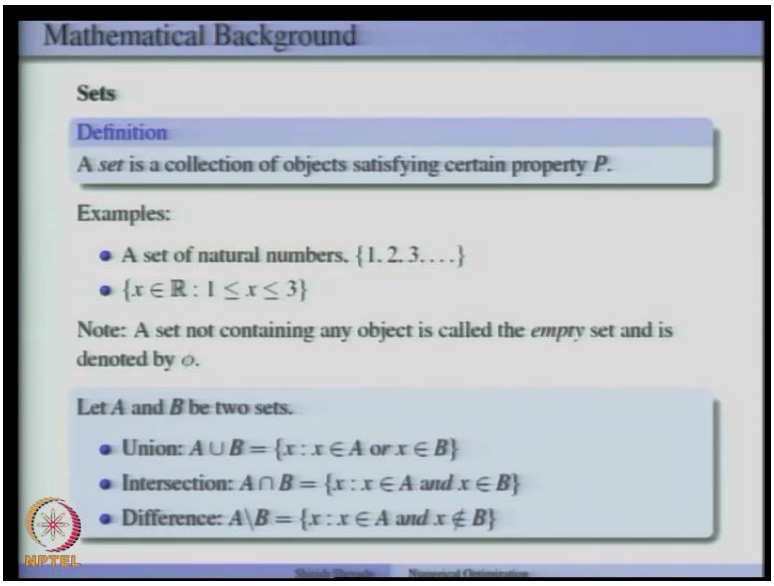
Numerical Optimization
Shirish K. Shevade
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Lecture - 2
Mathematical Optimization

Welcome back to this series of lectures on Numerical Optimization. So, in the last lecture we discussed about, how to formulate optimization or a mathematical programming problem. And we also saw some ways to solve some of those problems using graphical method. So, in this course we will mainly you know worry about solving mathematical optimization problems using either analytical methods or numerical methods.

Now, all these optimization methods will require some background some mathematical background on sets, linear algebra, differential calculus. So, in the next one or two lectures, we will spend some time studying about some of the background needed, some of the mathematical background needed for this course. So, I will try to give some important results which will be used in this course and for some details one can always refer to some of the references given which will be given at the end of this lecture.

(Refer Slide Time: 01:34)

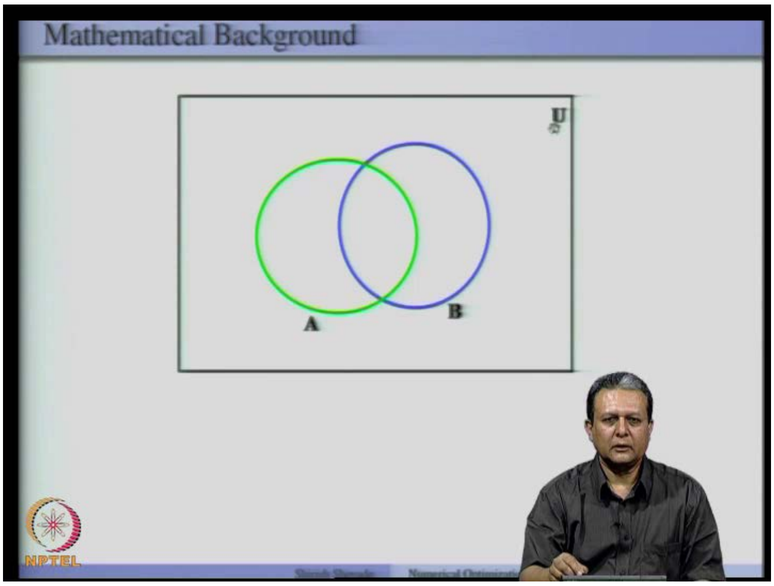


So, let us start with some mathematical background, we will start with the definition of sets as all of you know that a set is a collection of objects which satisfy certain property. So, for example, we can have a set of natural numbers, which take the values 1, 2, 3 and so on. One could have a set of real numbers or one could have as mentioned here, a set whose elements are real numbers and the elements take the values in the range 1, 2, 3 both 1 and 3 inclusive.

So, so, this is the property that needs to be satisfied the satisfied by the elements of the set. Now, in this case the property is that the elements have to be natural numbers. So, one can define a set using a certain property, now there could be some cases where set does not contain any object, such sets are called empty sets and we will denote them by the letter phi. Now, let us look at some operations on sets. So, suppose we consider two sets A and B, which are part of some universal set U and the union operation is defined like this where we take either element of x or element of B. So, this the elements x will satisfy this property that that will be denoted as A union B.

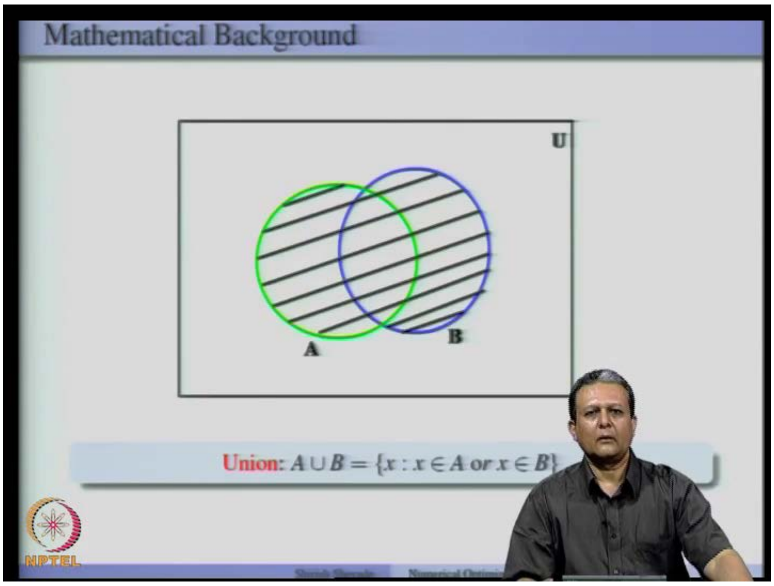
Now, similarly we have the intersection of two sets A and B denoted like A intersection B, where the elements of the set have to belong to both the sets. So, the intersection A of A and B is set of those elements which are both in A and B. So, the difference between A and B is also be can be called A minus B sometimes denoted by slash backslash. So, this is the set of elements of x which a set of elements of A which do not belong to B. So, if you collect all such elements of A which do not belong to B that will give us a difference between A and B. Now, we will show this using some figures.

(Refer Slide Time: 03:59)



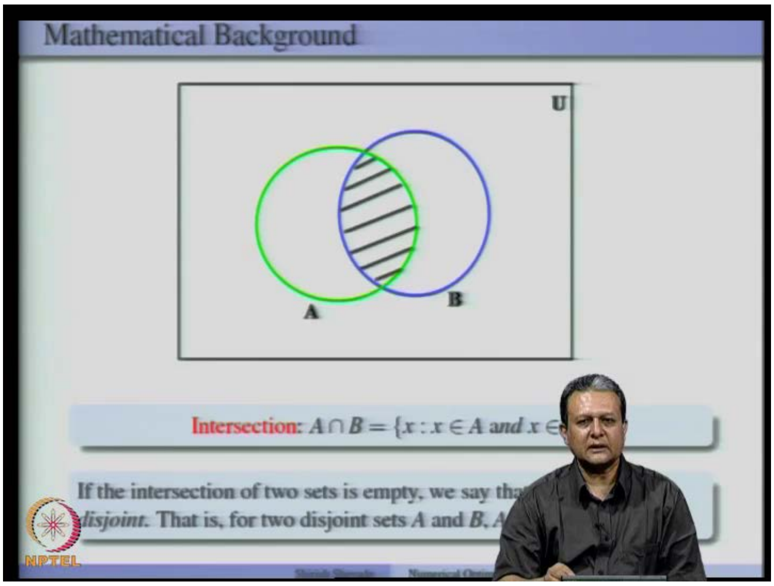
So, let us consider two sets A and B, which are part of the universal set U. So, the set A consist of all the elements inside this circle green circle and the set B consist of all the elements which are inside this blue circle. So, these are the given sets A and B.

(Refer Slide Time: 04:23)



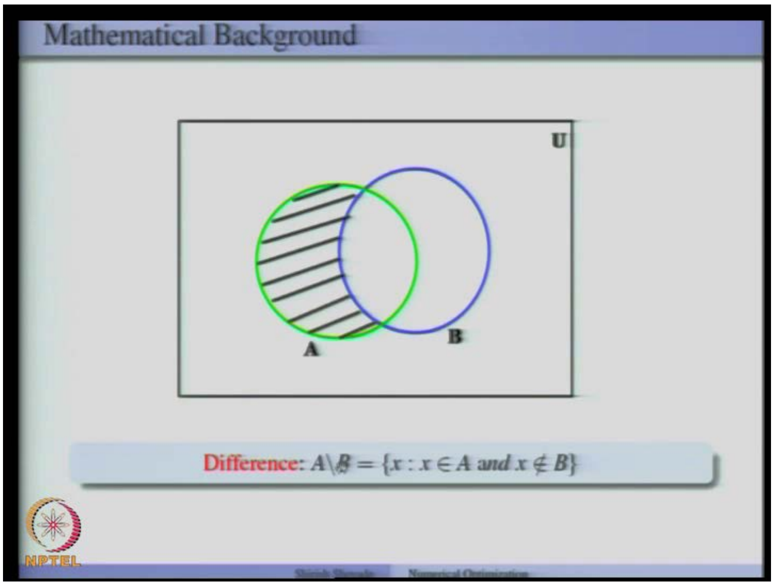
Now, now let us look at the union operation, so the A union B is basically the set of all elements which either belong to A or belong to B. So, the shaded region here tells you that, this is the union of this is the union of two sets A and B.

(Refer Slide Time: 04:48)



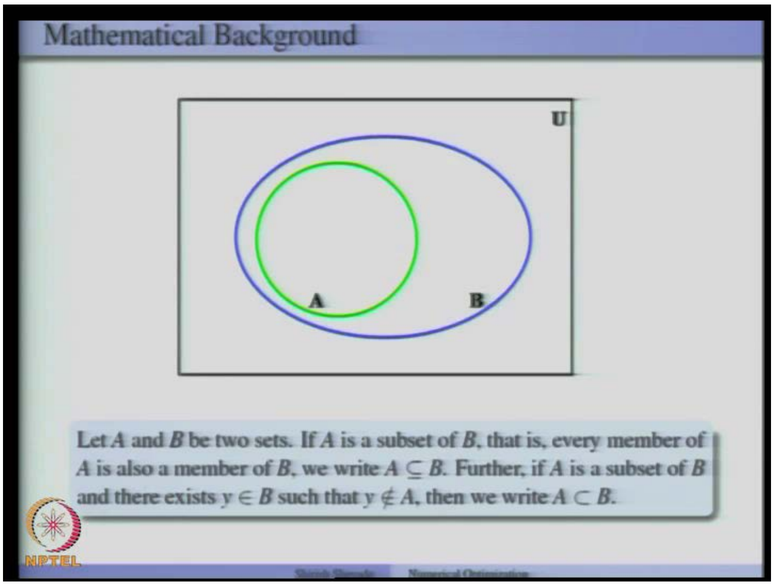
Now, let us look at the intersection. So, intersection is essentially those elements which are which belong to both A and B. So, you will see that the the shaded region here denotes the elements, which are part of the set A as well as the part of set B, so this is called a intersection. Now, if the intersection of two sets is empty, we say that the sets are disjoint. So, for example, set A is some were here and the set B is on the other side and they do not have anything in common then we say that they are disjoint, and for disjoint sets we say that A intersection B is a null set.

(Refer Slide Time: 05:37)



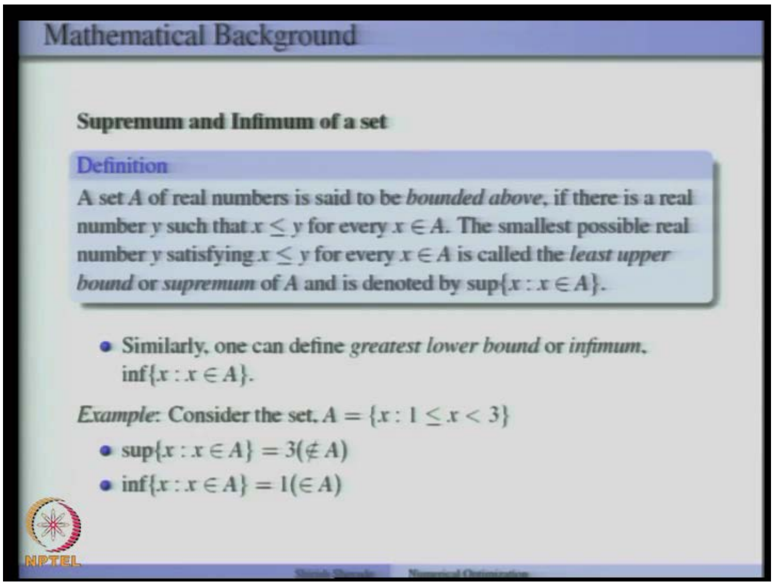
Now, let us look at the difference between A and B. So, the shaded region here denotes those the elements of the set A, which do not belong to the set B, so this is called the difference between the set A and a set B. Similarly, if you want if you want to find out the difference between B and A, it will be the other part of the shaded region the elements of B, which are not part of A. So, this is called the difference between the sets A and B and will use backslash to denote that difference.

(Refer Slide Time: 06:13)



So, suppose we have A and B as two sets, which are given like this. Now, we say that A is the subset of B that is every member of A is also a member of B, and we write this as A subset of B. Now if it, so happens that there exist some elements in B, which are not part of A for example, some elements here are not part of A. So, if such a thing happens then we say that A is a strict subset of B. So, if A is a subset of B and if there exist some element y which, belongs to B such that y does not belong to A; then we write A as a subset of B a strict subset of B. So, the strict subset will be denoted by this symbol.

(Refer Slide Time: 07:05)



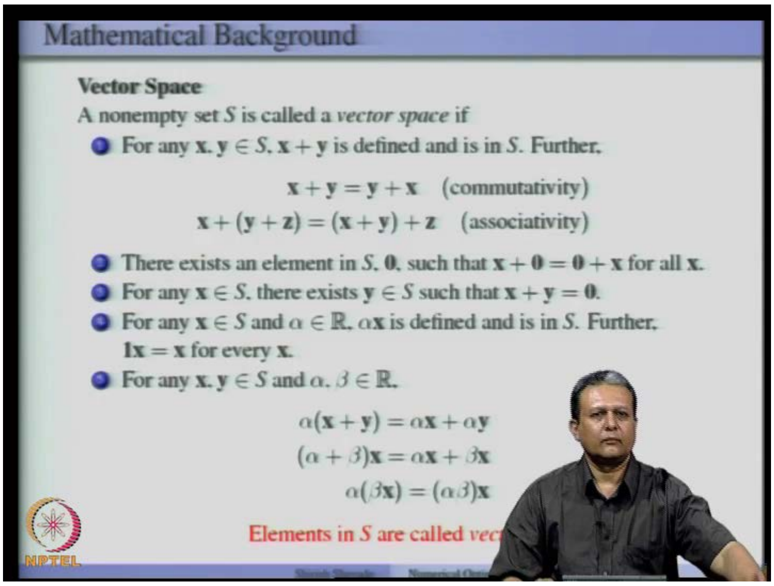
Now, we look at the supremum and infimum of a set, now let us consider a set of real numbers. So, that set let us consider a subset of real numbers; now that set is said to be bounded above, if there is a real number y such that every element of x of A is less than or equal to y. Now, among all possible such y's if you find the smallest possible real number such that x is less than or equal to y for every x in the set A, then it is called a least upper bound or supremum of A and is denoted by sup x such that x belongs to A.

So, you will see that, we first find the any number y such that x is less than or equal to y for all x belongs to A. And then among all such y's if you take the smallest possible real number y such that this property holds that for every x belongs to A, x less than or equal to y then we get the supremum of the least upper bound of A. Now. along similar lines one can define what is called greatest lower bound or infimum.

So, in that case one has to look for the element y such that x is greater than or equal to y for every x belongs to A, and among them if you find out the largest number such that x is greater than or equal to y for every x belongs to A then it is called a greatest lower bound or infimum. Now, here is an example, so let us consider a case set S set A where 1 is the elements of the set satisfy this property where 1 is less is than or equal to x x less than 3. Now so, which means that we include 1 in the set, but we will not include 3 in the set, but any real number less than 3 is and greater than or equal to 1 is permitted in the set.

Now, if you take the supremum of this set, now any number greater than 3 is a upper bound for this set, but then among all those numbers, what is the least upper bound and the least upper bound or supremum is three. And you will note that this 3 does not belong to the set A, so a supremum need not belong to the set. And similarly one can define the infimum of this set and it turns out that the infimum that is the greatest lower bound of the set is one and in this case it does belong to the set A. So, supremum and infimum, they need not belong to the set they can belong to the set depends on the definition of the set.

(Refer Slide Time: 10:07)



Now, we move on to the concept of vector space. So, we have so far studied the sets, so some of those concepts could be useful in studying the vector spaces. Now, a nonempty set S is called a vector space if it satisfy certain property. So, let us look at some of those properties in detail. Now, for any two numbers x and y of the set S, first of all there is a operation plus which is defined and such that x plus y always is in the set S.

Now, further the this addition operation should be commutative for example, x plus y should be equal to y plus x, and also it should be associated for example, if you take y plus z and the resultant element. If you x it is same as adding x and y first and then adding the element set, so this is called associativity. So, for any x y, the addition operation is defined and is in the set S x plus y is in the set S and commutivity and associativity holds.

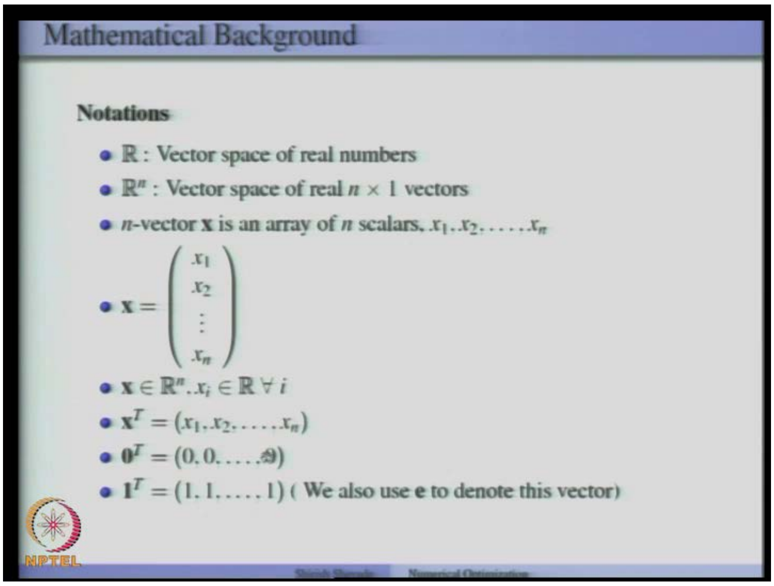
Now, further there should exist an element in S, which is called let us denoted by 0 such that x plus 0 is equal to 0 plus x for all x in the set S. So, this 0 is called identity for addition so that means, if you add that identity element to this to any element x you will get x. Similarly, the there exists for any x belong to the set S there exists y such that x plus y is equal to 0, so this y is called the additive inverse of x. So, the element when it it is added with additive inverse you get the identity for the addition.

Now, for any x belongs to S and alpha is set of real numbers alpha x should be defined and that should be also in the set S and one this this is the one one scalar one multiplied by x should gives as x for every x. So, this is the called multiplicative identity, so this is identity for the 0 is the identity for the addition operation and one is the identity for the multiplication operation. Now, here I have mentioned alpha belongs to R in fact for a vector space alpha can come from any field, but for this course will be restricting ourselves only to the field of real numbers, so that is why I have mention alpha to belong to R.

Now, further for any x y in the set S and alpha beta from the field of real numbers. So, alpha, so if you find out x plus y and then multiplied by alpha. So, it is same as finding alpha x plus alpha y. So, multiplication distributes over addition and similarly alpha plus beta combined together multiplied by x is same as alpha x plus beta x. And if you multiply a vector x by beta and then by alpha it is same as good as multiplying alpha and beta together and then multiplying them by the vector x.

So, all these properties should hold where x y or any two elements of the set S and alpha beta they come from the field of real numbers. Now, these elements of the set S they are called the vectors. So, the vector space is the space formed using vector now there are some standard examples of vector spaces for example, the space of real numbers n dimensional space of real numbers, two-dimensional space of real numbers these are some standard examples of vector spaces.

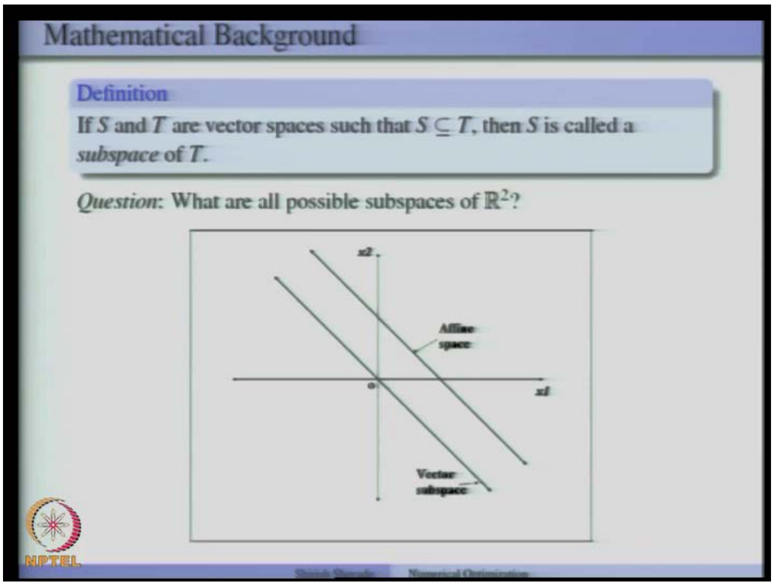
(Refer Slide Time: 14:51)



Now, here are some notations that, we will use in this course of course, some of them I have already used, but let me specify them in detail. So, R denotes the vector space of real numbers and R n is a vector space of real n dimensional vectors. So, an n dimensional vector x can be written as a column vector consisting of n elements, now each of these elements is a real number. So, x belong to R n means that each of the x i's belong to R and there are n such x i's I going from 1 to n. Now, the transpose of a vector will be denoted by a R o vector. So, the vector will typical will we will denoted vector by column vector and transpose by a row vector.

Now, there are some special cases where we have 0 vector. So, 0 vector is a vector which contents all zeros and the the vector one which contents all ones. Sometimes we will use the letter e to denote this one vector, now the number of elements here depends on the the situation, so have not specified exactly the number of elements here. So, based on the context one can decide what is the number of element in the 0 and 1 vector.

(Refer Slide Time: 15:25)

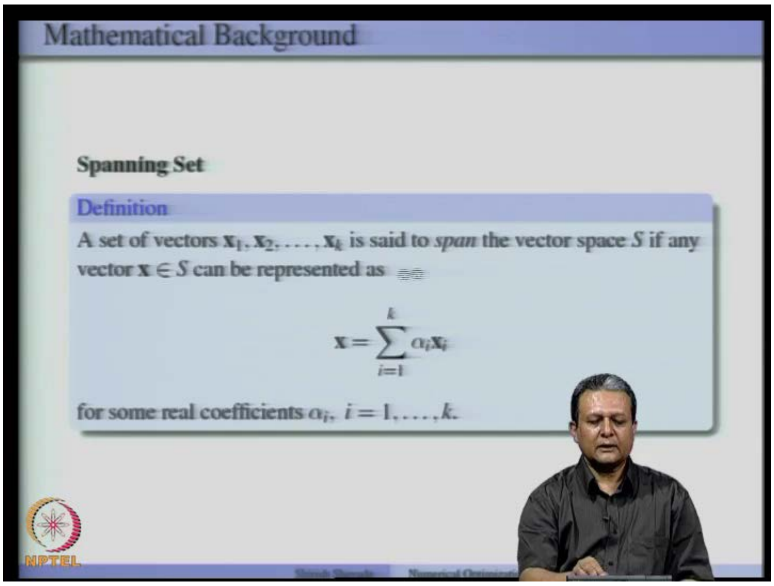


Now, if S and T are vector space S such that S is the subset of T, then S is called a subspace of T. So, here is one question that what are the possible subspaces of R 2, now if we look at the possible subspace of R 2 then the origin, a subspace containing only the 0 vector is a subspace of R 2, then R 2 itself is a subspace of R 2. And then what we have is the set of lines, which pass through the origin, so this is one line which passes through the origin this is a vector space which is a subspace of R 2.

Now, it is easy to characterize our subspace, you can think of it as space where if suppose x y are any two vectors in the vector space and alpha and beta or any two real numbers coming from the field of real numbers, then alpha x plus beta y should always belong to the subspace. Now, going by this definition, you will realize that if you put y equal to minus x and alpha and beta to be 1, then x minus x that is 0 should also belong to the subspace. So, a 0 vector should always belong to the subspace. So, sometimes note this also called linear subspace.

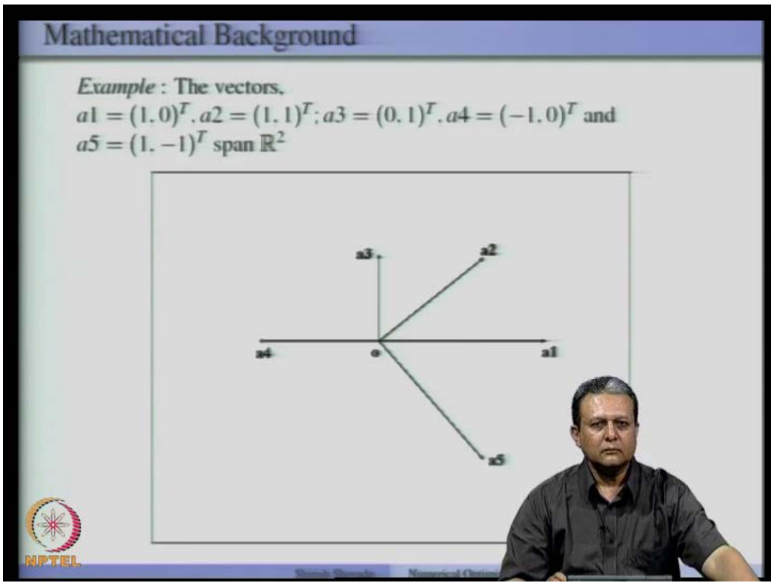
Now, if you translate this subspace that is then it no longer remains a subspace because the origin does not line that such spaces are called affine spaces. So, affine space is just a translation of a linear subspace. Now, in R 3 if you want to write out write down the possible subspaces, so R 3 itself is a subspace of R 3, then you have a 0 vector, which is a trivial subspace of R 3. And then the set of lines which pass through origin form of subspace of R 3, and set of planes passing through the origin also form of a subspace of R 3.

(Refer Slide Time: 18:57)



Now, each sub vector space is span by is the set of vectors. So, let us look at now what is called a spanning set. So, a set of vectors x 1 to x k is said to span the vector space S, if any vector x belong to S can be represented as a linear combination of those vectors. So, so, if you are given this set x 1 to x k, then any vector in the vector space can be represented as a linear combination of those vectors, where these alphas are some real coefficients. So, such a set is said to be the spanning set of the vector set.

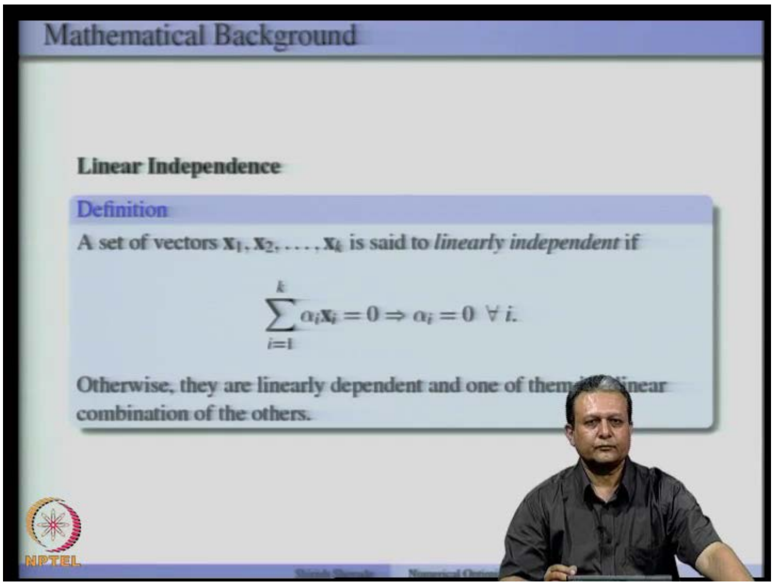
(Refer Slide Time: 19:47)



Now, here is one example, so we have 5 vectors given here, so a 1 which is 1 0, a 2 which is 1 1, then a 3 which is 0 1, a 4 minus 1 0 and a 5 1 minus 1. So, these are the 5 vectors in two dimensional space and they span the 2 the two dimensional space of real numbers. So, what it means is that you take any vector in the space of in the two dimensional space of real numbers now that vector can always be represented as a represented as a linear combination of each of these vectors. Now, that linear combination what we saw it may so happen that, we will not require all the vectors. Some of the alphas could be 0, when we specify a linear combination. So, for example, if we take a 2 a 2 itself represented as a 1 plus a 3, and we do not require a 4 and a 5 if we use a 1 and a 3 to represent a 2. So, similarly one can work out other of examples of vectors, which can be represented using any of this these vectors. Now obviously, the next question would be that, what is the minimum size of the spanning side that is needed to span a given vector space.

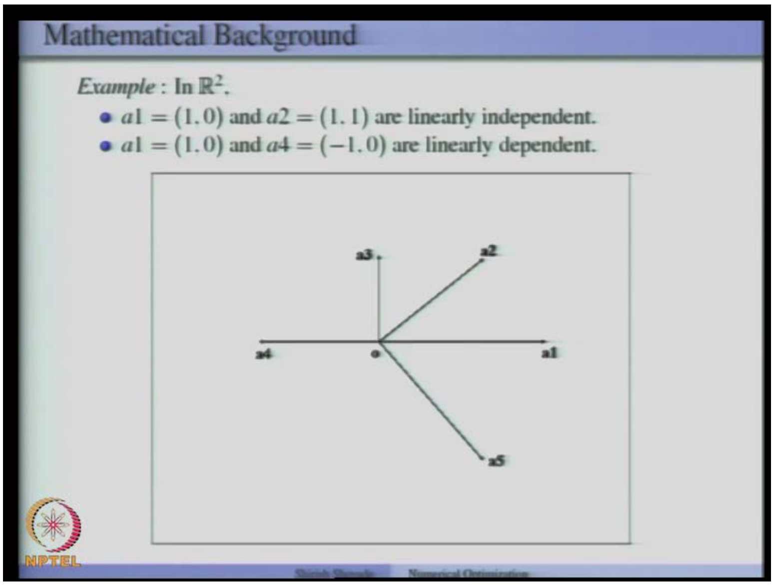
So, it is clear that if your are just given one vector say a 1, a 1 you cannot be used to represent any vector in the two dimensional space. On the other hand if your given say a one and a three, we can represent any vector in the two dimensional space. So, then then what is the role of a 2 a 4 a 5. So, so, is there any redundancy in this set of vectors which span the space R o.

(Refer Slide Time: 21:54)



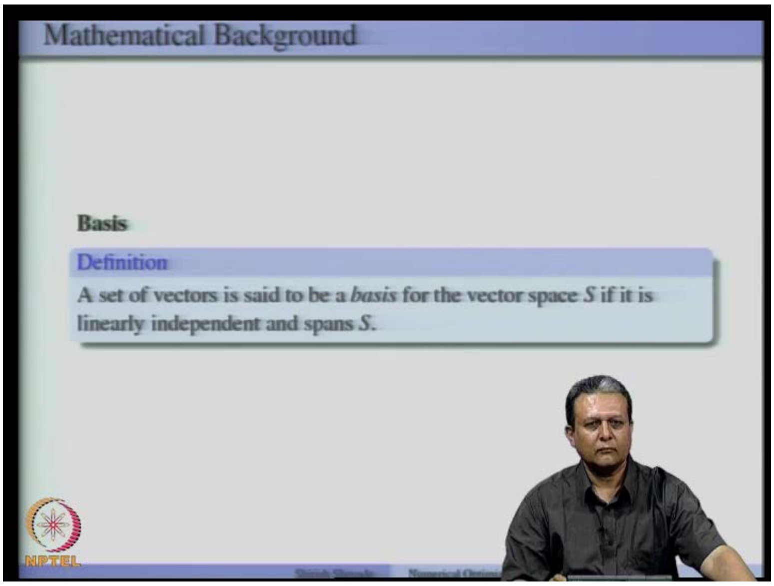
So, we have to study the notion of linear independence for this purpose. Now, a set of vectors x 1 to x k is said to be linearly dependent, if the linear combination of those x i equal to 0 means that each of the real coefficient each of the coefficients alphas are 0 for all I going from 1 to k. Now, if they are not linearly independent then they said to be linearly dependent and one of them can be written as a linear combination of the others. So, this is the important definition for a linear independent of vectors. So, the linear combination of those vectors is 0 implies that each of the real coefficients alphas has to be 0, so such a set of vectors is said to be linearly independent.

(Refer Slide Time: 22:48)



Now, if we look at the vectors a 1 and a 2, so this is 1 0 and this is 1 1 they are linearly independent. So, one can show that, if we take alpha 1 a 1 plus alpha 2 a 2 equal to 0 that will result in alpha 1 and alpha 2 to be 0. On the other hand, if you take the vectors a 1 and a 4. So, if a 1 is 1 0 and a 4 is minus 1 0 then they are linearly dependent. So, one can easily see that a 4 can be written as minus of a 1. So, so, a 1 minus a 4 is equal to a 1 plus a 4 is equal to 0, so and this will happen when alpha 1 equal to 1 and alpha 2 is equal to 1. So, sigma, so a 1 plus a 4 is equal to 0 essentially means that the that vectors are linearly dependent.

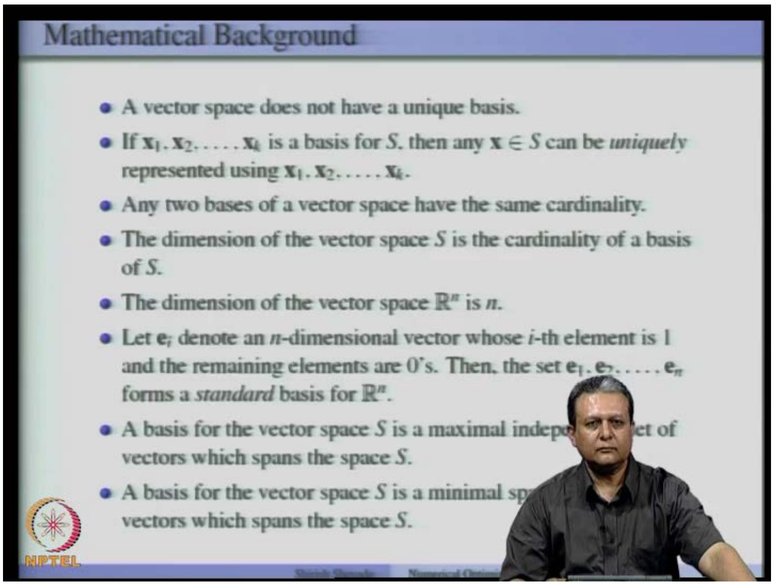
(Refer Slide Time: 23:53)



Now, a set of vectors is said to be a basis for the vector space, if they are linearly independent as span that sets spans S. So, if you look at the previous example a 1 and a 2 are linearly independent. And if a 1 2 also span the space S that means, that if suppose any vector x in two-dimensional space can be written as a linear combination of this a 1 and a 2, then a 1 and a 2 turns out to be a basis for the space.

So, you will see that a 1 and a 2 form a basis, similarly a 1 and a 3 also form a basis. So, this shows that the basis are not unique. So, I can have a 1 a 2 as a basis or a 1 a 3 as a basis or a 4 a 3 also as a basis or a 2 a 5 as a basis and so on or a 2 a 3 as a basis. So, the only thing that I have to ensure is that, they are linearly independent and those vectors should span the space S.

(Refer Slide Time: 25:16)



So, as we saw that a vector space need not have a unique basis one can have multiple basis, but suppose if you fix a basis say x 1 to x k or a vector space S, then any x in that space S is uniquely represented using x 1 to x k. So, once if you fix a basis a vector is always uniquely represented. So, the alpha 1 alpha 2 alpha k will be unique representation of x with respect to the basis x 1 to x k.

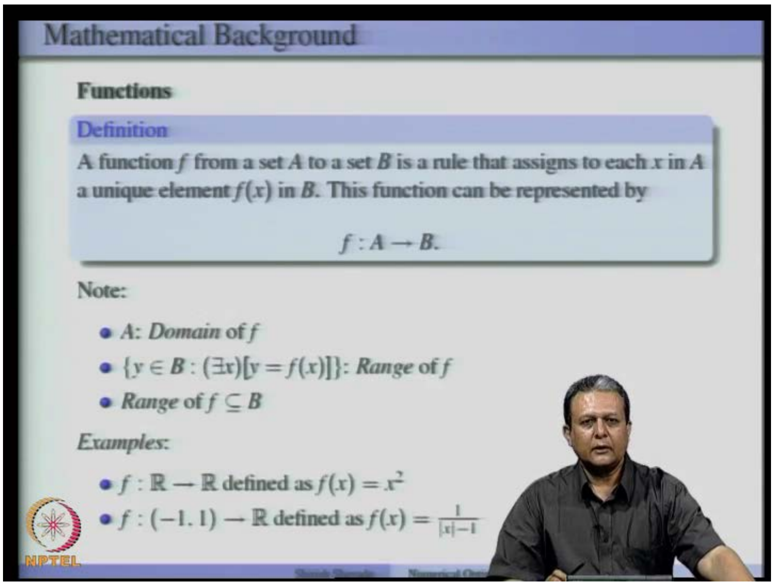
Now, if you take any two basis of a vector space they have the same cardinality. So, the number of basis could be same they should be same for a given vector space and dimensional dimension of a vector space is the cardinality of a basis. So, the set of linear independent vectors maximum linearly independent vectors we span thus set S that is the basis and then the cardinality of that set is called a dimension. Now, when we looked at the example in the two dimensional space, we saw that a 1 and a 2 or a 1 and a 3 are enough to represent any vector in the two dimensional space. And further a 1 and a 2 or a 1 and a 3 are linearly independent, so they form a basis. So, the dimensionality of this space is two because they can form a basis of size two.

Similarly, extending those ideas to high dimensional spaces we can say that the dimension of the vector space R n is n. Now, if e i denotes a n dimensional vector whose I th element is one and the remaining elements are zeros; then the set e 1 e 1 to e n forms a standard basis of R n. So, in each of this e S the i th element is one and the rest of the elements are 0, they form a standard basis. So, if you look at the two dimensional example this a 1 and a 3 they form a standard basis of R.

Now, a basis or a vector set S is the maximal independent set of vectors which spans the space S. So, it is a maximal independent set which spans the set, so if you add anything any extra vector to this basis; then it becomes a a set of linearly dependent vectors. So, adding any extra thing to the set S will make it linearly independent; and one more remark about this basis is that a basis for the vector space S is a minimal spanning set of vectors, which spans the space S.

So, this is the minimal spanning set S, so if you take out any element from the set S, those vector space cannot be any element of the vector space S cannot be constructed using the remaining element. So, these are the necessary elements to span the space S. So, these are two important concepts that the some maximal independent set of vectors which spans S and a minimal spanning set. So, you cannot add anything because then it will become dependent and you cannot remove anything from that set of basis because then it will not span the space S.

(Refer Slide Time: 29:14)

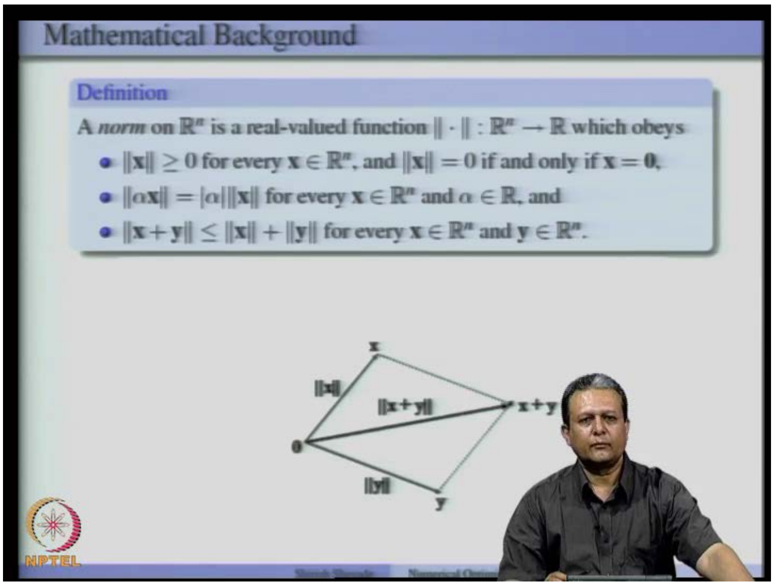


Now, after having studied this vectors and vector spaces we look at some other preliminaries. So, let us look at the functions, a function f is always defined from a set A to B and the function assigns each element in x in A a unique unique element f x in B. So, we will denote this as f is a function, where it takes a element from the set A and assigns it an element f of x in B and f from A to B is denoted like this. Now, we call this a as the domain of f.

Now so, the function takes an element of a and assigns it some value, so the values taken by the function in the set B that is called the range of f. So, remember that the range of f is always A subset of the set B. So, here are some examples of the functions. So, suppose f is defined from R to R. So, the domains as well as the range are the set of real numbers and the function is defined as f of x equal to x square.

So, this function will take every, every element from the domain and assign a value x square to it and that will be the function value. One can also define a function say from the open interval minus 1 to 1 R, where the function is defined as 1 over mod x minus 1. So, this you can see that this function is not defined at the end points of this interval, while it is defined at any intermediate points, so this the definition of a function.

(Refer Slide Time: 31:22)

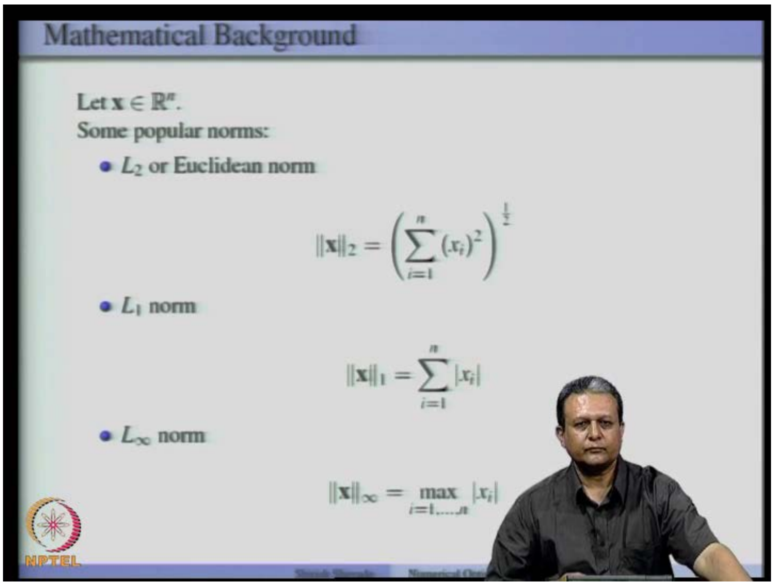


So, we now define what is called a norm on R n. So, a norm on R n is a real-valued function. So, it takes a n dimensional vector and assigns a real number to it and then it satisfies certain properties. So, the first property that the norm should satisfy is that a norm is always non-negative. So, you take any vector x in R n the norm has always to be non-negative and norm is 0, if and only if x is a 0 vector. So, only for 0 vectors, so norm is 0 otherwise it is a positive quantity. Now, if you take a vector x belong to R n and alpha a real number then the norm of alpha x can be written as mod alpha into norm x.

Now, we have seen that norm x is always a non-negative quantity and norm also has to be a negative quantity. So, this we have to take mod alpha here, so mod alpha is always a non-negative quantity. So, this holds for every x in R n and alpha belongs to R. The third important property that the norm should satisfy is that norm of x plus y should be less than or equal to norm x plus norm y for every x and y in R n. Now, this property is called triangle inequality and it is clear from this figure, so we have a vector x. So, norm x is the length of this vector we have vector y norm y the length of this vector.

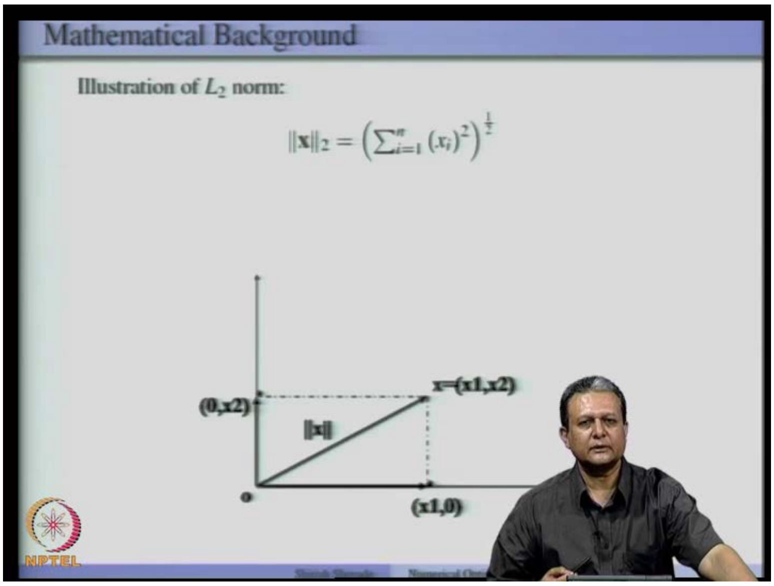
So, suppose if you add x plus y, so the resultant vector is x plus y and its length is norm of x plus y. Now, this vector is y, so using triangle inequality that we have studied in earlier classes, we note that the the side of a triangle is always less than or equal to the sum of the sides of sum of the other two sides of the triangle. So, norm of x plus y is always less than are equal to norm x plus norm y. So, a norm should always satisfy these properties will will see more details about these norms now.

(Refer Slide Time: 33:55)



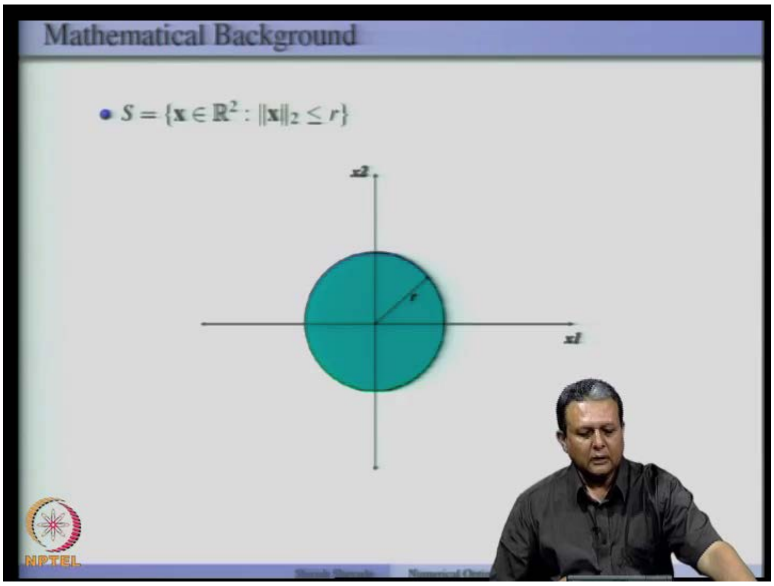
So, let x be a vector in n dimensional space, now here are some definitions of some popular norms. So, one is called L 2 or Euclidean norm we know about this norm. So, norm x, so L 2 norm of x, so you take square of each element add them and then take a square root, so this is called the L 2 norm of the vector x. The L 1 norm you take an absolute value of each of the elements and in sum them up that will give us the L 1 norm of x. So, that is a another norm which is sometimes is that is called L infinity norm, where we among all possible values we take maximum of mod x i, and that will be the infinity norm of the vector x. So, I will give some illustration about some these norms.

(Refer Slide Time: 34:54)



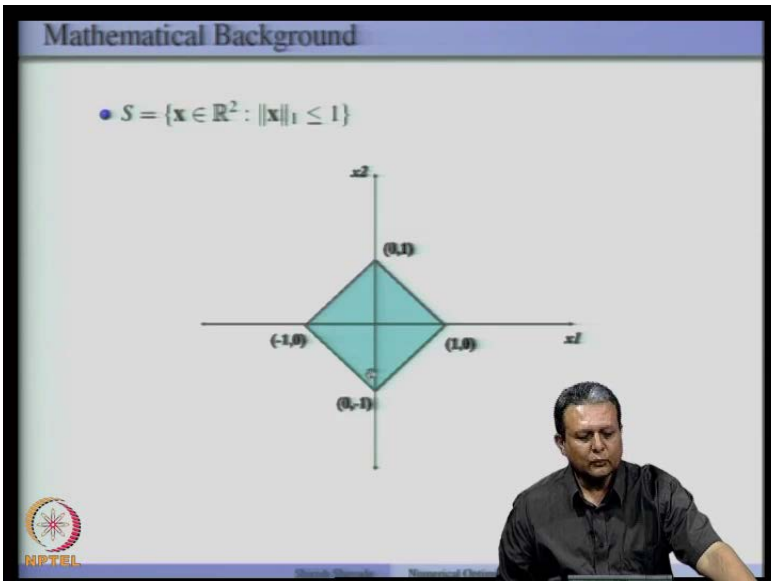
So, let us take a two dimensional space, so you have two axis here and there is a origin here and vector x, which has two components x 1 and x 2. So, so, the norm of x is easy to find, so this distance, so if we you drop a perpendicular from x to this horizontal axis. So, this distance is x 1, so from 0 to this point the distance is x 1 and from this point to this the distance is x 2 which is as good as saying that we drop a perpendicular from x the vertical axis, so this distance is x 2. So, the horizontal distance is x 1 the vertical distance is x 2, so norm x is square root of x 1 square plus x 2 square using pythagoras theorem, so that we have showed here. So, this is a illustration of a L 2 norm.

(Refer Slide Time: 36:03)



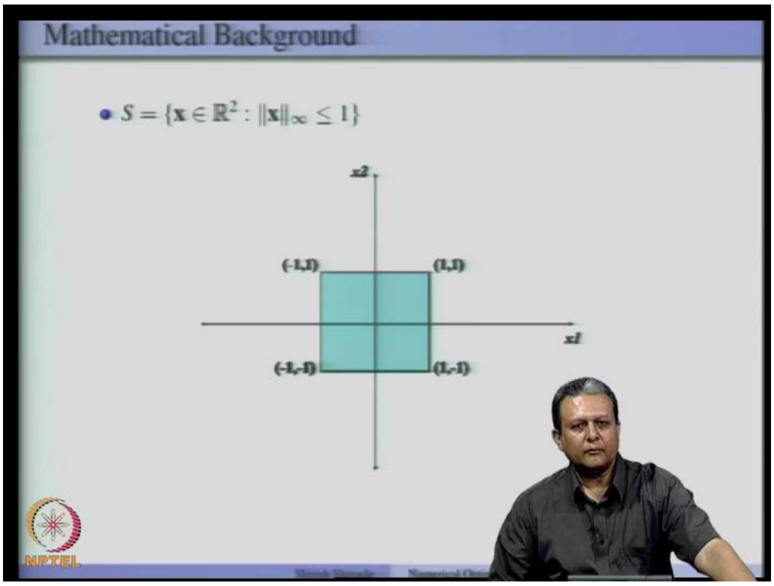
Now, we look some sets, so suppose we again consider a two dimensional space of real numbers, now the set is shown here is the set of points whose two norm is less than or equal to R. So, the points on this circle as well as the points in the interior of the circle, they constitute the set S where the two norm of x is less than or equal to R.

(Refer Slide Time: 36:47)



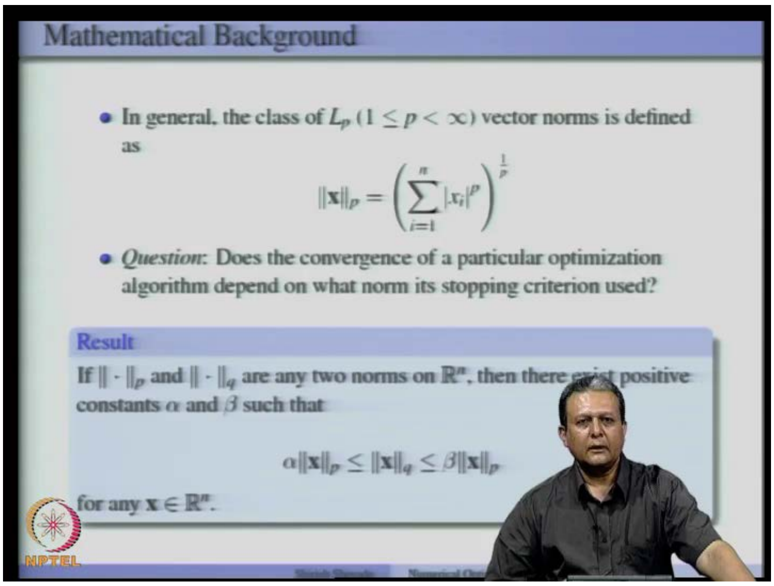
Now, the two norm x is less than or equal to R means that we are talking about the circle of radius R centered around origin. Now, when it comes to one norm the things are different, so you will see that again we consider a two-dimensional space of real numbers and the region shown here is the set of points whose one norm is less than or equal to 1. So, you will see that unlike the previous case this is a diamond shaped object centered around origin.

(Refer Slide Time: 37:20)



Now, if we look at the L infinity norm, so this is a square, whose end points are 1 1 and minus 1 minus 1. So, the shaded region or the colored region here shows that the set of all points whose L infinity norm is less than or equal to 1. So, you will see that depending upon the definition of the norm the distance major has a different notion.

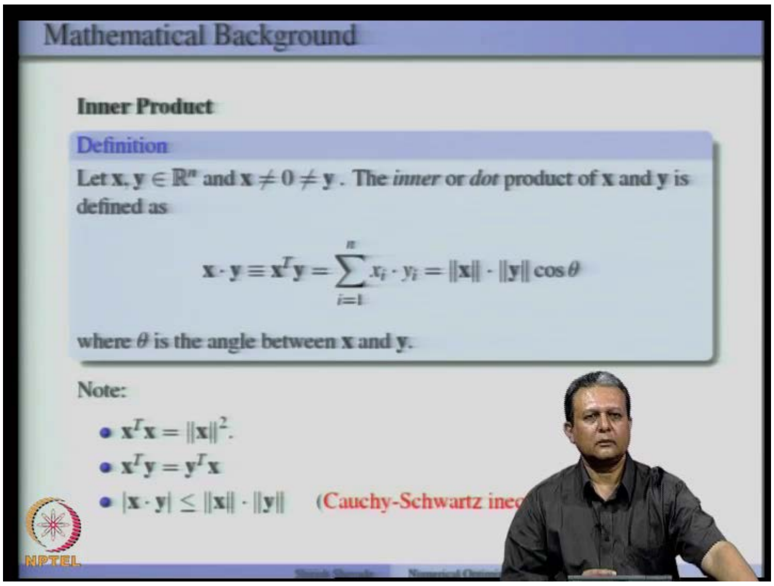
(Refer Slide Time: 37:55)



So, in general you can define a L p norm where p is a finite quantity greater than or equal to 1. So, you can define l p norm as you take the mod of the individual component and take a pth power of that sum them up over all the n components of the vector and then take a pth root of that quantity. So, we have studied, so many different types of norms and different norms exist as we see here, now the question is that typically in optimization algorithms, we use some norm to find out the distance of a current point from the solution. So, the obvious question that one would ask is that there is a convergence of a particular optimization algorithm depend on which norm is used for what as the stopping criteria.

So, there is a important result which holds in R n not in infinite dimensional space. So, we say that if suppose we have two norms norm p and norm q, which are defined on R n then there exists a positive constants alpha and beta, such that alpha pth norm of x is less than or equal to qth norm of x less than or equal to beta pth norm of x for any x in R n. So, which means that that if we consider qth norm that is bounded below and above by the p th norm of the same vector with appropriate constants and find beta. So, the result of optimization algorithm does not depend on the norm, that we use in your stopping criteria because of this important result.

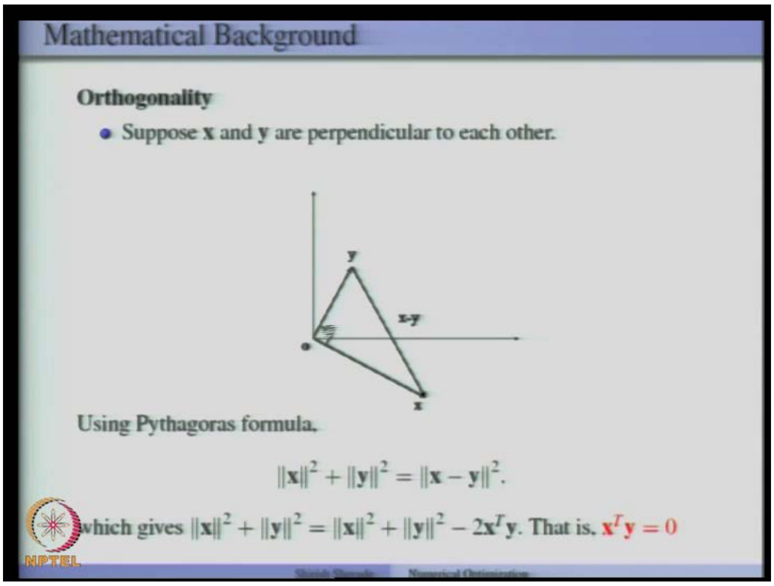
(Refer Slide Time: 39:58)



So, we now look at some other definitions, so the first one is the inner product of two vectors. So, let us consider two non zero vectors in n dimensional space then the inner product or dot product of these two vectors. So, we will denote either by x dot y or x transpose y. So, we can will use either of these notations when we refer to the inner product of x and y and that is defined as take the product of the individual components of the two vectors and then sum them up. So, component-wise product of the two vectors and then summing of up those products will give us the inner product of two vectors x and y and that can also be written as 2 norm of x into 2 norm of y into cos theta; where theta is the angle between the two vectors x and y. So, this is how the inner product or dot product of 2 vectors is defined.

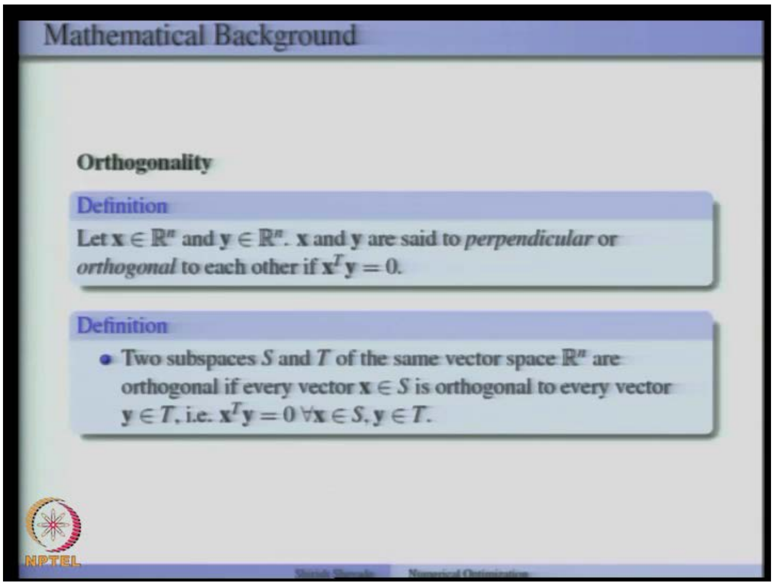
Now, there are some remarks that I would like to make here, first one is that x transpose x x is nothing but norm x square. So, that is obvious from this definition that, when you take a dot product of the vector with with respect to itself. Then, the angle between the two vectors is 0 cos 0 is 1, so x transpose x is nothing but norm x square. Similarly, the angle between x and y is same as the angle between y and x. So, x transpose y is nothing but y transpose x now there is another important property that one should keep in mind is the mod of x dot y is less than or equal to norm x and into norm y, this is called Cauchy-Schwartz inequality. So, this property is obvious from the definition of the inner product. So, we know that cos theta is always in the range minus 1 to 1. So, mod of x dot y has to be less than or equal to norm x into norm y. So, this is called Cauchy-Schwartz inequality we will use it sometime during this course.

(Refer Slide Time: 42:30)



The next concept is about the orthogonality of vectors, now suppose x and y are perpendicular to each other. So, here is a vector x and here is a vector y, which are perpendicular to each other. Now, I have taken a vector which is x minus y, so if you take y plus x minus y you get x. Now, since the two vectors are perpendicular we can use Pythagoras formula. So, norm x square plus norm y square is equal to norm of x minus y square. Now, if you expand the right side what we get is norm x square plus norm y square minus 2 into x transpose y. So, equating this 2 you will see that x transpose y has to be 0, so which means that if x and y are perpendicular to each other then x transpose y has to be 0.

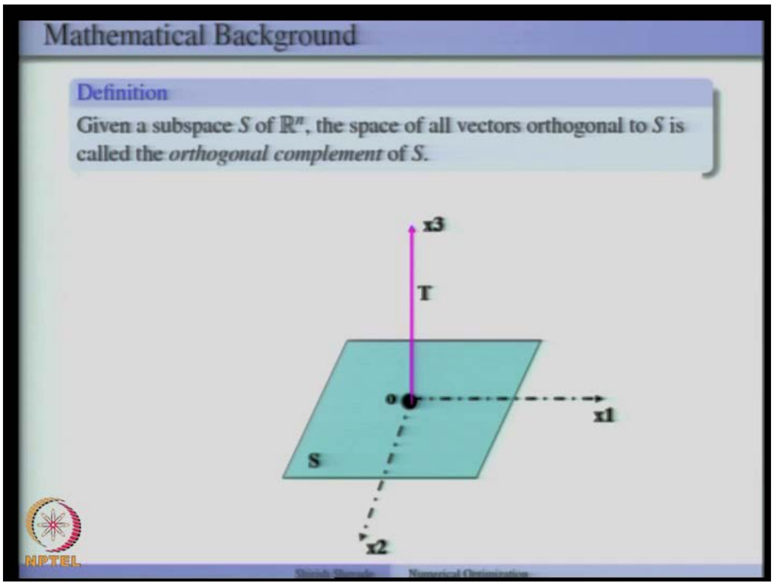
(Refer Slide Time: 43:31)



So, here is the definition of orthogonality, so let x be a n dimensional vector y also be a n dimensional vector x and y are set to be orthogonal or perpendicular to each other if x transpose y equal to 0. So, we will use the term perpendicular or orthogonal interchangeably. Now, suppose if we have two subspaces S and T, which of the same vector space R n the 2 subspaces are said to be orthogonal if every vector x belong to S is orthogonal to every vector y belong to T. So, that means, that if you take any x in S and any y in T then x transpose y equal to 0.

So, suppose if you take a 2 dimensional space and we take a horizontal axis as the one space one subspace and vertical axis as another subspace. Now, clearly their subspaces because they pass through the origin now you will see that any vector x on the horizontal axis is perpendicular to any vector y on the vertical axis. So, such subspaces are called orthogonal subspaces are said to be orthogonal to each other.

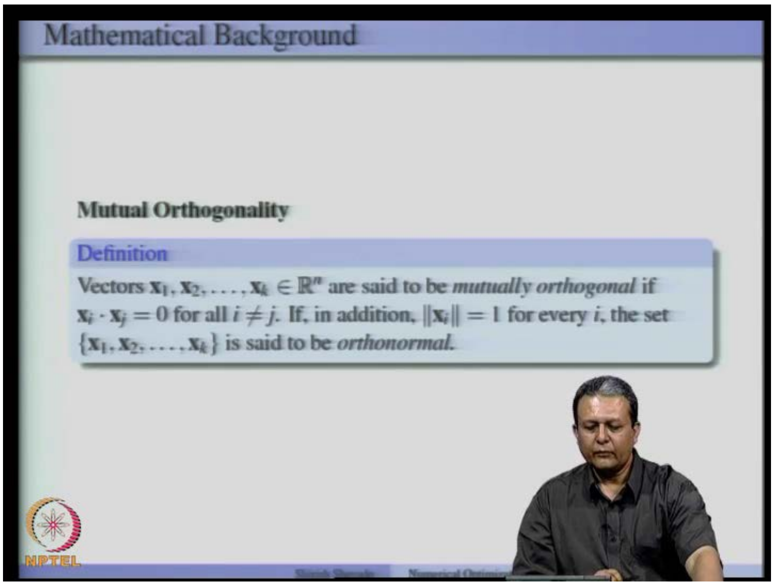
(Refer Slide Time: 45:10)



Now, let us look at the definition of orthogonal complement now suppose given a space S, the space of all vectors orthogonal to S is called the orthogonal complement of S. So, suppose if I take a three dimensional space now S is suppose a space spanned by the vectors x 1 and x 2. So, you can think of it as a horizontal plane and T is the subspace spanned by the third axis which is x 3. Now, these two, so any vector in the space S is orthogonal to any vector in the space T for further S and T together span the three dimensional space. So, there is a difference between the orthogonal subspaces and orthogonal complements.

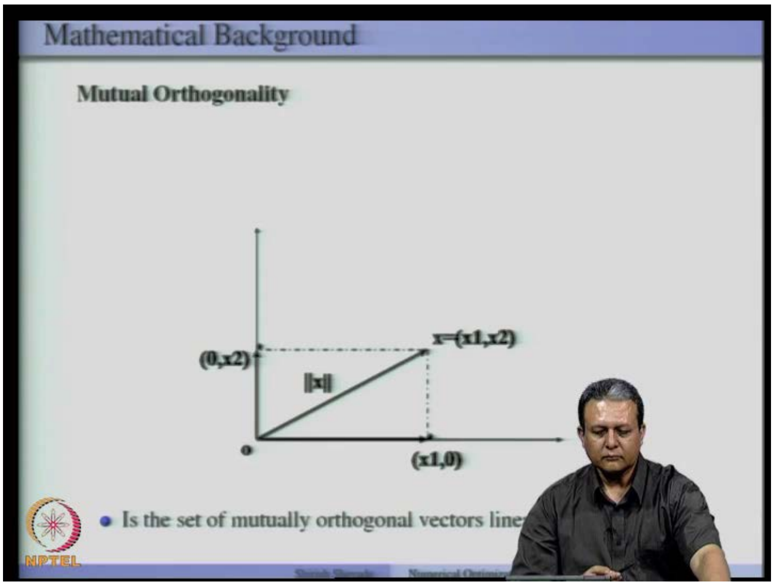
So, in orthogonal complement the two subspaces spanned by entire that in the in this case the two subspaces S and T span the three dimensional space. On the other hand, if you just take x 1 as 1 subspace and x 3 as another subspace, then they do not although the each vector on x 1 is orthogonal to each vector on x 3 they together do not span R 3. So, they they are just orthogonal to each other, but in this case since S and T together span a three dimensional space, we say that S and T are orthogonal complements of each other.

(Refer Slide Time: 47:04)



Now, let us look at mutual orthogonality, now vectors x 1 to x k are said to be mutually orthogonal if they are pairwise orthogonal. So, you take any two vectors two different vectors the dot product is 0. Now, if further if the norm of each vector is 1 for every i, then the set x 1 to x k is said to be orthonormal. So, for the orthonormal set x i transpose x j i not equal to j is equal to 0 and x i transpose x i equal to 1.

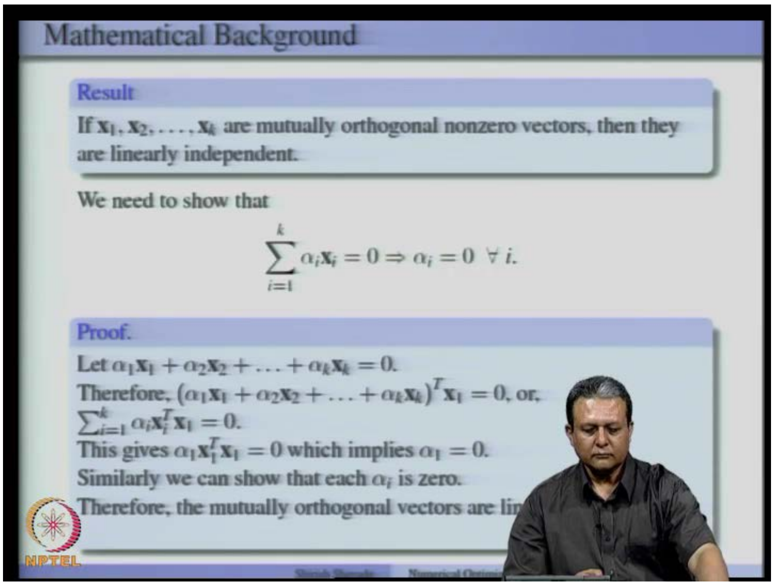
(Refer Slide Time: 47:55)



Now, the next question is that is the set of mutually orthogonal vectors linearly independent. So, you have a vector space whose basis is horizontal axis and vertical axis.

So, we are talking about two-dimensional vector space. Now, these two basis vectors are mutually orthogonal, assume that the norm of each of the vectors is 1, now is these set of mutually orthogonal vectors linearly independent the answer is yes.

(Refer Slide Time: 48:42)

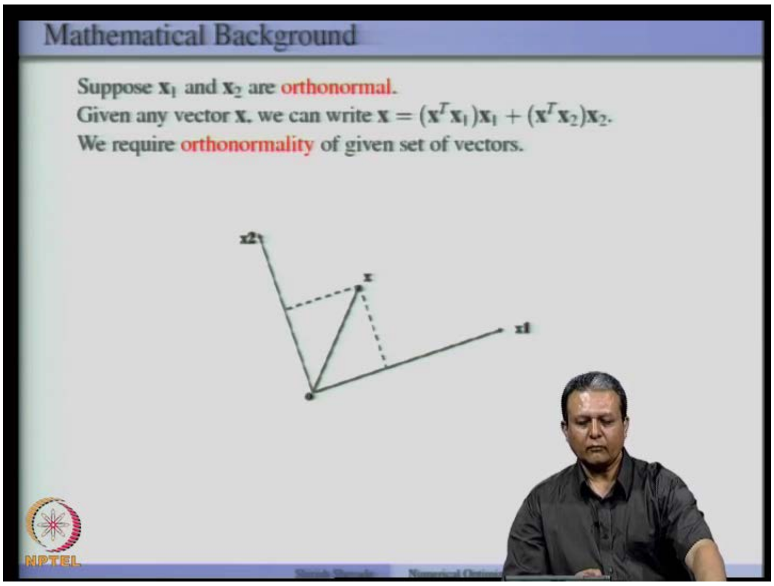


If you have a set of mutually orthogonal non zero vectors then they are indeed linearly independent. So, we will show this result now to show this result what we need to show is that the set of vectors are linearly independent. So, which means that sigma alpha x I equal to 0 implies alpha equal to 0 for all i. So, here is a small proof of this result. So, let us start with the left hand side, so we say that alpha 1 x 1 plus alpha 2 x 2 plus alpha k x k is equal to 0 and finally, we have to show that the right side (( )).

Now, what we do is that, we take a dot product of each of the vectors with respect to x 1. Now, remember that what is given is that x 1 to x k are mutually orthogonal non 0 vectors. So, we can expand this to write it as alpha 1 x 1 transpose x 1 plus alpha 2 x 2 transpose x 1 and so on plus alpha k x k transpose x 1 equal to 0. So, which can be shortly using the short form can be written as alpha sigma alpha i x i transpose x 1 equal to 0. Now, we are given that the set of vectors are mutually orthogonal. So, x 1 transpose x 2 equal to 0 x 1 transpose x k equal to 0 except x 1 transpose x 1 and x 1 transpose x 1 is 1, because they are orthogonal.

So, which, so each of this k minus one comes vanishes and what we are left with this only alpha 1 x 1 transpose x 1 equal to 0, and since x 1 has a norm one the only way this is possible is when alpha alpha one equal to 0. Now, similarly we can show that each of the alpha case is 0 by multiplying by x 2 to x 2. So, by doing that we will show that all all alphas are 0, and since all alphas are 0 this condition is satisfied and then we can said that the mutually orthogonal vectors are linearly independent.

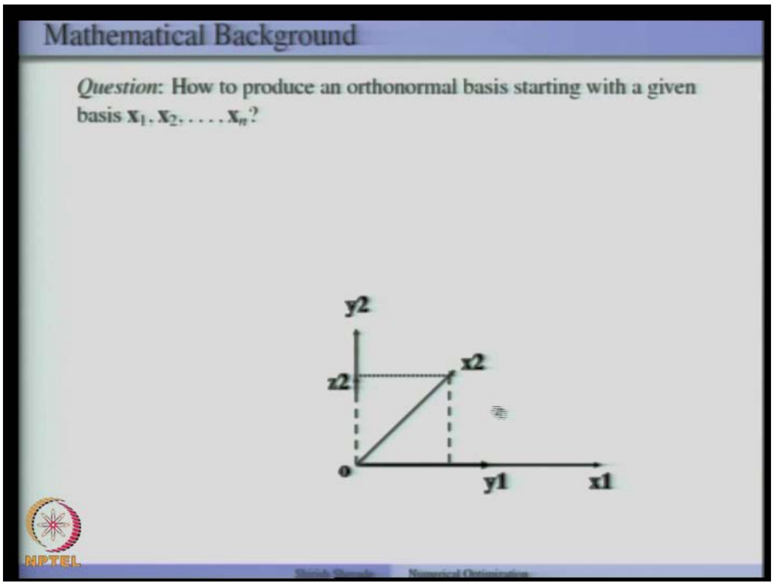
(Refer Slide Time: 51:16)



Now, suppose x 1 and x 2 are orthonormal, so which means that they are perpendicular to each other and each of them has a unique norm. Now, if we take any vector x we can write it as the component of x along x 1 into the vector x 1 plus the component of x along x 2 into the vector x 2. So, the component of x along x 1 is x transpose x 1 and this direction is x 1.

Similarly the component of x along x 2 is x transpose x 2 and this direction is x 2. So, x can be written as x transpose x 1 into x 1 plus x transpose x 2 into x 2. Now, this is a simple form in which x can be written, so that it is it can be written as a linear combination of x 1 and x 2 and the individual components also can be found out easily. Now, all this requires that x 1 and x 2 need to be orthonormal only then in we can write this in a easy way.

(Refer Slide Time: 52:45)



Now, what happens when the set of vectors the set of basis, which is given to us is not orthogonal, so how to generate and orthogonal basis orthonormal basis from the set x 1 to x 2. So, here you will see that unlike the previous case, where x 1 and x 2 are orthonormal, here they are not orthonormal. So, they are not perpendicular to each other and neither they are perpendicular to each other nor they have unique norm.

Now, from this we are interested in generating a basis, which is orthonormal and that basis I have shown here, so which is y 1 and y 2. So, there are two important things that you have to note here. So, 1 is that y 1 and y 2, which are generated using x 1 x 2 they are orthonormal to each other, they are orthogonal to each other and not only that y 1 and y 2.

They have unique norm, so which means that they form an orthonormal basis for a space, which is spanned by x 1 to x 2. Now, in general suppose you are given n basis x 1 to x n, how do we form a orthonormal basis y 1 to y n such that y 1 to y n are mutually orthogonal and each of the y's have unique norm. And that procedure is commonly known as a Gram-Schmidt procedure, so will study that procedure next time.

But, before I sign off I just would like to say some important fact here is that this finding this orthonormal basis is very important in the sum of the optimization algorithms; especially when we try optimize along each dimension and we when we optimize along one dimension optimize an objective function along one dimension, we want to make sure that next dimension that we add is orthonormal to the first one. So, finding the orthonormal basis is a very important concept from the optimization view point and we will study that in the next class.

Thank you.