IIT Madras

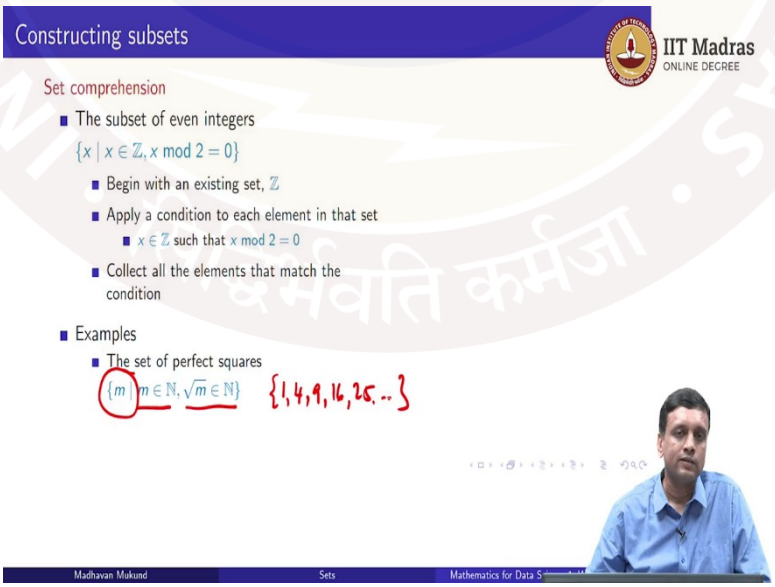
ONLINE DEGREE

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.

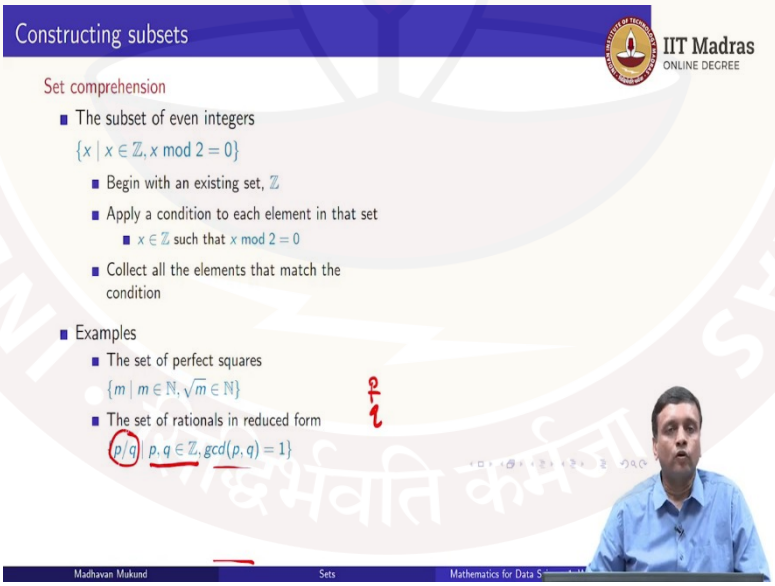
Now we assume that we already have a set like Z or N or Q or R for which we know what elements are. So, we do not have to describe how to pick out element we know what those elements are. What we are now giving is a description of how to choose elements which satisfy a given property. So, let us look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

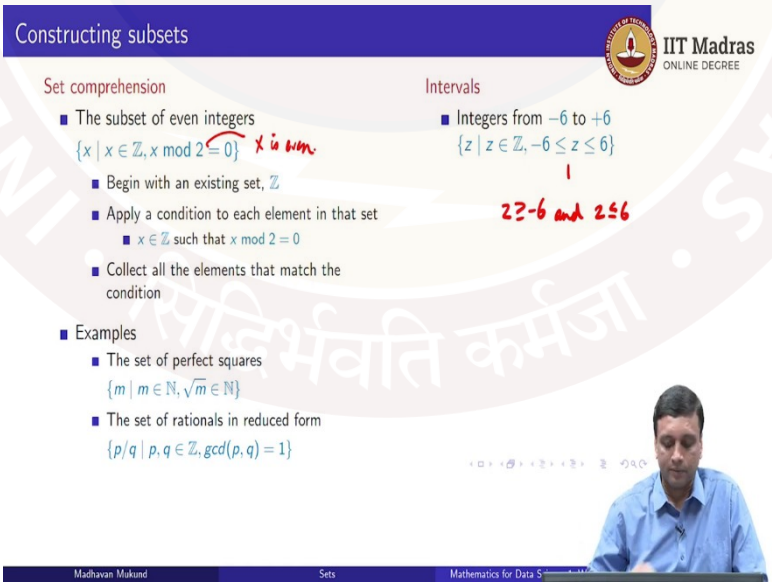
rationals. For example, $\frac{3}{5}$is in reduced form, $\frac{6}{10}$is not in reduced form because I can; cancel the 2 and get $\frac{3}{5}$.

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)

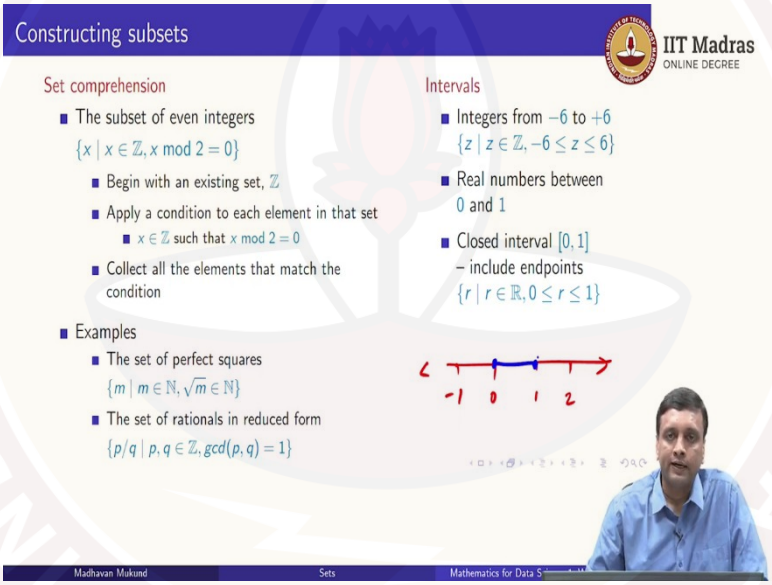


One of the things that we will often use with respect to numbers is to define intervals of numbers between something and something else. So, for instance if you are looking at the integers; we might want the integers from some lower limit to some upper limit. This for

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)

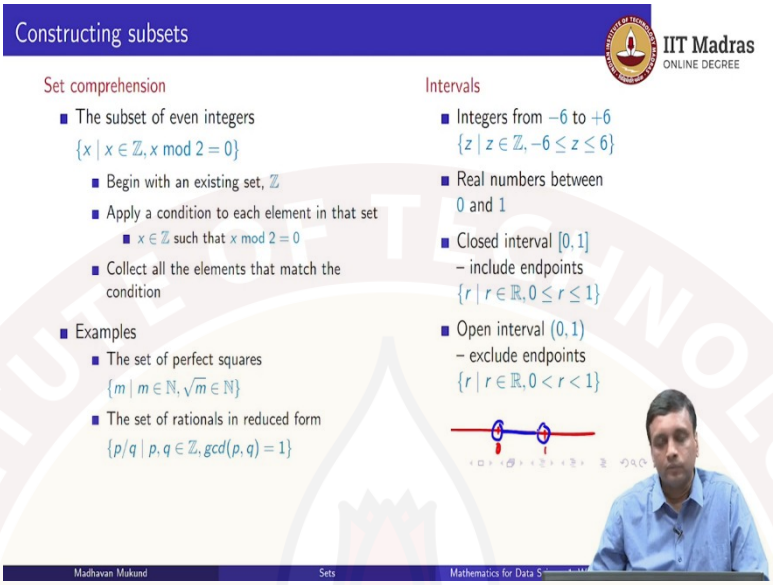


So, intervals are more interesting when we talk about real numbers and one of the intervals that we really often want to talk about is the interval between 0 and 1. So, 0 to 1 is quite interesting because we will often talk about probabilities for instance and probabilities range between 0 and 1. So, 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 $0 \leq r \leq 1$.

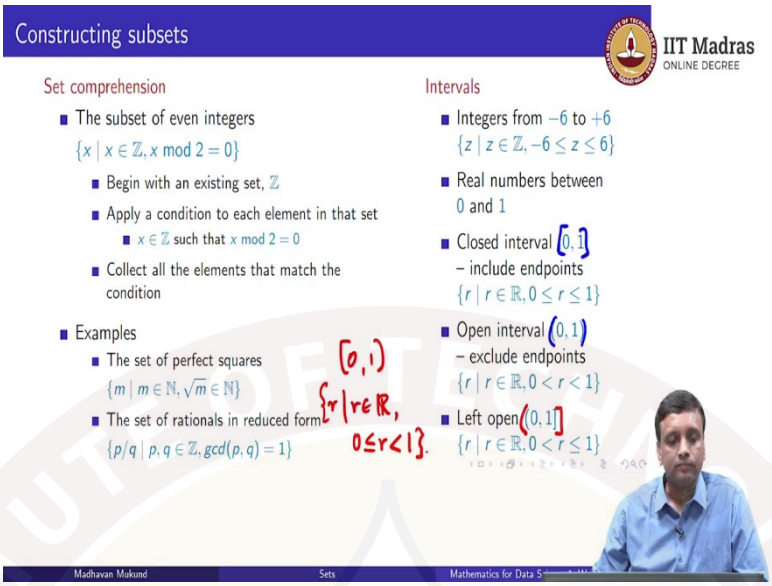
(Refer Slide Time: 07:46)



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.

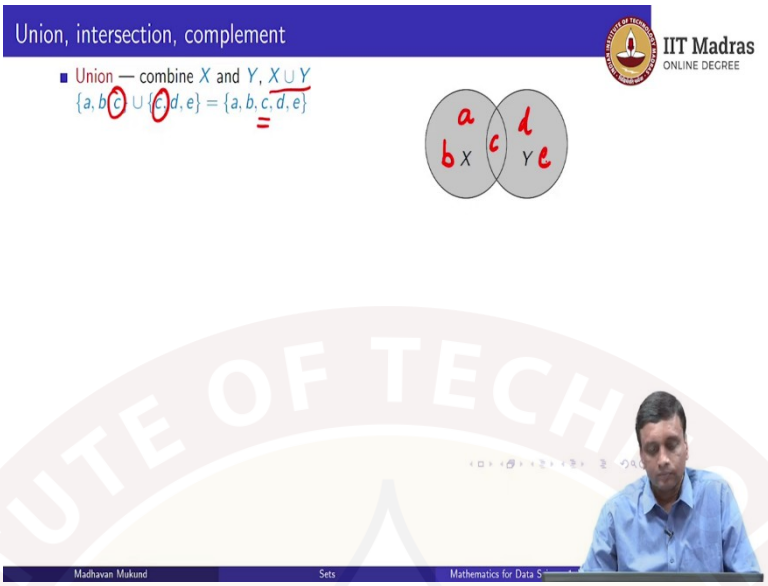
(Refer Slide Time: 08:29)



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.

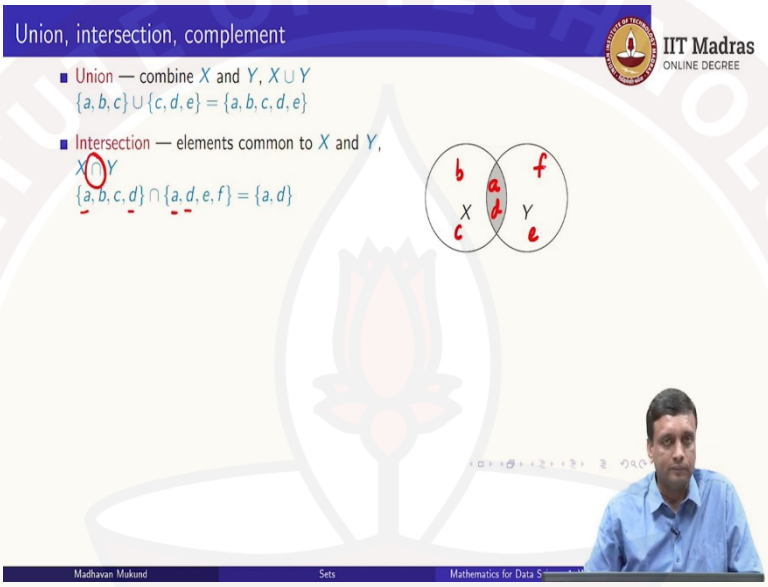
So, therefore, the cardinality of the union will in general be less than the cardinality of the two sets put together. So, here we have two-three element sets, we take the union we get a five element set not a six element set because there are some elements which are common and the symbol for union is this $\cup$ 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)

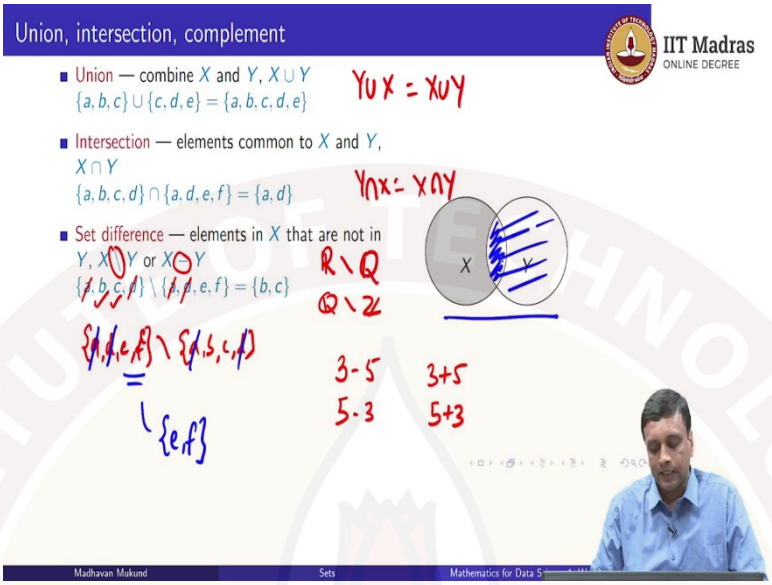


So, the corresponding thing which takes up only the elements which occur in both sets as you know is called intersection. So, intersection is written with the upside down version of the union sign right. So, X intersection Y is written like $X \cap Y$. So, here for instance we look at elements which 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.

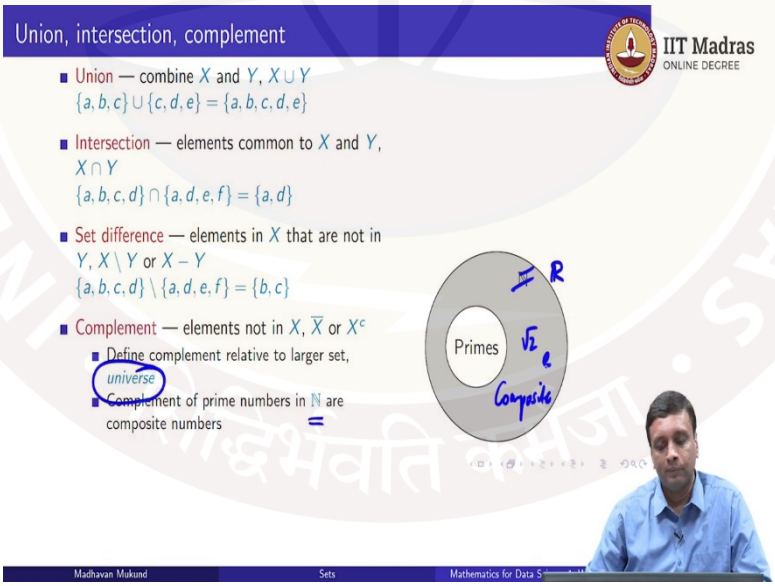
So, here we have that the first set minus the second set has b and c because those are the 2 elements in the first set which are not in the second set. Now, this is like subtraction not symmetric in the sense that you know that 3 - 5 is not the same as 5 - 3 unlike 3 + 5 right. So, 3 + 5 is the same 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.

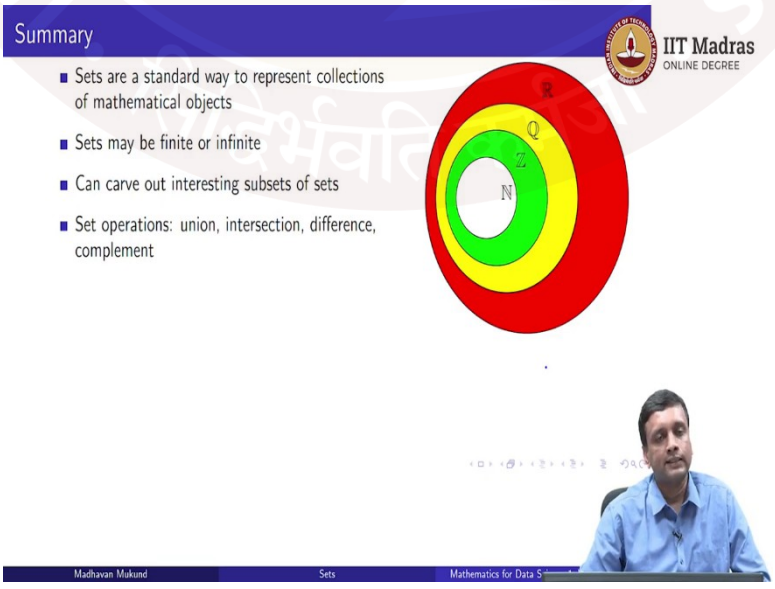
So, either it is prime or it is a composite. So, the composite numbers are disjoint from the primes and they are all the numbers that are not prime. So, this is what we mean by complement. Complement means the opposite side it means everything else, but complement is not very straightforward in set theory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