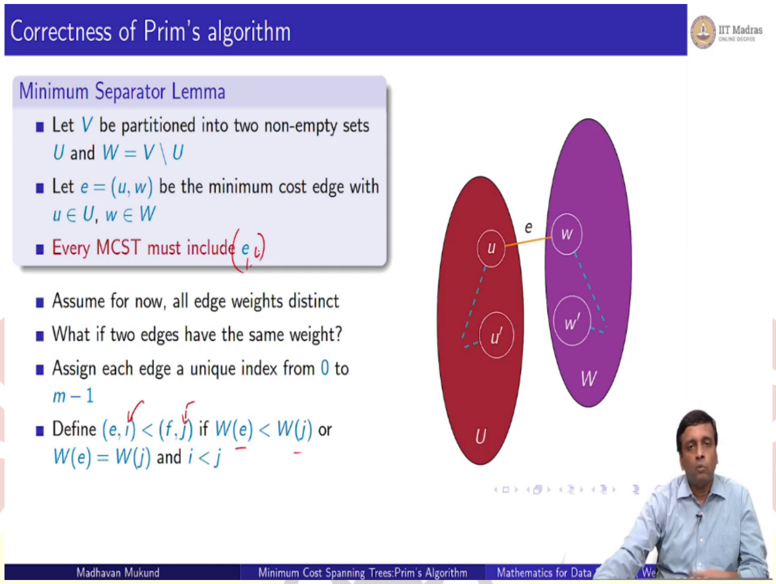
So, I am not using this edge e which I have want but I have to cross somewhere. So, there must be some other edge. Let me call it f where this path crosses. You cannot go from the left side from the red side to the purple side from capital U to capital W without crossing from this side to that side via some edge. And if I assume that I have not taken the edge e which I am interested in which is the smallest edge connecting U to W. I must have taken some other edge.

So, this is what the picture looks like. So, now we can see what happens. So, this is connected, so now supposing I had drop this edge and instead keep this edge. Then I claim that everything that was connected before is still connected anything that I can reach I can now go from there and reach. So, thing that I could reach by following this U to W edge I have this long path which goes via u, w and coming back to w prime and I can do it.

So, therefore, the connectivity does not change but I have replaced f by a smaller edge. So, therefore if I replace f by e I have got a smaller or a cheaper spanning tree in terms of cost. So, this contradicts the assumption that I have started out with a minimum cost spanning tree. I started of with an MCST and I have told you get a smaller one. So, therefore, this could not being the case.

So, either T was not in MCST or T was an MCST but these assumptions that e did not belong as false. So, therefore we have establish through this contradiction this lemma that says that if I take any partition of vertex set and I find the smallest edge going back in fourth across those that partition that must belong to every MCST that you build. So, if you choose that edge you are okay if you do not choose that edge you are in trouble. So, just lets dispose of this case of distinct edge weights before we look at Prim's Algorithm.

(Refer Slide Time: 11:25)

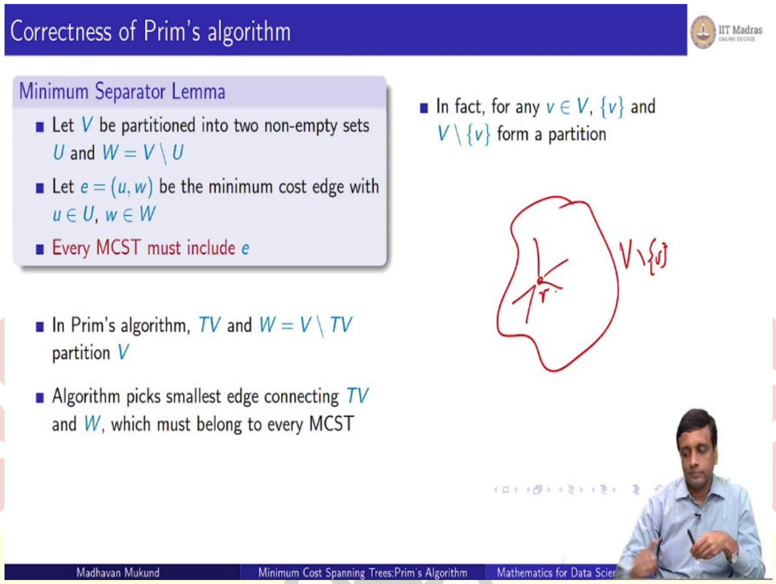


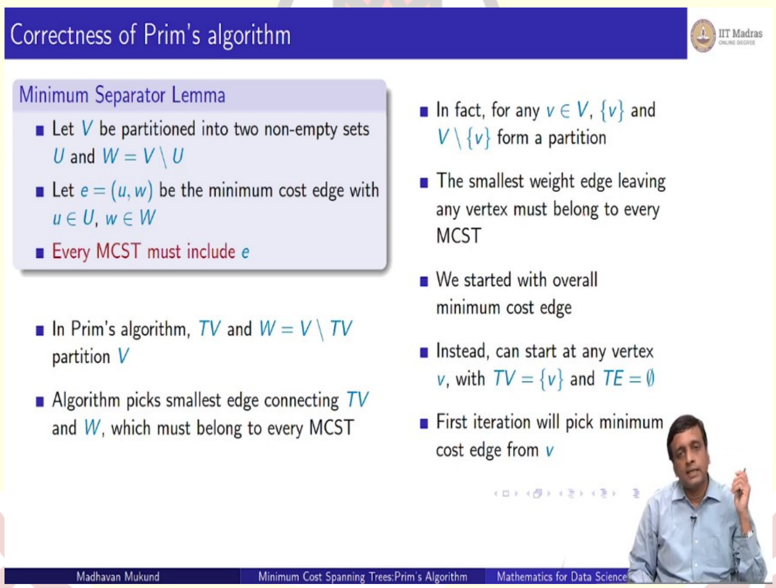
So, what if two edges have the same weight it is not a big problem. Because you just need to have a strategy to choose one or the other. So, you need to arbitrarily decide it one has smaller than the other in order to make this thing work. Which one will go into every tree? So, if you fix a strategy, so fix a strategy basically you assume that you have numbered the vertices in some fixed way from 0 to m-1.

And then you just decide that the ordering in such that either an edge. So, this is the numbering, so e with number i and f with number j now. So, supposing e and f are two edges I would have assigned each of them a number a different number between 0 and m-1. So, I look at now e comma i and f comma j and if the weight of e is smaller than the weight of j, I declare that e is this edge is smaller. But if they have equal weight than I will look at the index and I will say the index i is smaller than the index j then e comma i is smaller than f comma j. So, this gives me a weigh of kind of ordering all the equal vertices.

So, then my the lemma will say that it must include the smallest, smallest in terms of the ordering. So, that is what we say so it is not a big deal so it can be done. So, this lemma holds in general even if the graph has weights which repeat. So, now what is the impact of this lemma on Prim's Algorithm.

(Refer Slide Time: 12:36)





So, in Prime's Algorithm, what we have at any given point is the set of vertices which are in the tree. So if I look at the set of vertices which are in the tree. And the set of vertices which are not in the tree. Let me call it W so this is the set-operation. So, V-TV, so I used it there also. So, this means all the elements which I get by removing the elements of TV from V. S, W and TV together have all the vertices and they are disjoint so it is a partition.

So, therefore, now if I look at what Prim's Algorithm does, it says I am currently looking at a tree which I have built and I am looking at all the edges which go out. Which connect my current tree to vertices are not in the tree. And I pick the smallest one. But that is nothing but this minimum separator. For this particular partition I am picking the smallest one. So, therefore, the one that I am picking has to belong to us. So, I am not picking anything wrong.

So, every, every edge that Prim's Algorithm picks is guaranteed to belong to every MCST by this lemma. And since it does that and it picks exactly n-1 edges overall. All the edges are necessary and therefore they all are part of an MCST. So, that is correctness argument. So, in fact there is a slightly strongest statement we can make. If we look at a vertex V and we look at everything else the whole set-V.

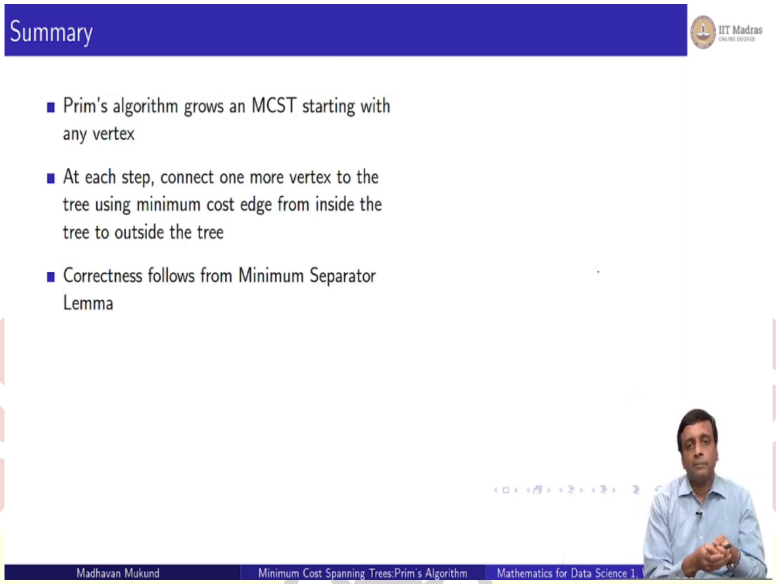
Then if I look at the vertices the edges coming out of V they connect the vertex V to its neighbors and all those neighbors belong to the other partition. So, by this lemma among those edges which are coming out of a vertex all of them are disconnect, connecting the partition containing V alone to the remaining things. The smallest edge leaving a vertex is this minimum separator.

So minimum separator separating V alone form everything else and by this lemma that is smallest edge must belong to every MCST. So, basically if I started a vertex and I look at all the edges which are connected to it. And I pick the smallest one then that smallest one is guaranteed to be in every spanning tree, every minimum spanning tree. So, actually therefore it does not really matter that Prim's Algorithm started with this minimum cost edge that is the bonus.

The minimum cost edge is of course going to be the minimum separator between the partition of the two end points. So, that is fine but you can start anywhere. So, you can start with any vertex and we know that from that vertex the smallest edge leaving it is a minimum cost separator between it and the rest. So, I can start with the vertex TV set to be just a single vertex and no adjust. And then I can apply Prim's Algorithm.

What Prim's Algorithm will first discover is the smallest edge which connects V to one of its neighbors. So, that will by my first edge and so on. So, Prim's Algorithm will work from any starting point, the first iteration will pick the minimum cost edge leaving v which by this lemma is correct.

(Refer Slide Time: 15:53)



So, Prim's Algorithm is a natural way to build a minimum cost spanning tree starting at any vertex as we saw. The way we first presented it was starting with the minimum cost edge but starting at any vertex you can build an MCST because of this minimum separator lemma. So, at each edge what you do is you take the tree have already constructed. And we pick the minimum edge connecting that tree to the rest.

So, we extend a tree one edge at a time. And at each point because we have going from inside to outside the new edge is guaranteed to keep it a tree. And finally every edge that we get is guaranteed to be required by the minimum separator lemma. So, every edge that we add was required to be in the tree and therefore overall we have added no useless edges. So, we must have got a minimum cost spanning tree.

INDIAN INSTITUTE OF TECHNOLOGY MADRAS

सिद्धिर्भवति कर्मजा

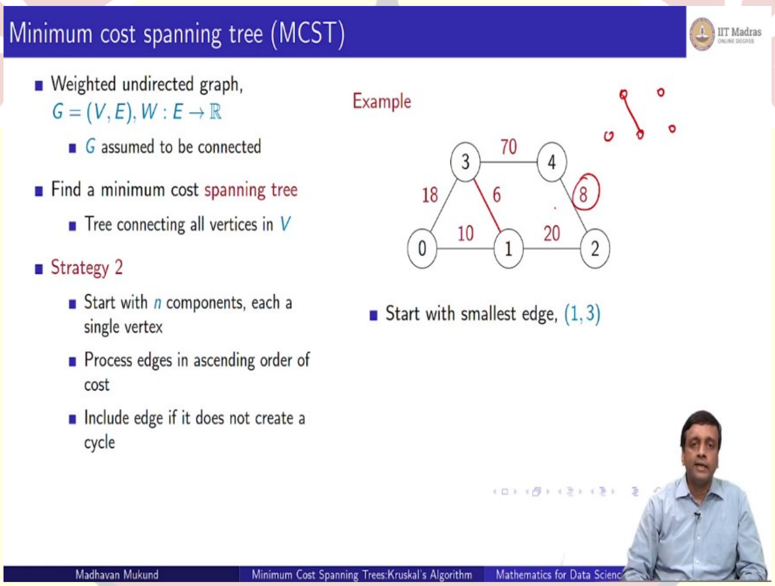
IIT Madras

ONLINE DEGREE

Mathematics for Data Science 1
Professor. Madhavan Mukund
Chennai Mathematical Institute
Lecture No. 12.7
Minimum Cost Spanning Tress: Kruskal's Algorithm

We are looking at minimum cost spanning tress and we saw one strategy the Prim's strategy which tries to incrementally grow a tree starting with one vertex or one edge until you get an overall tree which is minimum cost. The other strategy is called Kruskal's Algorithm.

(Refer Slide Time: 0:32)

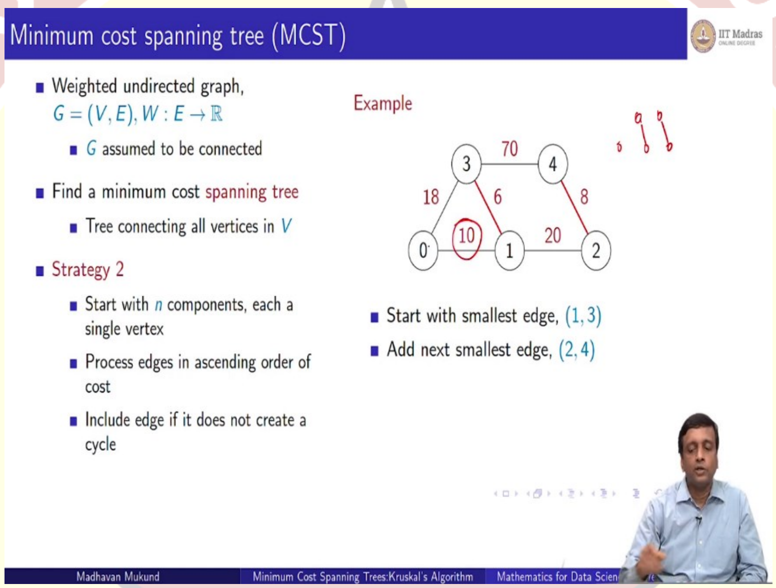


So, remember that we have working with weighted undirected graphs. So, we have a weight function associating a number with every edge. And we assume the graph is connected. And we want to find the minimum cost spanning tree which connects all the vertices in V. So, in Kruskal's Algorithm what we do is we start with all the vertices disconnected forming n components. And then we try to merge components. We try to connect components by the smallest edge that we have which connects to components 2.

So, let us do an example and then we will do it in more detail. So, this our familiar example so here the first thing that we do in Kruskal's Algorithm is to sort the edges in ascending order. So, we will sort the edges in ascending order and then we start with the smallest edge. So, the smallest edge is in this case. So, initially we are imagining that we have this disconnected graph consisting of these five vertices which are just siting an isolation. Then we bring this one edge in and we will create this component.

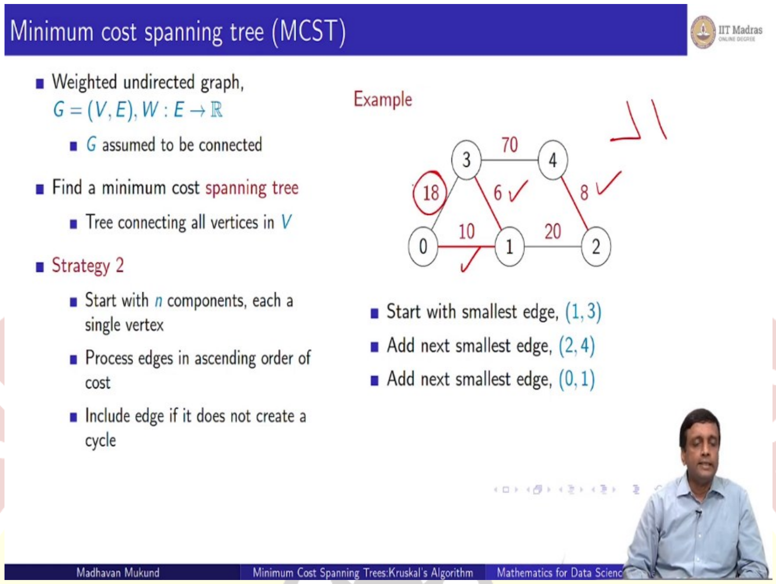
So, now we have 4 components one component has the vertices 1 and 3. And the other three are these isolate components. Now in Prim's Algorithm you would take this component and you would extend the tree. In Kruskal's Algorithm you just take the smallest edge which connects two components and makes them into a larger component. So, in this case I jump from 6 to 8. That is the next smallest edge and it is connecting two components which are separate at the moment.

(Refer Slide Time: 1:57)



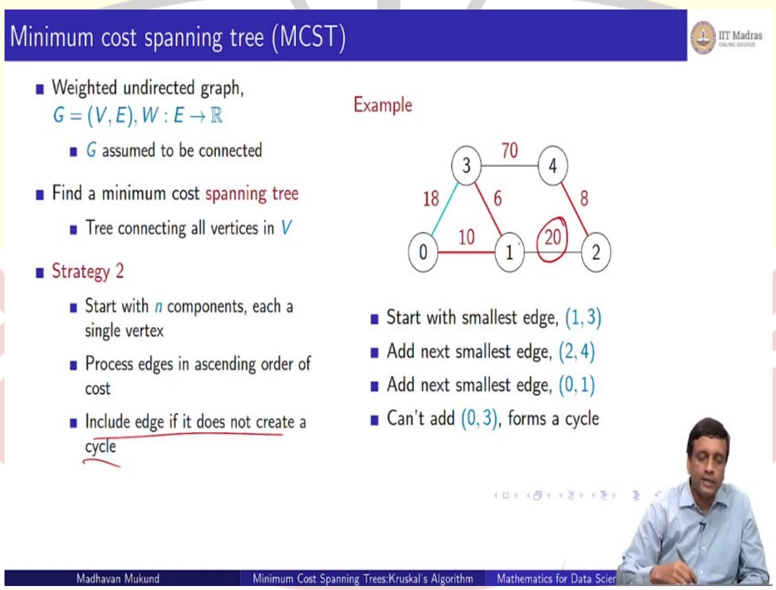
So, I am allowed to add that and I get now these two components I have three components, I have now constructed this kind of a graph. Where I have 0 separate 1 and 3 and then 2 and 4. Now what is the next one? The next one is this 10, so it will connect 0 to 1.

(Refer Slide Time: 2:16)



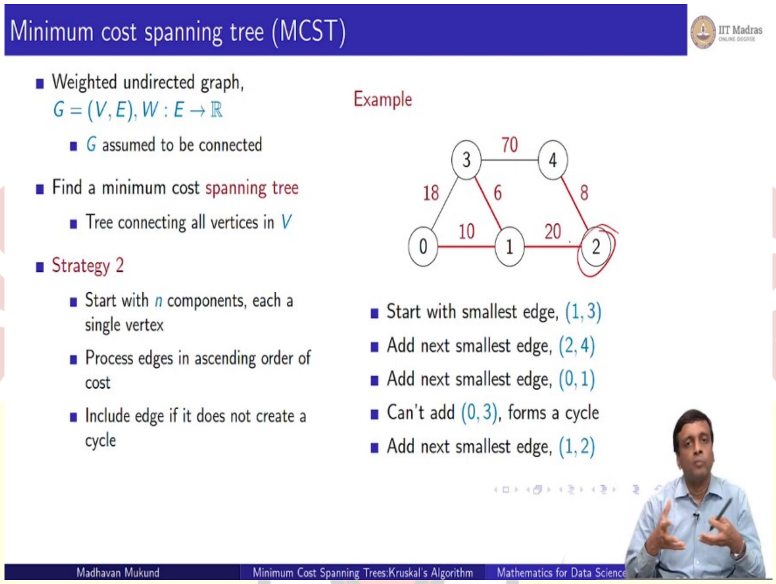
So, I am allowed to add it. So, now I have this kind of situation, so I still have overall disconnected graph it is not a tree yet. And now I have to decide what to do. So, I first is 6, 8, and 10. So, the next vertex in ascending order of cost is 18.

(Refer Slide Time: 2:38)



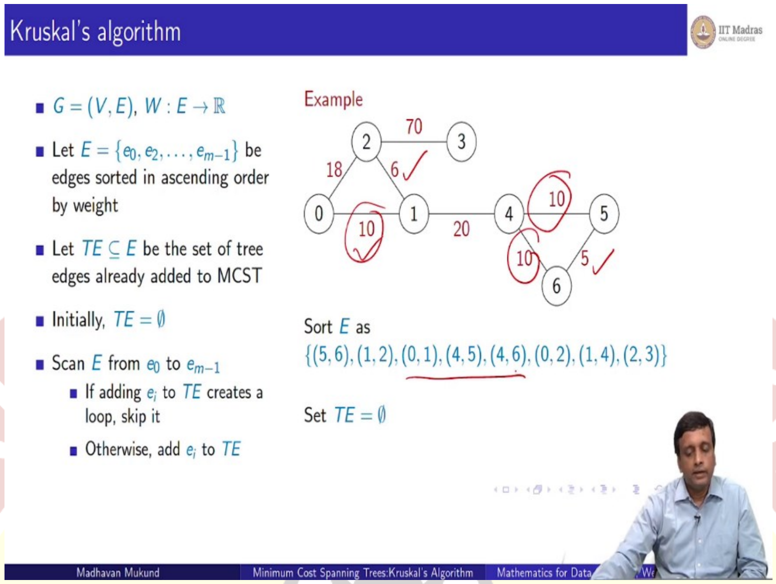
Who I discover now is at when I try to add 18 it does not connect two different components. It connects two vertices in the same component. So, it forms a cycle and this is not good. So, Kruskal's Algorithm says include an edge if it does not create a cycle. So, 18 gets discarded. So, we do not take the edge 0, 3 and we proceed to the next stage. So, the next stage in ascending order is 20.

(Refer Slide Time: 3:01)



So, we add that and we have found essentially the same spanning tree we found from Prim's Algorithm but in a different sequence. Remember in Prim's Algorithm we first added 6 then we added 10 then we added 20 and finally we added 8 because we could only add 8 when we reached one end point of 8 that namely this vertex 2. Where as in Kruskal's Algorithm we added 8 upfront and we create these disjoint components. And as we go along this components kind of merge together and they form a tree.

(Refer Slide Time: 3:28)



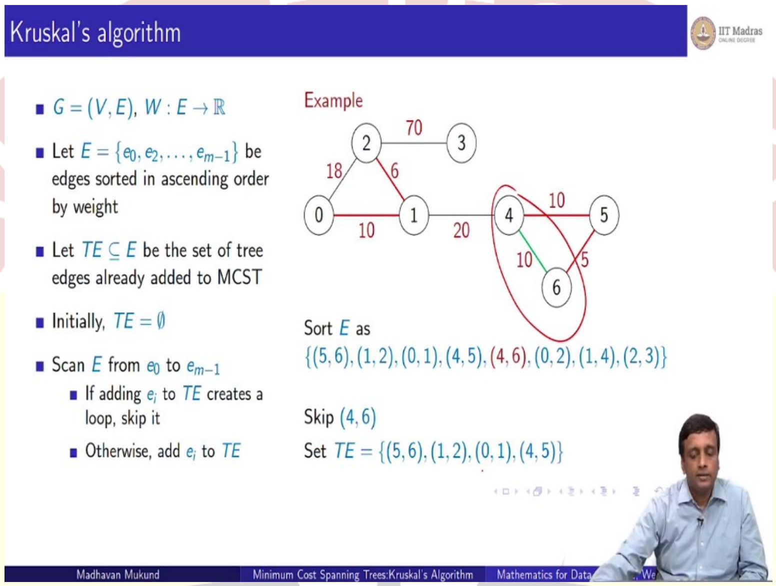
So, here is the little bit more formal definition of Kruskal's Algorithm. So, we have assume that our edges are arranged in this ascending order. So, e0 is the smallest edge this is e1, e0, e1 upto em-1. So, I sorted in ascending order by weight. And now what you do is you keep track of like we did not Prim's Algorithm. We keep track of the set of edges that we have added. And implicitly the set of edges also tells us what are the components.

So, you can if you are trying to actually write this as code you will actually keep track for the components also which is a little bit TDS when you are programming this. But mathematically we know the edges then we can compute the components by just doing kind of the reachability on each component. And finding out which components are. So, initially the set of edge is empty. And now we scan all the edges from the smallest edge to the largest edge.

And if adding it creates a loop or a cycle we skip it otherwise we add it. So, here is the little bit more complicated graph to try and see how this works. So, these are my 0 to 6 or 7 vertices and I have some 8 edges in them. So, I sort the edges so this is my smallest edge and this is my second smallest edge and so on and this is my third. Now I have three edges which are of equal size. So, these 3 edges, so I fix some ordering we discuss that if you have equal weight edges you just fix some ordering on them.

So, I have fixed some ordering which is basically based the lexicographic ordering of the end points. So, 0, 1 comes before 4, 5 because 0 comes in 4, 4 and 4, 5 comes before 4, 6 because 5 come before 6. So, I have just chosen this ordering you can choose any ordering. So, you fix an order in which you are going to process the vertices. Such that this an ascending order if the equal vertices you choose some way to group them so that they are in some fixed order. So, initially, now my edge set is empty and I pick the smallest one.

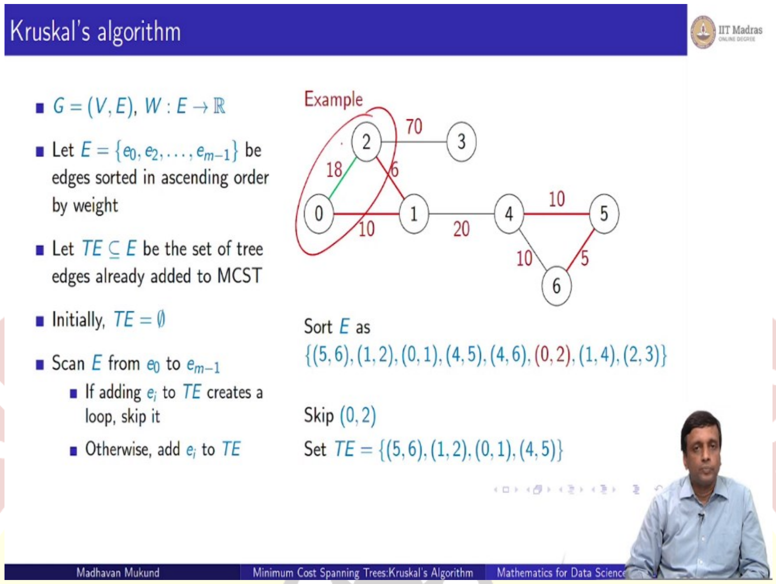
(Refer Slide Time: 5:21)



So, now I process it from smallest to largest. So, I pick the smallest one no problem it does not create any I have got no components before this. So, I now created one component, so I can add it and I have 5, 6 as the tree edge. Now I look at the next stage so remember this is useful that I have already sorted it. So, I know the next stage I have to process is 1, 2. So, I look at 1, 2 again no problem it connects two different components so I am fine. Then I look at 0, 1. 0, 1 also connects two different components so I am fine.

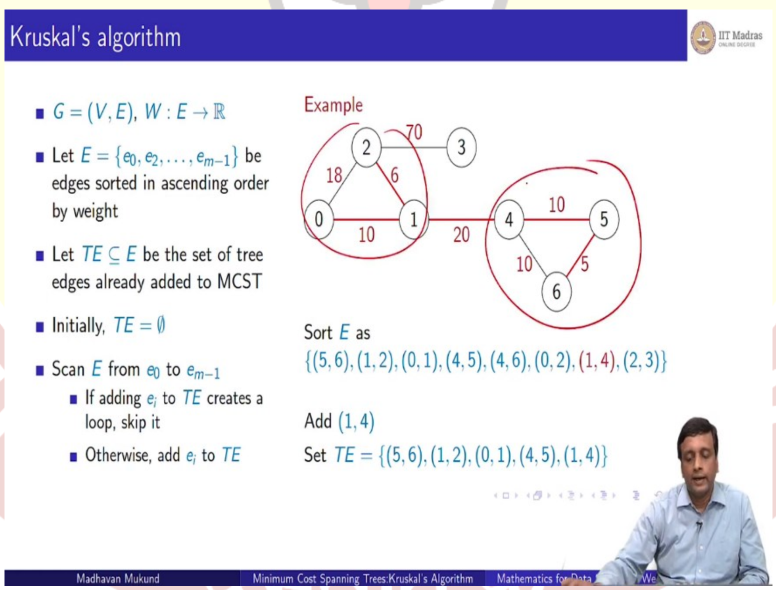
So, now I have come to the 10 block. So, now I look at 4, 5. 4, 5 also connects two different things because 5, 6 was m component and 4 will m component I am fine. Now I come to 4, 6 and 4, 6 now is actually lying within a component. So, I cannot use 4, 6, 5 skip it. So, I do not add to the edge set. So, I leave the edge set unchanged.

(Refer Slide Time: 6:17)



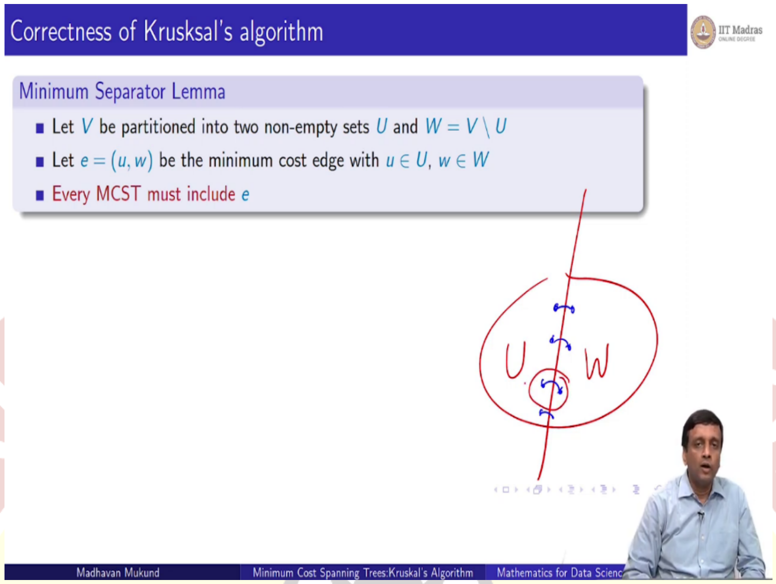
Now I come to 0, 2 same problem 0, 2 is connecting two vertices which are already in the dame component. So, I have to skip that also.

(Refer Slide Time: 6:27)



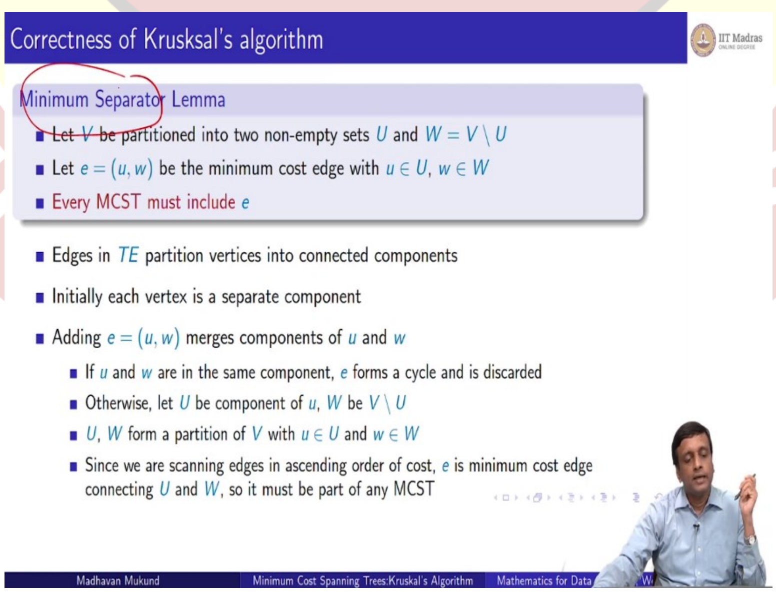
So, now after 18 I go to 20. And that is not a problem because it takes this component and this component and connects them they are two different components. So, I add that and finally that edge 70 which connects 2 to 3 is added. So, this is how Kruskal's Algorithm works.

(Refer Slide Time: 6:40)



So, the reason that Kruskal's Algorithm is correct is exactly the same reason that Prim's Algorithm is correct is because of this minimum separator lemma remember what the minimum separator lemma says if I take my graph and I split it so that there are some vertices on this side and some on the other side. There are this partition into U and W no matter how I partition it. If none sit and look at all the edges which cross and then among these. If I pick the smallest one then that smallest one must belong to every MCST.

(Refer Slide Time: 7:14)



So, in our think the edges that we have found so far in Kruskal's Algorithm partition the vertices into connected components. So, now initially each vertex is in a separate component. And when I add an edge u, w it merges the components of u and w. So, basically the thing is that if I am keeping track of which vertices are in the same component and I can do that incrementally. Even I add an edge I can grow the component I can say the component containing u now contains w also. So, it is not a problem.

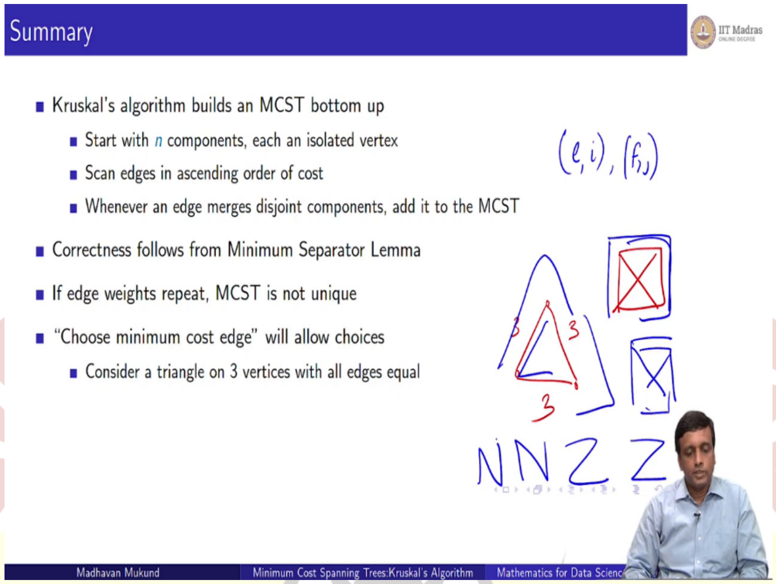
So, I do this as I go along, so if I discover that the edge I want to add actually falls within the component both end points is the same component. It will then form a cycle and then I discover it. On the other hand if it connects two different components then we can apply this lemma to argue that what we are doing is correct because we look at the component containing the starting point. So, let capital U be the component containing small u. And let W be the rest.

Now W has because we are assuming that U and W are in u small u and small w are different components, I have U which contains the small u and this W which is the rest it contains small w. Now we are processing these edges in ascending order. So, since these are in different components I have not yet connected these two. So, I have not found any edges yet between 0 between capital U and capital W.

Because if I have found them I would have already connected these components. So, the reason they are not connected is because I have not found them yet. But among the edges which I have not looked at the edge I am looking at right now is the smallest one. Because I am doing it in ascending order. So, this forms a partition and we have forming this we are scanning in ascending order.

So, this edge that I am looking at must be the smallest edge connecting capital U to capital W. So, it is satisfies this property of this minimum separator. It is a minimum separator of capital U and capital W. So, what Kruskal's Algorithm would do? (())(9:15) pick it up an edit. And it is correct because its separator says that any such edge which is the minimum separator between this partition and that partition. And must be there in every MCST. So, Kruskal's Algorithm is correct for the same reason that Prim's Algorithm is correct because of this minimum separator lemma.

(Refer Slide Time: 9:33)



So, the difference between Prim's Algorithm and Kruskal's Algorithm is basically Kruskal's Algorithm is kind of assembles the tree bottom up. So, it takes all this just the connected things and then it goes around putting them together. Whereas, the Prim's Algorithm starts somewhere and it grows a tree gradually to cover the whole thing. So, they have different strategies but both of them owe that correctness to that one lemma which says that whenever I partition the vertices into two disjoint sets.

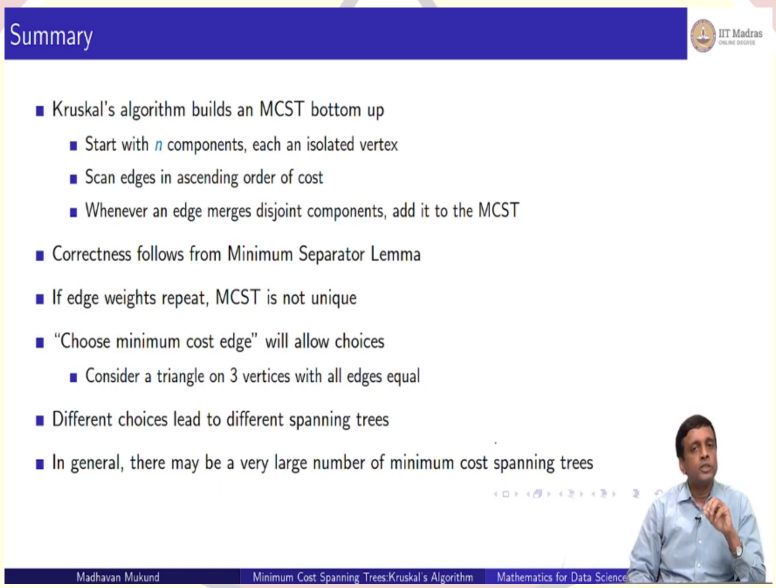
The smallest edge connecting these two partitions must be there in every MCST. Now if there are repeated edge weights then we already saw in the unweighted case that there are many spanning tress. If there are repeated the same weight repeats, then we may not get a unique spanning tree. So, for instance supposing I take a very simple graph which is just this 3 vertex graph and I put weight 3 here, here and here.

Then, any of these would be a spanning tree that I could take this pair of edges or I could take this pair of edges or I could make this pair of edges. So, all of them would be minimum cost spanning tress. And where it comes in our algorithm means when we say choose the minimum cost edge remember we have said that we might actually have to specify it as some f comma j and so on.

So, this ordering that we choose will decide which one will get picked up. So, that is why we get different choices. So, here you can see that this triangle on four vertices. And three vertices has actually three spanning tress. If I do a more complicated thing for instance supposing I take a square. Pick the four two diagonals then what are the spanning tress well the spanning tress. So, there are some obvious tress like this one. So, this is a spanning tree. When I take 3 edges around 4 vertices or spanning tree will have 3 edges.

But these are also spanning tress. Two edges and a diagonal this z is also a spanning tree. Two edges and diagonal connecting them. How many of them are there well there are 4 ways of going around the outside I can have this, I can have this, I can have this. Then there are some 4 ways of choosing the corner to include and the then the z can also be in many ways. So, it is many different orientations. So you can see that with 6 edges I can get a anonymous number of spanning.

(Refer Slide Time: 11:48)



So, in general different choices lead to different spanning trees. And there are not unique edge weight, then I could get a very large number of spanning. So, depending how I have chosen to order this equal edge weights. Prim's Algorithm and Kruskal's Algorithm will pick one particular one out of these. It will not give you all of them it will give you one of them. And if they are disjoint in the sense that it not disjoint if they are distinct that is if all the edge weights are different then using the minimum separator lemma you can argue that every choice on Prim's Algorithm is forced. Every choice in Kruskal's Algorithm is forced and they will give you exactly the same spanning trees.

So, as long as the edge weights are disjoint it does not matter whether use Prim's or Kruskal's you will get the same spanning tree. But if the edge weights repeat then depending on how you choose to order the vertices. The two algorithm might produce different spanning trees with the same minimum cost but different set of edges. So, keep that in mind.