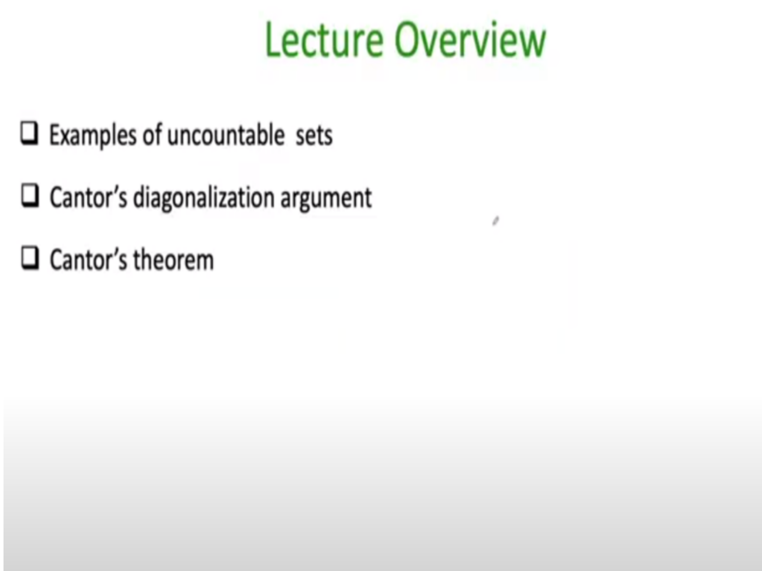
Discrete Mathematics
Prof. Ashish Choudhry
IIIT, Bangalore

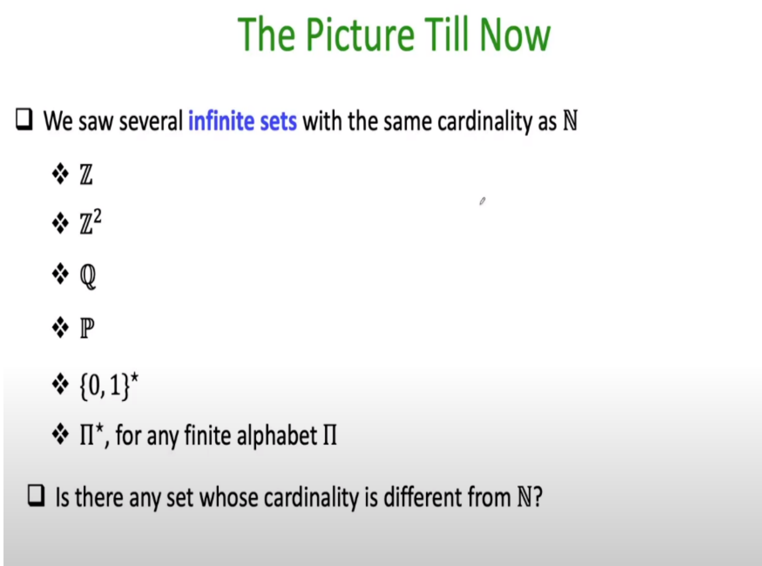
Module No # 05
Lecture No # 29
Cantor's Diagonalization Argument

(Refer Slide Time: 00:24)



Hello everyone welcome to this lecture just a quick recap. In the last lecture we saw various examples of countably finite sets. So we will continue the discussion on countably infinite sets and the plan for this lecture is as follows. In this lecture we will see several other examples of uncountable sets and we will discuss about Cantor's diagonalization argument and Cantor's theorem.

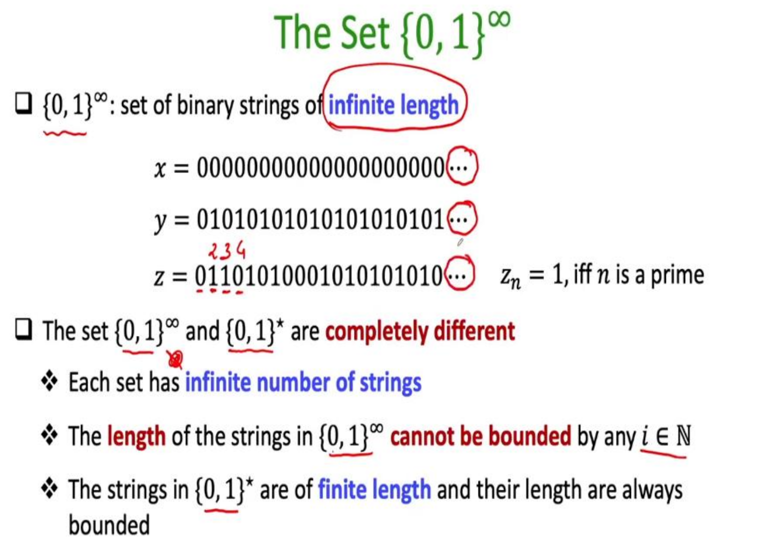
(Refer Slide Time: 00:51)



So as I said earlier in the last lecture we saw examples of several countably infinite sets. And the nice thing about those set is that their cardinality is same as set of positive integers. So we saw several such sets it may be the set of integers, the 2 dimensional plane, integer plane, set of rational numbers, set of prime numbers, set of all binary strings of finite length, and for any finite alphabet the set of all possible strings of finite length over the alphabet.

So it might look like that for every infinite set, somehow we can show that its cardinality is same as set of positive integers. But the interesting part here is that is not the case and the focus of this lecture is to argue about the existence of infinite sets whose cardinality is different from that of set of positive integers.

(Refer Slide Time: 02:05)



So we begin with our first set namely set of all binary strings but of infinite length. And this set is denoted by this notation $\{0, 1\}^{\infty}$. So; some examples of binary strings of infinite length if I consider the string x equal to 0 0 0 0 and the sequence of 0's which never ends then that is a binary string whose length is infinite. Its length is infinite because the characters in the string will never end.

Similarly if I consider this binary string where I have alternate 0's and 1's and the sequence goes forever then that is again it is an example of a binary string which has a infinite length. Similarly if I consider this binary string consisting of 0's and 1's where at the nth position the bit is 1 provided n is a prime number otherwise the bit is 0. So for example 1 is not prime so that is why the first position I have bit 0, 2 is prime the integer 2 is prime.

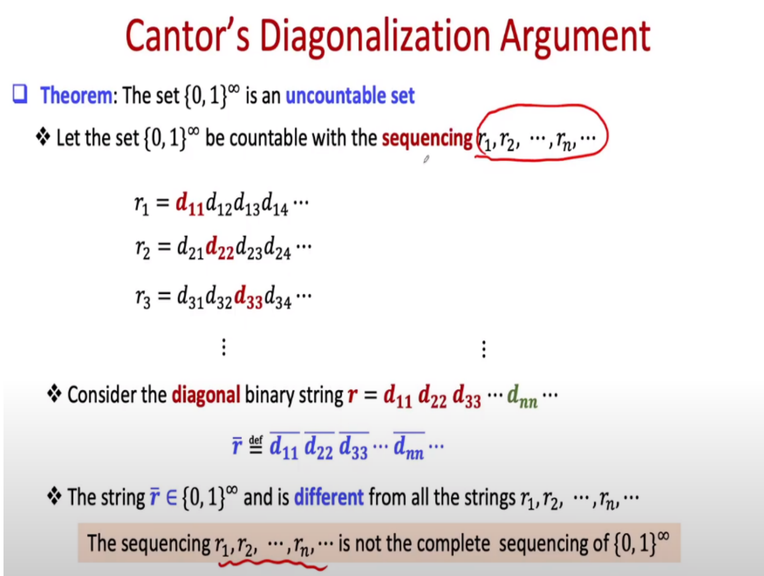
So that is why at the second position I have bit 1, the integer 3 is prime that is why at the third position I had the bit 1, the integer 4 is not prime that is why I have at fourth position the bit is 0 and so on. And again this is an example of an infinite length binary string. So before proceeding further you might be wondering is there any difference between the set of all binary strings of finite length, namely the set $\{0, 1\}^*$ and the set that we are considering right now, namely the set of binary strings of infinite length. It turns out that indeed these two sets are completely different. The difference is in the terms of the length of the strings in the individual sets. So when it comes to the set of; first of all both the sets has the infinite number of sets remember that. Even if I consider the set $\{0, 1\}^*$, the number of the number of strings in that set is infinite.

However the length of each string in that set will be finite. So the difference, the primary difference between the 2 sets is the following. If I consider the set $\{0, 1\}^{\infty}$ then the property of the set is that the length of any string in this set cannot be bounded by a natural number. Because you take any string in this set we can never say what will be the end digit or end bit of the string because the sequence of characters in each string in the set will be an infinite sequence.

Whereas if I consider the set $\{0, 1\}^*$ then the property of this set is that each string in the set is of finite length. That length might be arbitrary large, it might be enormously large positive number but it is a bounded quantity. That means we will know that it starts with certain bit and it ends with a certain bit. It is not the case that its end bit is not known. So that is the primary difference between these 2 sets.

That is why you can see here in the examples that I have listed down. We do not know what exactly is the end bit of these strings. That is why we have dot written down here.

(Refer Slide Time: 05:43)



So now what we are going to discuss is a very beautiful result, very fundamental result attributed to Cantor is called a Cantor's Diagonalization argument and using this diagonalization argument is we are going to prove is that the set of all binary strings of infinite length is an uncountable set. That means we cannot enumerate out or we cannot list down the elements of this set. So the proof will be by contradiction; we will use a proof by contradiction mechanism here.

So we are supposed to prove that this set is an uncountable set. But we believe the contrary, we assume the contrary and we assume that the set is countable and if it is countable then it must be having a sequencing of elements of the set. Imagine that the sequence is this r_1, r_2, r_n and so on that means we know that what is the first element in this set, the second element in the set and so on.

So remember each element in the set $\{0, 1\}^{\infty}$ is a binary string of infinite length. So imagine r_1 is of this form : it has first bit d_{11} second bit d_{12} third bit is d_{13} forth bit is d_{14} and so on. And it is an infinite length string because it is the number of $\{0, 1\}^{\infty}$. Similarly imagine that the bits of the strings r_2 are d_{21}, d_{22}, d_{23}, d_{24} and so on. Similarly the bits of the string r_3 are d_{31}, d_{32} and so on.

And since the sequence is an infinite sequence because there are infinitely many elements in this set, this list of elements of the set $\{0, 1\}^{\infty}$ will go on and that is why I will have dot here. Now what I am going to do here is I have to arrive at a contradiction the way I am going to arrive the contradiction is that I will show that there exist at least 1 string which is of infinite length and which is binary and which is not there in the sequence of binary strings that we are assuming exists.

We are assuming that we have a sequence r_1, r_2, r_n and so on and that sequence is the enumeration of all the elements in the set $\{0, 1\}^{\infty}$. But what I am going to show is, I am going to show the existence of one string which is going to the missed in the sequencing which will show that the sequencing which we are assuming to exist does not exist actually. So what, exactly that string, So you consider the binary string r which is obtained by focusing on the bits along the diagonal here.

So remember this diagonal is an infinite diagonal because I have more elements to follow and the sequencing I am just for r_4 I am consider d_{44} and so on. So I am considering the diagonal binary string, its is an infinitely long binary string. And what I consider here is now I consider a new string which I denote as $\bar{r}$ and $\bar{r}$ is obtained by complementing each of the bits in this diagonal binary string.

So d_{11} is 0, I take its complement. So this is denoted by $\overline{d_{11}}$ and so on. Now my claim is that the string $\bar{r}$ which I have constructed here which I am considering here definitely is a member of the set $\{0, 1\}^{\infty}$ why? First of all it is a binary string and its length is infinity. Because the diagonal here which I am considering here; the diagonal binary string is of infinite length. So the complement of that string is also will be infinite length.

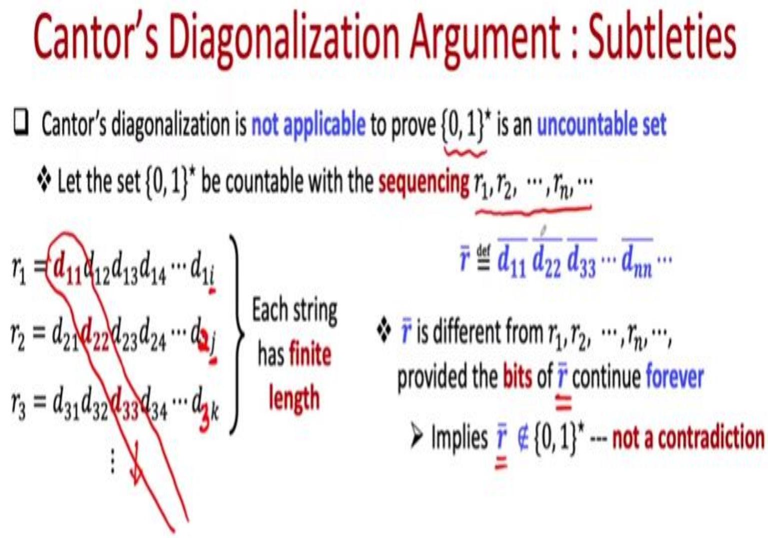
So definitely $\bar{r}$ is the member of the set $\{0, 1\}^{\infty}$. But the interesting thing here is that the string $\bar{r}$ I am considering here will be different from all the string r_1, r_2, r_3, r_4 and so on in the sequencing which I am assuming to exist. So you can verify that. So if I consider the first string r_1 in my list, $\bar{r}$ is definitely different from r_1 because the first bit of $\bar{r}$ and first bit of r_1 are different.

They are complement to each other. Similarly if I consider the second string r_2 in my sequencing it will be different from $\bar{r}$ because the second bit of $\bar{r}$ will be different from r_2. And this process will continue; you take any string in the sequencing which you are assuming there will be at least 1 bit in that string in the sequencing which will be different from the corresponding bit in r compliment or $\bar{r}$.

So that shows that the sequencing that we are assuming is not the complete sequencing of the set or the complete sequencing of the elements of the set, $\{0, 1\}^{\infty}$. We are definitely missing some elements from the set $\{0, 1\}^{\infty}$ which we are not writing out. And that shows that your set $\{0, 1\}^{\infty}$ is an uncountable set because as per the definition countably infinite set if the set is countably infinite there must be some valid sequencing some sequencing of the element of that set.

So it does not matter what is the sequencing, if you show me any sequencing for the set $\{0, 1\}^{\infty}$ I will show you the existence of 1 string which will be missed in that sequencing. It will show that no sequencing of the elements of the set $\{0, 1\}^{\infty}$ is possible.

(Refer Slide Time: 11:59)



So there are few subtleties involved here when running the Cantor's diagonalization argument. It might look to you that why I cannot Cantor's diagonalization argument to even prove that the set {0, 1}* is also uncountable. So {0, 1}* remember is the set of all binary strings, there will be infinitely many elements in the set {0, 1}* but the length of each string in this set will be finite. So let us see where exactly the Cantor's diagonalization argument fails when we try to run it for {0, 1}*.

So we will start assuming that we have a sequencing for enumerating out the elements for the set {0,1}* where the elements where the listing is r_1, r_2, r_n and so on. So again I will focus on the individual bits of r_1, r_2, r_3, and so on. The important point here to note is that since each of the strings r_1, r_2, r_3 are the members of the set {0, 1}* their length will be finite. I will know that d_{1i} is a bit. I will know that i is, actually some number, some natural number, some positive number.

It is not the case that the bits of r_1 will keep on going forever. In the same way if I consider the string r_2 I know j is a natural number. Sorry for the typo here this should be d_{2j} and similarly this should be d_{3k} and so on. So when I consider the bits of the string r_2, I know that there are finitely many bits in r_2 that means j is a natural number. Similarly for r_3 I know there are finitely many bits in r_3 that means k is a natural number and so on.

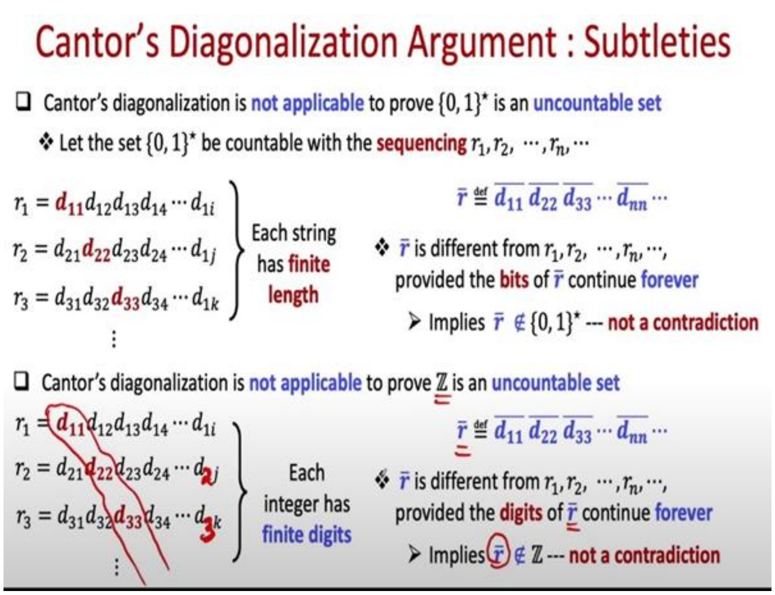
Although the number of elements in the set {0, 1}* is infinite that is why the sequencing will keep on going forever. So as per the Cantor's diagonalization argument I will consider the binary string are compliment. It will be the compliment of the diagonal string here. So the diagonal string will be like this and it will continue forever. Definitely $\bar{r}$ that I have constructed here will be different from each of the strings r_1, r_2, r_n in the sequencing, provided the bits of r complement or $\bar{r}$ continue forever.

That means only when $\bar{r}$ goes forever that means the bits of the $\bar{r}$ keeps on going and going never ends then only I can claim that the $\bar{r}$ is different from each of the strings r_1, r_2, r_n in my sequencing. But notice that is the case when string $\bar{r}$ which I am constructing here which is not an element of {0, 1}*. It is because if at all $\bar{r}$ belongs to {0, 1}* then its length has to be finite.

I cannot say that the bits of the string $\bar{r}$ continue forever. That is the characteristic of binary string which has infinite length. But if at all $\bar{r}$ is of finite length that means it stops some where and if it is of finite length definitely it will be appearing somewhere in my sequencing and that is where the Cantor's diagonalization argument fails. Whereas when if we consider the argument for the set $\{0, 1\}^{\infty}$ there was no restriction on the bits of $\bar{r}$ it was allowed to go forever.

So, that is the point where the Cantor's diagonalization argument fails when we try to run it for the set {0, 1}*. Let us see another subtlety to make my point more clear.

(Refer Slide Time: 16:06)



So let us try to see whether we can run the Cantor's diagonalization argument to prove that the set of integers is an uncountable set. Remember it is a countable set and we had shown sequencing, we know how to sequence, list down the elements of the set of integers. So as per the Cantor's diagonalization argument the proof will be contradiction and so I will assume that I have sequencing for listing down all the integers.

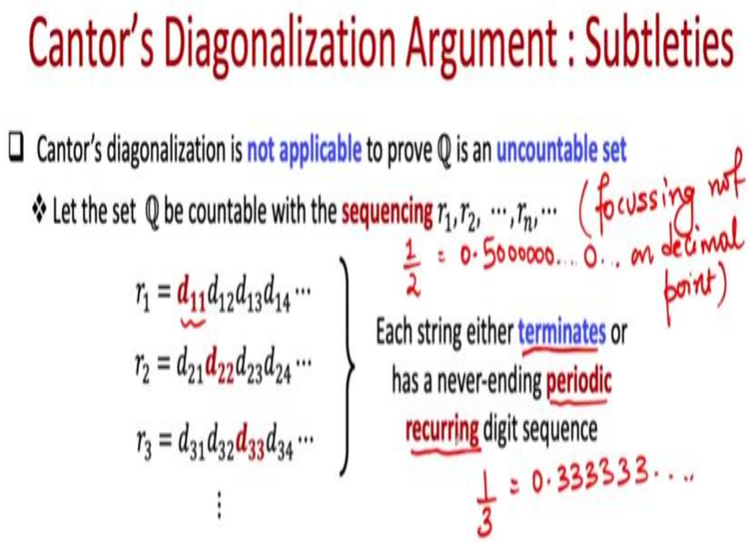
So r_1, r_2, r_3, r_n will be the elements in the sequencing and then each integer can be considered in terms of its decimal representation, namely the digits that we have in that integer. So imagine that digit of r_1 are d_{11}, d_{12}, d_{13}, d_{14}, d_{15} namely r_1 has i number of decimal digits. Similarly say r_2 the integer r_2 has j number of decimal digits. Again sorry for the typo error this should be d_{2j}.

And similarly assume that r_3 has k number of decimal digits and so on. Again remember each integer has a magnitude. When I say it has a magnitude that means its possible only when the number of digits in an integer is finite. You cannot have an integer which has infinitely many digits in its decimal representation. That is not a valid integer at all because you do not know what exactly will be the magnitude of that integer.

So each integer will have finitely many digits that might be arbitrary large that is the different thing. And the number of elements in the set of integer it might be infinite that also fine. But the property is that the length or the number of digits in each integer will be finite. So again if I consider the diagonal digits here and flip them and obtain new string of decimal digits say $\bar{r}$. Then I can say that $\bar{r}$ is different from all the integers in my sequencing provided that the digits of $\bar{r}$ continue forever.

Then only I can say that the $\bar{r}$ is different from r_1 and r_2 and r_3 and r_4 and every integer in my sequencing. But if that is the case that means if I allow the digits of $\bar{r}$ to continue forever then that is not a valid integer because every integer has to stop after certain number of digits. That might be enormously large quantity but it has to stop somewhere. I cannot have an integer whose digits continue forever. So the resultant sequence of the digits $\bar{r}$ which I will construct as per the Cantor's diagonalization argument will not be a member of the integer set.

(Refer Slide Time: 19:14)



So i am not getting a contradiction that I now have a new integer which is not there in my sequence. In the same way I cannot run Cantor's diagonalization argument to prove that the set of rational numbers is uncountable. Remember the set of rational number are countable even though it is an infinite set and we know how to enumerate or list down the element of the set of rational number. Again let see where exactly the Cantor's diagonalization argument will fail for the set of rational numbers.

So we will assume that we will have sequencing or listing of the set of rational numbers and let r_1, r_2, r_3, be the rational numbers, listing of the rational numbers as per my sequencing. So again what I have done here is I am not focusing here in my sequencing. I am not focusing on the decimal point, that is just for simplicity. I am focusing on the remaining part whatever is appearing after the decimal point in the decimal representation of your rational numbers and focusing only the digits here and not focusing on the decimal points here.

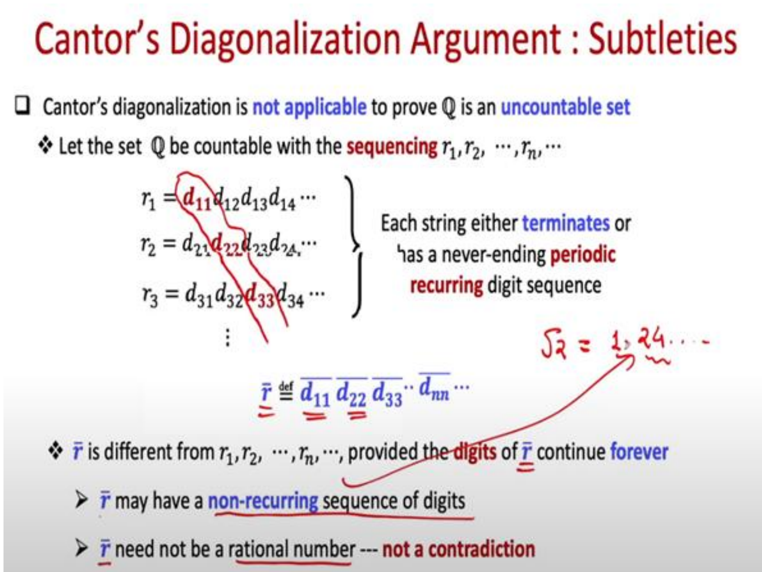
That is just for simplicity; so again the digits of the first rational number I am assuming them to be d_{11}, d_{12}, d_{13}, d_{14} and so on. So now in this case what is happening is, the number of digits in the decimal representation of any rational number might be finite or it might be infinite. It might be finite in the sense it might terminate after some point.

Say for instance if I consider the rational number say 1 over 2. Then 1 over 2 is 0.5 so it terminates somewhere. But 0.5, I can imagine as 0.500000 followed by infinite number of 0. So even if the decimal representation terminates after certain positions or the given rational number I can imagine that I append it by infinite number of 0 and hence the, number of digits continue forever. Whereas I may have rational number where in the decimal representation the number digits never terminates.

But it will have a never ending periodic recurring digit sequence. So for example if I consider the rational number 1 over 3 then 1 over 3 is 0.333333333 never terminates. So that means when I am writing down the decimal representation of the rational numbers in my sequencing I can imagine that for every rational number in its decimal representation the digits continue forever.

But if it terminates it continues forever or it does not terminate in that case it will be a never ending recurring sequence. It would not be a case that it does not terminate and simultaneously it does not recur that will not be the case for a rational number. If at all, the decimal digits does not terminate then it will be definitely a recurring somewhere.

(Refer Slide Time: 23:07)



So now, again it will focus on the digits along the diagonal entries here and we will flip them. We will flip them in the sense that what you can imagine is that the $\overline{d_{11}}$ here represents any digit different from d_{11}. So for the notion of the compliment make sense in the context of binary strings here but we can generalize it here and assume that $\overline{d_{11}}$ represents any digit different from d_{11}.

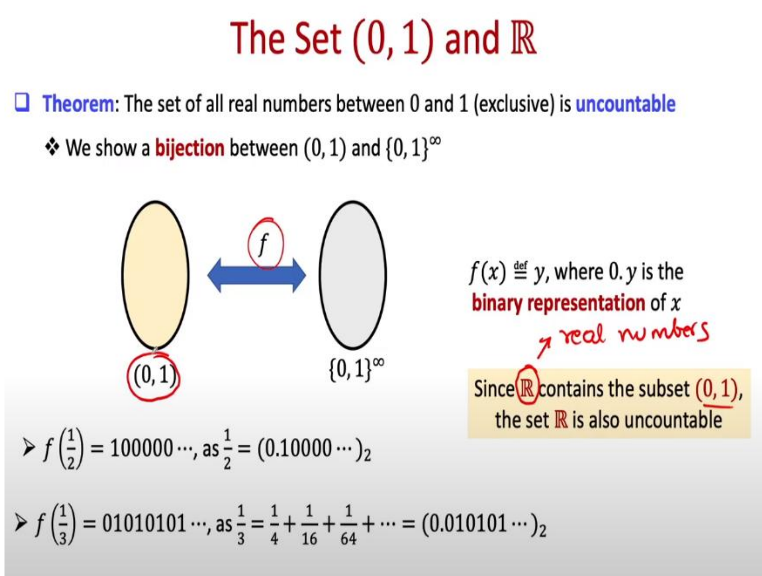
Similarly $\overline{d_{22}}$ represents any digit different from d_{22} and so on. So that will be new sequence, a new string of decimal digit which we have constructed here. Definitely this new string of decimal digits which we have constructed will be different from all the sequence of decimal digits in your sequencing provided the digits of this $\bar{r}$ continue forever. Now the question is can I say that this $\bar{r}$ definitely is a rational number.

Only in that case I can arrive at a contradiction. But the point here is that there is no guarantee that the sequence of digits in this string $\bar{r}$ has a periodic recurring digit sequence. So what I am saying is that it may be possible that in $\bar{r}$ you have a non-recurring sequence of digits. That means even though the sequence of digits does not stop we do not get any recurrence or any periodic recurrence or you do not get any recurrence of periodic recurring sequence in $\bar{r}$. That means, this $\bar{r}$ need not be a rational number because we do not know whether all these digit $\overline{d_{11}}$, $\overline{d_{22}}$, $\overline{d_{33}}$, $\overline{d_{nn}}$ they are distinct or they are going to be repeated. We do not have any guarantee what so ever. So that is why we do not get the contradiction, we do not get the guarantee that this string of digits $\bar{r}$ represents a rational number.

It may represent an irrational number so for instance if you consider √2 is a rational number then you know that if I consider the decimal representation of √2. And if I focus on the sequence of the digits in the decimal representation of √2. Then it is a never ending sequence and you do not have any recurring or any periodic digit recurring sequence in this sequence of decimal digits.

So your $\bar{r}$ may be a sequence like this, you do not know. So you cannot say that the $\bar{r}$ that you are obtaining here is definitely a rational number. So again Cantor’s diagonalization argument fails if we try to apply it to the set of rational numbers.

(Refer Slide Time: 26:20)



So now we know at least one set may be the set $\{0, 1\}^{\infty}$ which is not countable. Now we will see some other sets as well which are not countable. So what we are going to show here is first the set of real numbers between 0 and 1 but excluding 0 and 1 is uncountable. So the set is denoted by (0, 1) so this is the representation of the set of all real numbers between 0 and 1 excluding 0 and 1.

So how I am going to show it is uncountable? Well I have already shown that the set $\{0 , 1\}^{\infty}$ is an uncountable set. I will show you now a bijection between the set $\{0, 1\}^{\infty}$ and the set of all real number between 0 and 1. That will automatically show that; conclude that the set of all real numbers between 0 and 1 has the same cardinality as the set $\{0, 1\}^{\infty}$.

So what is the bijection? Bijection is very simple. So if I take any x, any real number between 0 and 1 excluding 0 and 1 that will have a binary representation. So let the binary representation of that real number be 0.y where y is a binary string. So what will be the function f? The function f(x) will be y, that means I will just chop off the 0 here and the point here and just I will focus binary representation that means the bits in the representation y and that will be the mapping of x as per the function.

So how exactly this function will look like. So if you consider say for instance x = 0.5. So since the mapping of 0.5 will be the binary string 1 followed by infinite number of 0's why? Because 1 over 2 in binary; can represented as 0.1 and I can always put infinitely many 0 even though 1 over 2 is 0.1. So this will be my y and that is why f(x) will be mapped to I chop off the 0 at this point.

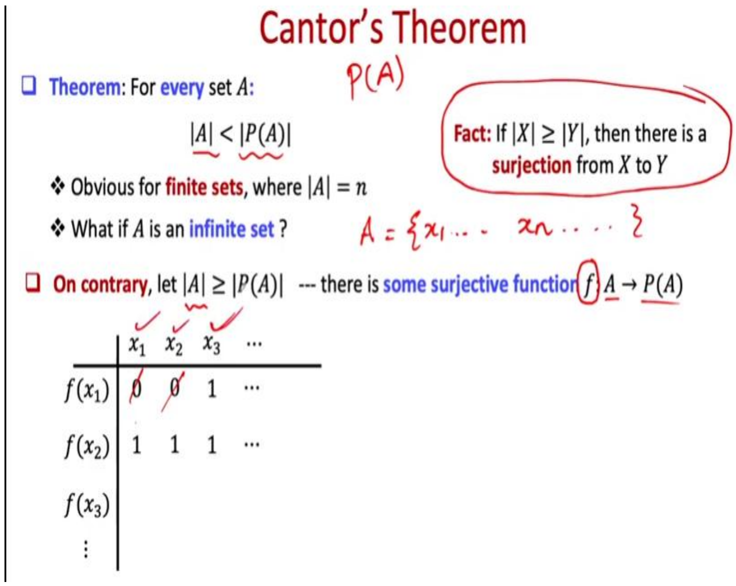
In the same way if I consider x = 1 over 3 then 1 over 3 can be represented as this infinite sum 1 over 4 + 1 over 6 + 1 over 64 and so on and this an infinite sum and it is in a geometric progression, if you apply the rule of summation of infinite geometric series you will get the sum that is nothing but 1 over 3. But now if I focus on binary representation of 1 over 4, 1 over 16 and so on, this will be a binary representation.

So anything after the binary point here it will my y and I will chop off this and that is why my x will be now mapped to this y and so on. So it is easy to see that this function f is indeed a bijection and I am leaving that as an exercise for you. Because the simple fact is you take any x, any real number it will have unique binary representation that is all that is a simple observation here.

Now if I consider the set of real numbers, this set R denotes the set of real numbers then it contains the subset (0, 1) it also includes all the real number between 0 and 1 also. And since (0, 1) the set of all real numbers between 0 and 1 is uncountable and remember we had argued that, if we had a set with a subset which is uncountable then the whole super set will also be uncountable. That automatically shows that the set of real numbers is also uncountable.

And intuitively the main reason that the set of real numbers is uncountable is that it has irrational numbers as well which we cannot enumerate out.

(Refer Slide Time: 30:54)



So Cantor proved a very interesting result as well. He showed that you take any set A then the cardinality of that set is strictly less than the cardinality of its power set. So remember the notation P(A) denotes the power set of A. Where the power set is the set of all subsets of that set. So of course this statement is true if your set of A is finite namely if your set A has n number of elements then its power set will have 2^n elements.

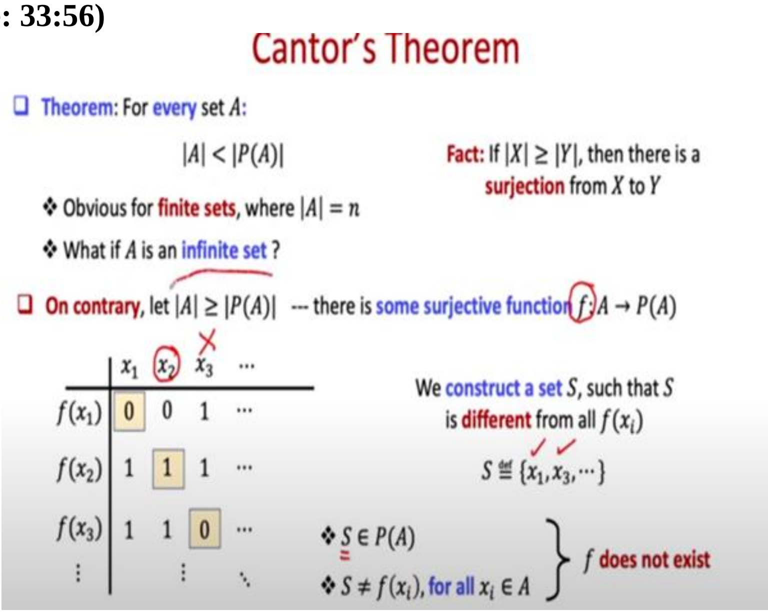
And we can always prove that n is always strictly less than 2^n. What if A is an infinite set can we conclude that this theorem is true that is even for infinite set and Cantor showed yes, so the proof is again is contradiction and we will run the diagonalization argument here as well. So we will assume that: let the cardinality of the set A be greater than equal to the cardinality of its power set.

Now before proceeding with fact which we will be using in this proof is the following. If you have the sets X and Y and if the cardinality of X is greater than equal to the cardinality of Y then there always exist a surjection from X to Y. This is a very simple fact which you can prove very easily. So I am not going to the proof of that; we are going to utilize this fact in this proof. So I am assuming here that the cardinality of A is greater than equal to the cardinality of its power set.

That means there will be some surjective function from the set A to the power of set of A. I do not know what exactly is the structure of the surjective function but I denote the surjective function by f. So now what I have done here is let the elements of A be x_1, x_2, x_3 and x_n and so on. It is an infinite set, so it has infinitely many elements, so I am assuming that elements of set A can be listed down as x_1, x_2, x_3 and so on. And I have listed down $f(x_1)$, $f(x_2)$, $f(x_3)$ and so on.

So each of the f value is nothing but a subset of A set that means it will be the element of the power set. So depending upon which elements from the set A are present in $f(x_1)$ accordingly I have put the entry 0's and 1's. So for example here I mean to say that $f(x_1)$ it is a set which does not have x_1, it does not have x_2 but it has x_3 and so on. Similarly $f(x_2)$ is a subset which has x_1 , it has x_2, it has $x_3$and I have listed down $f(x_1)$, $f(x_2)$, $f(x_3)$ and so on. That is the interpretation of 0's and 1's in the table,

(Refer Slide Time: 33:56)



Similarly $f(x_3)$ is the subset which has x_1, which has x_2 but it does not have x_3 and so on. Now what I am going to show is since I am assuming that the function f is surjective function and I have to arrive at a contradiction. I will show that this f actually does not exist and how do I show that f is does not exist? I have to show that f is actually not a valid surjective function. To do that I will show you a subset which belongs to the powerset namely I will show you a subset of A set which do not have any pre image.

That means it will be different from, $f(x_1)$, it will be different from $f(x_2)$, it will be different from any $f(x_i)$ which shows that the set S, the subset S do not have any pre image, hence contradicting that function f is a subjective function. So how do I construct that subset S; again I run the diagonalization argument, so I focus on the diagonal entries here and the elements of my set S will be constructed depending upon the diagonal entries.

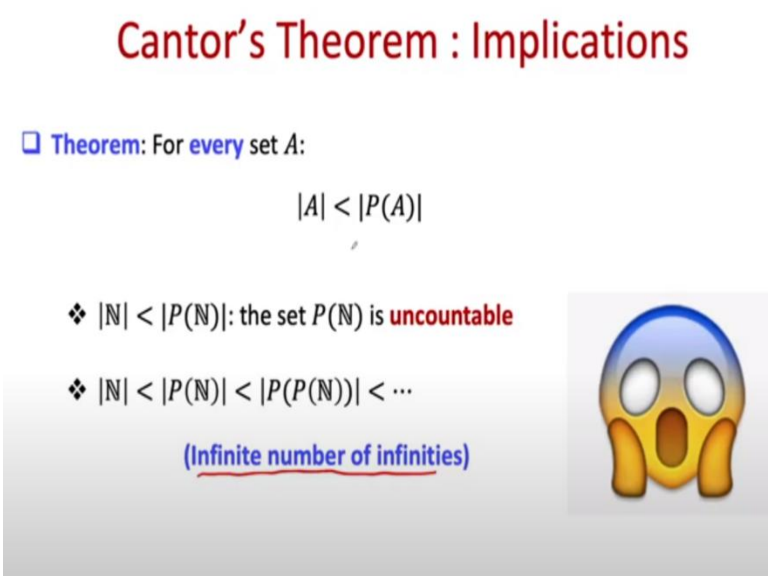
So in this example so the first diagonal entry is 0 so I will include x_1 because here the entry is 0 so I will flip it and I will make it 1 that means I have to include x_1. The second entry is 1 along the diagonal 1 so I will flip it and make it 0 that means I have to exclude x_2. The third entry along the diagonal is 0 so I will flip it and will make it 1 that means I have to include x_3 and so on. So that is the way I have constructed set S here.

So now you can check here that indeed the set S I have constructed will be an element of the power set because it is a subset, it will have some of the element from the A set. I am not taking the element in the S set some from outside. So that is why it will be the element of the power set. But you can check here that the S will be not equal to $f(x_1)$ that means the set S will have at least 1 element which is not there in $f(x_1)$.

So for example x_1 is present in S but x_1 was not present in $f(x_1)$. Similarly the set S will be different from $f(x_2)$ why? Because $f(x_2)$ has x_2 but I have not included x_2 in S. Similarly $f(x_3)$ will be different from set S why? Because $f(x_3)$, does not have x_3 but I have included x_3 in S. So the way I have constructed the set S it will be different from the image of $f(x_1)$, $f(x_2)$, $f(x_3)$ and so on.

That means S will not be the image of any x_i and hence S is not a valid surjective function. That means whatever I assumed here namely I have assumed existence of the surjective function from the set A to its power set which is not a valid assumption. This is not a valid assumption because I made here a wrong assumption that the cardinality of the set A is greater than equal to the cardinality of its power set. So that means indeed the statement in this theorem is the correct statement.

(Refer Slide Time: 37:32)



So what is the implication of the Cantor's theorem it has a very beautiful implication. So this is the statement of the cantor's theorem. So if I apply it over the set A being the set of positive integers or the set of natural number then I obtain the fact that the cardinality of the set of natural number is strictly less than the cardinality of its power set. That means the cardinality of the set of natural number is $\aleph_0$.

But what I am showing here is that its $\aleph_0$ is strictly less than the cardinality of the power set of the natural number. That means the power set of the set of natural number is uncountable. Now if I treat the power set of natural number as the set A then the power set of this power set will have more cardinality and this process will keep on going forever. So what basically Cantor showed is that there are infinite number of infinities.

You do not have only one infinity so $\aleph_0$ is one of the infinity it is one of the infinite quantities. But you can have now an infinite number of infinities because now you have a hierarchy of infinite quantities. So that is the very interesting fact about the cardinality of infinite sets. So that brings me to the end of this lecture. Just to summarize this lecture we saw some uncountable sets and we proved that those sets are uncountable by using Cantor's diagonalization argument. Thank you.