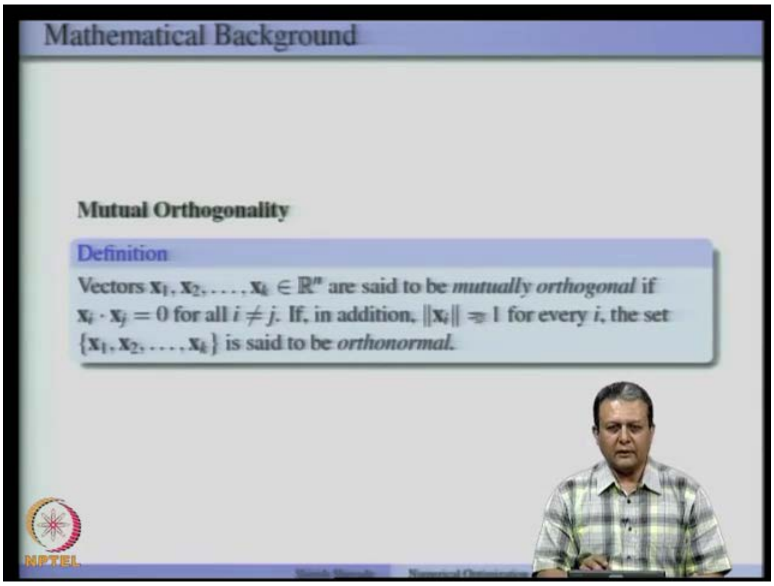
Numerical Optimization
Prof. Shirish K. Shevade
Department of Computer Science and Automation
Indian Institute of Science, Bangalore

Lecture - 3
Mathematical Background (contd.)

Welcome to this third lecture. So, this is the continuation of our previous lecture, where we were talking about vector spaces. Then, we talked about various properties of vector space, sub space, basis dimension of a vector space. Then, we moved on to mutual orthogonality of vectors.

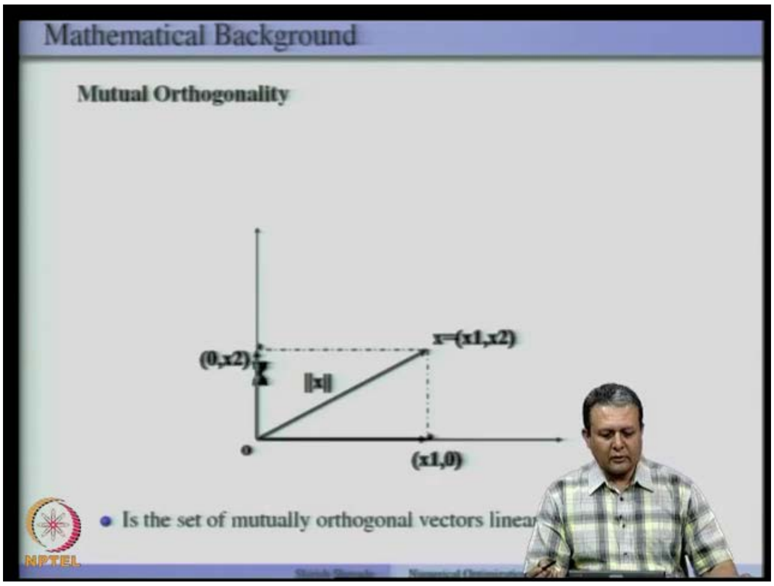
(Refer Slide Time: 00:43)



So, vectors x 1, x 2 up to x k are said to be mutually orthogonal, if you take any distinct vectors in that set, there dot product is 0. Now, if in addition, the norm of each vector is 1, then we say that the set of vectors is ortho normal. So, for ortho normal vectors, if you take a, if you take any distinct vectors in the set dot product is 0, the norm of individual each of the vectors the set is 1.

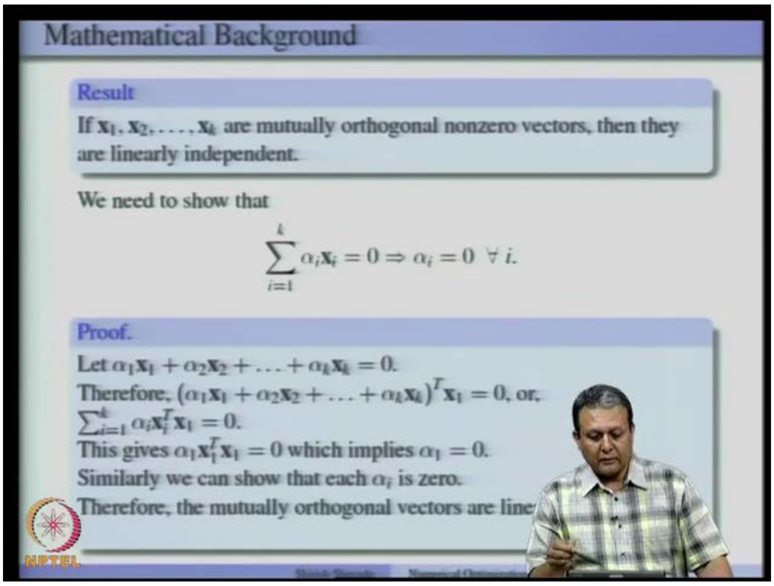
.

(Refer Slide Time: 01:20)



Now, what is the advantage of this ortho-normality? So, suppose I have this vector, which is ortho normal to this vector. Then, what I can do is that I can project given vector x along each of those directions and those components. So, the vector x can be represented as the component x 1 into the ortho normal vector in this direction plus the component x 2 into the ortho normal vector in this direction.

(Refer Slide Time: 01:55)

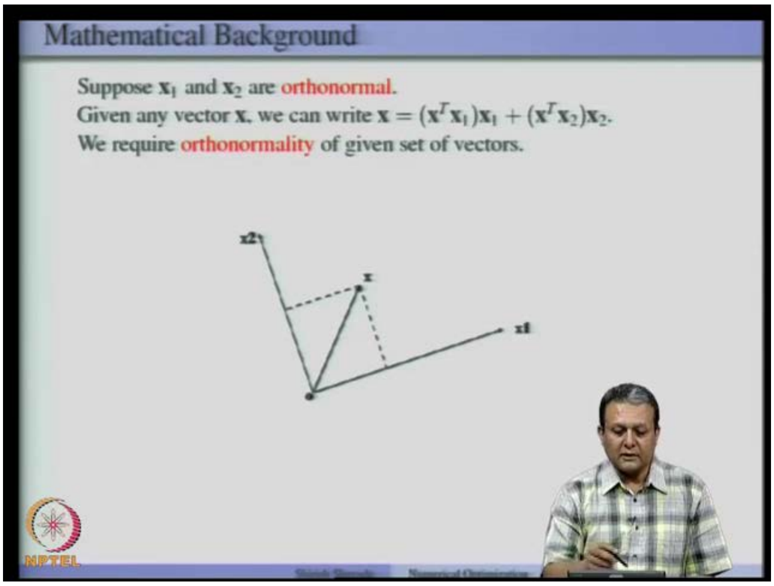


Now, we also saw last time that this set of ortho normal vectors are linearly independent, and we showed that for linear independents. We need to show that if the left hand side

.

holds, then of the left hand side of the implication holds. Then, the right hand side should hold. So, we started with this left hand side. Then, we took one vector x 1 and took a dot product of this with x 1. Then, we show that alpha 1 has to be 0 as x 1 has the unit norm or x 1 has a non-zero norm because we are talking about non zero vectors. So, alpha 1 is 0, which essentially implies that all alpha i are 0 by taking the dot products with the respect to vectors. Then, we show that these orthogonal non zero vectors are indeed linearly independent.

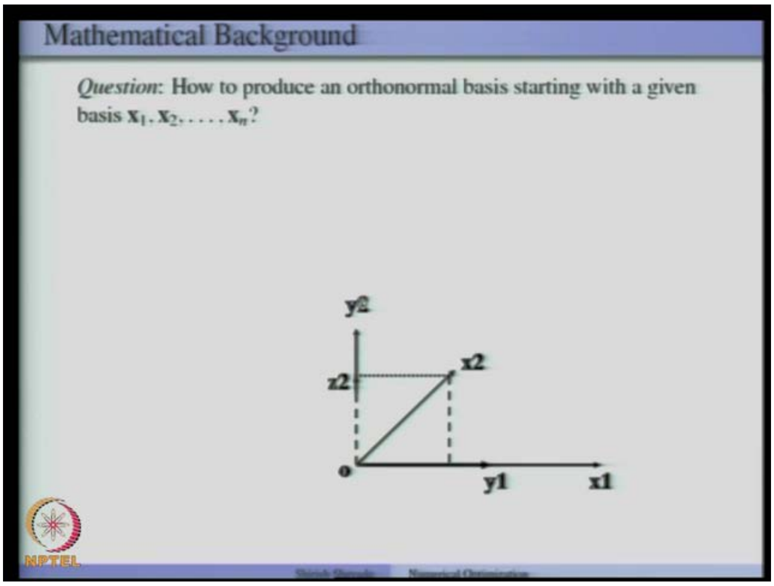
(Refer Slide Time: 02:47)



Now, given a set of ortho normal vectors as we saw that if we have vector x, then the x can be easily represented using x 1 and x 2. So, what we need to do is that we need to take a dot product of x with respect to x 1 and dot product of x with respect to x 2 and then this component into x 1 plus this component x 2 will give us the vector x. But then, we started with the assumption that x 1 and x 2 are ortho normal. That may not always be the case in practice.

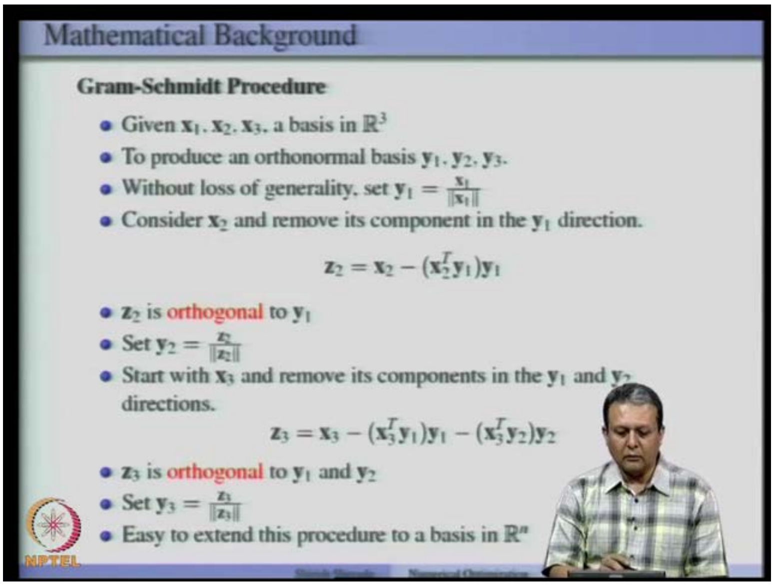
.

(Refer Slide Time: 03:23)



So, how to produce an ortho normal basis starting with any arbitrary basis x 1 to x n for a vector space. So, here is the basis x 1 and x 2, you can see that they are not orthogonal and also they are not ortho normal. Now, from this, we need to generate the basis y 1 and y 2, which are ortho normal and the norm of the individual vectors is y. So, there is a well know procedure called gram Schmidt procedure to do this orthogonalization. We are going to study that procedure now.

(Refer Slide Time: 04:03)



.

So, let us take a simple example. Let us assume that we are given x 1, x 2, x 3, which is the basis in 3 dimensional space. Our aim is to produce an ortho normal basis y 1, y 2, y 3. So, what it means is that the norm of the individual components y 1, y 2, y 3 should be 1. They are this set of vectors is mutually orthogonal to each other. Now, without loss of generality, what we will do is that we will take the vector y 1 to be the first vector x 1 divided by norm of x 1. So, that means norm of y 1 is 1. Now, let us consider the vector x 2. So, if you look at this figure, so we had x 1. From that, we derived a vector y 1, which is along the same direction, but which are the unit norm.

Now, let us look at vector x 2, which is the second vector in the basis. What we do is that we remove the component of x 2, which in the direction y 1. What will be left with if the component z 2? So, in short the z 2 is nothing but x 2 minus the component of x 2 along the direction y 1 removed. So, this is explained here. Now, we have got z 2. You can see that z 2 is orthogonal to y 1. So, if you take a dot per out of z 2 with y 1, what you get is x 2 transfers x 1, x 2 transfers y 1 minus x 2 transfers y 1 into y 1 transfers y 1.

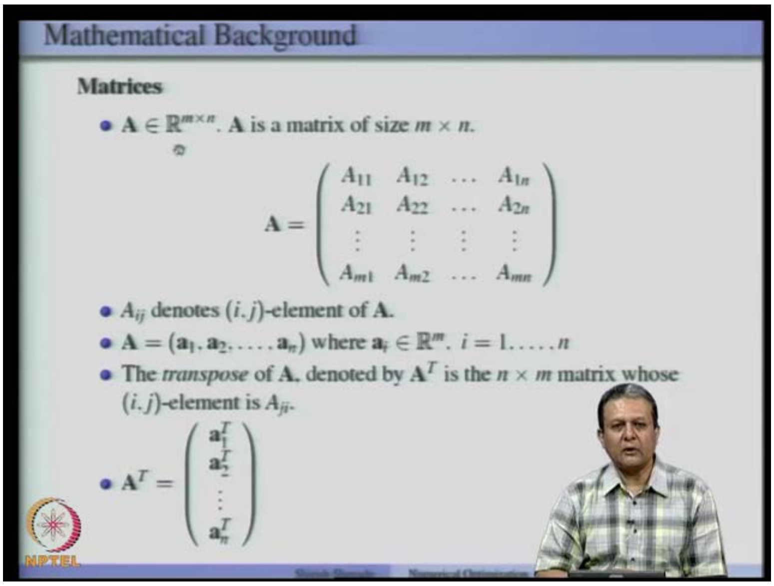
Now, y 1 transfers y 1 is 1. This is by design. So, x 2 transfers x 1 minus x 2 transfers x 1 will make it z 2 transfers y 1 to be 0. So, it clearly shows that z 2 is orthogonal to y 1. So, we have got a vector, which is orthogonal to y 1. Now, our next job is to make z 2 make define a vector y 2, which which is along the same direction at z 2, but has the unit norm. So, that is done using y 2 is equal to z 2 by norm rate 2. So, this will make it make y 2 unit norm vector. Now, we start with x 3. So, we take the third vector in the set x 3 and remove the components along the previous 3 2 difference directions y 1 and y 2.

So, the component of x 3 along the direction y 1 is removed from x 3. Then, the component of x 3 along a direction y 2 is also removed. What we get is z 3. Now, one can easily verify that z 3 is indeed orthogonal to both y 1 and y 2. So, we have got a direction, which is ortho normal to both y 1 and y 2. Now, we have to just make sure that the y 3 vector the unit norm vector.

So, we do do it in a similar way as we did earlier. So, y 3 is a z 3 by norm z 3. The procedure can be extended to a general basis in R n. So, given a general basis in R n, one can use a gram Schmidt to to get a ortho normal basis y 1 y R in R n. So, this procedure is going to be very useful in our optimization algorithms, in some of the optimization algorithms. So, that is why I spent some time discussing about this.

.

(Refer Slide Time: 07:38)

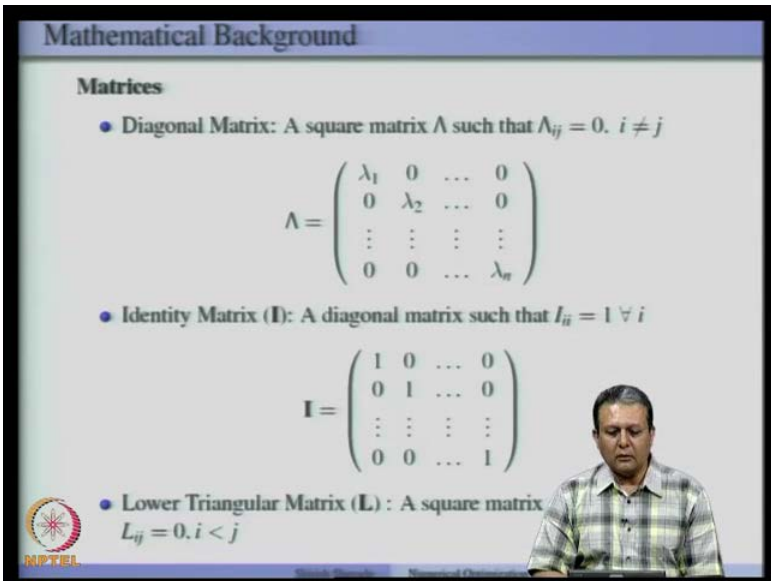


Now, we will go to the topic of matrices. So, here is the matrix, which is shown here. So, a matrix, which of dimension m by n means that it has m n elements. We in this of course, we will always be worried about the matrices with whose entries are real numbers. So, that is why we have a written test R m cross n. So, it is the m by n matrix consisting of real numbers.

Now, each A adds A denotes i j element of this matrix. So, you can think of matrix has a collections of vectors appended to one another. So, if if if A 1 denotes the first column, A 2 denotes the second column and A n denotes the n th column, then we can write the matrix A has A 1 A 2 up to A n where each A i is a vector in m dimensional space in this case. Now, the transpose of A will be denoted by A transfers and it is A n by m matrix whose i j entry is A j i. So, when we transpose it, this row will become, this first column will come the first row, the second column will become the second row and so on. So, a transpose in short can be written as the individual columns are transposed and put there.

.

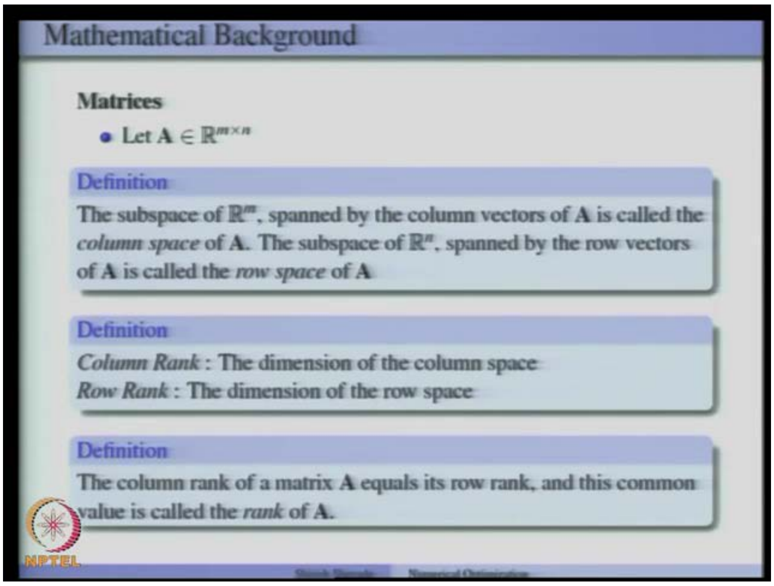
(Refer Slide Time: 09:18)



Now, we can define the square matrix lambda, where you have only non zero entries along the diagonal and 0 entries the half diagonal limits. Some of these lambdas could be 0, but not all will will be 0. So, such a matrix is called diagonal matrix, where half diagonal elements lambda is unequal to 0 for all i are not is equal to j. Now, a special case of this diagonal matrix is the identity matrix, which will be denoted by I, where all the diagonal elements are 1 and half diagonal elements are 0. So, another definition is a lower triangular matrix. So, it is a square matrix, where the entire upper diagonal, all the elements above the diagonal are 0 and above all the elements above the below the main diagonal not all of them are 0. So, such a matrix is called a lower triangular matrix. So, i j elements of that matrix is 0 is i less than j. So, that means that all the elements above the diagonal are 0.

.

(Refer Slide Time: 10:39)

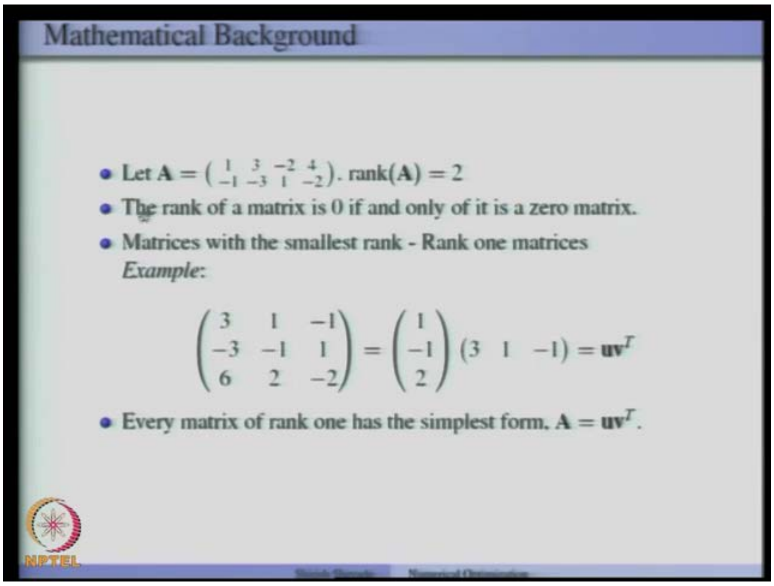


Now, let us look at some definitions. So, the subspace, let us consider matrix A, which is of the size m by n. Now, the subspace is spanned by the column matrix of A is called the column space of A and subspace spanned by the row vectors of A is called a row space of A. so, the column space is the subspace of R m because that is spanned by the vectors in R m and are row space is the subspace of R n.

Now, the dimension of the column space will be called the column rank. The dimension of the row space is called a row rank. So, for a given matrix, what is the rank of a matrix? So, here, we have one result, which says that the column rank of a matrix equals its row rank. This common value is called the rank of the matrix. So, for any embed matrix, the column rank is always is always equal to the row rank. This common value, which is column rank equal to row rank; that value is called the rank of the matrix.

.

(Refer Slide Time: 11:54)

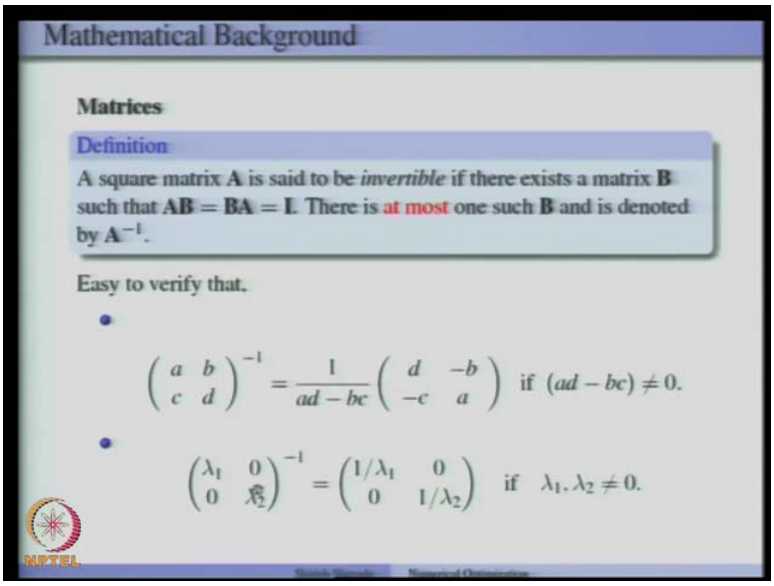


Now, if you consider matrix A, which is the 2 by 4 matrix, you will see that this matrix has a rank 2. For example, the second column is 3 times the first column. Then, the fourth column is A minus 2 times the third column. The first column and the third column are independent. So, you we will see that the rank of this matrix is 2. Now, only the 0 matrix matrix, which contains all 0 s has a rank 0. So, in our course, we will come across some situations where we will have to add some matrix of smallest rank to the existing, some existing matrix.

Now, the matrix with the smallest rank is non rival matrix with the smallest rank is a rank 1 matrix. So, one such example of a rank 1 matrix is given here. So, you will see that the first column is 3 times the second column. The third column in minus 1 times the second column. So, there is only 1 independent vector in the space span by the columns. Therefore, the rank of this matrix is 1. Now, this matrix can be written in this form. If you consider this vector as u and this vector as v, so you can write this as u v transpose. So, every rank 1 matrix can be written in the form A equal to u v transpose.

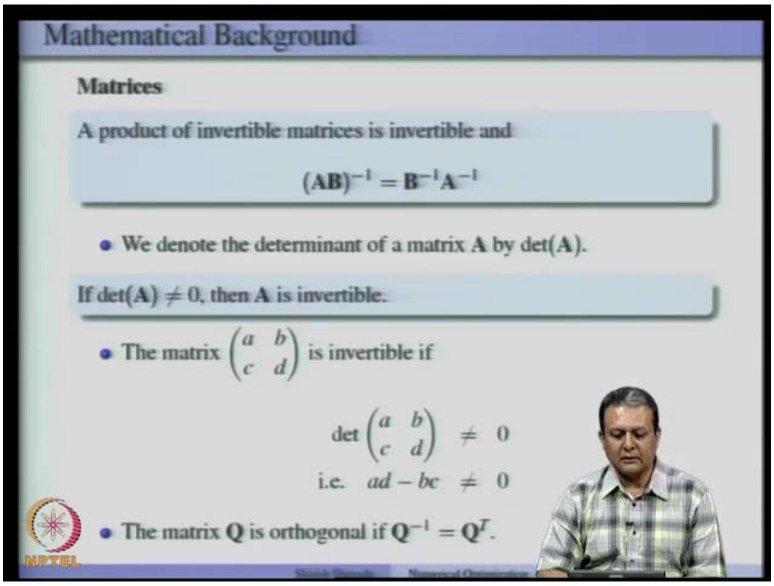
.

(Refer Slide Time: 13:38)



Now, let us assume that we have square matrix A. So, m will be equal to n. now, such a matrix is said to be invertible if there exists a matrix B such that A B is equal to B A is identity and there is at most 1 such B that is denoted by A inverse. For example, you can verify that the inverse of this matrix is nothing but 1 over a d minus b c into the matrix b minus b minus c as if a d minus b c is not 0. If lambda 1 and lambda 2 are not 0, then the inverse of this matrix is 1 over lambda 1, 1 over lambda 2. So, it is easy to see that diagonal matrix matrices, if they have non zero diagonal entries, they can be easily noted for 2 by 2 matrices. One can use the formula like this.

(Refer Slide Time: 14:42)

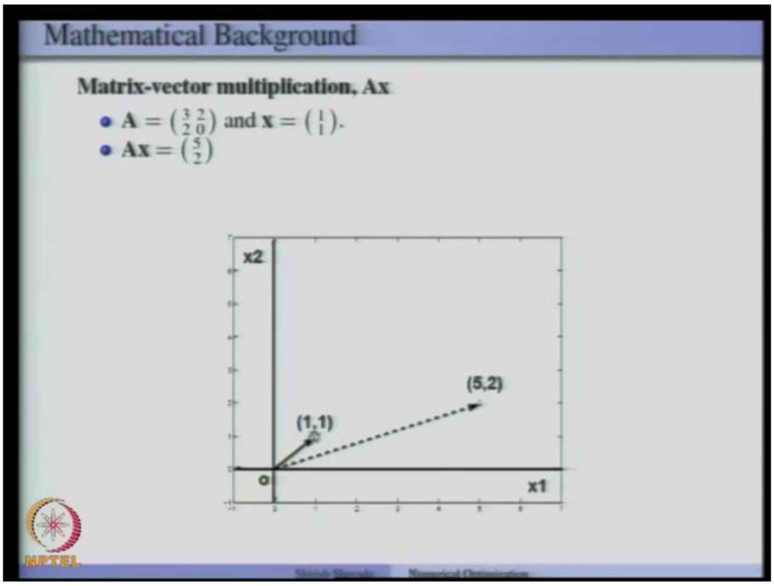


.

Now, a product, if you consider product of 2 matrices when is it invertible, so this product A B is invertible if the individual matrices are invertible. We have this result, which says that A B is inverse is nothing but B inverse into A inverse. So, you will see that the order of the matrices is just when we expand is inverse. So, this is the similar result can be extended if you have multi or you have if you have product of more than 2 matrices. So, the order gets reversed. So, we will denote the determinant of A matrix by determinant of A like this.

Now, there is the important result, which says that if the determinant of A matrices is non zero, then it is invertible. So, this means that this matrix a b c d is invertible if determinant of a b c d is not 0, which means that a d minus b c is not equal to 0. So, you will see that when we talked about this the same matrix earlier, we wrote that a d minus b c should not be 0. So, if the determinant is not 0, then the matrix is invertible. Here is one definition. The matrix Q is an orthogonal matrix if Q inverse is Q transpose. So, every column of this vector of this matrix Q is an ortho normal vector.

(Refer Slide Time: 16:25)



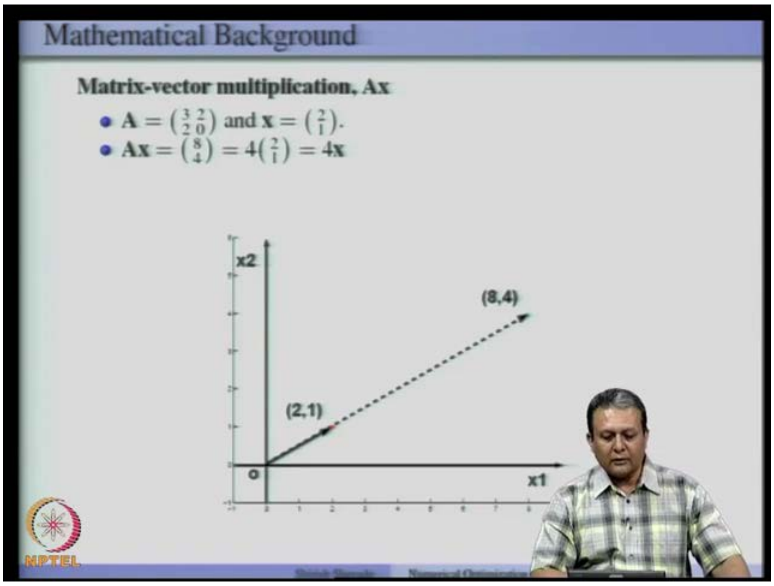
Now, here is an important term said that we will keep using in our course and which the matrix vector multiplications. So, suppose if you are given a matrix A, we will talk about the square matrices. So, if you are given the matrix A and A vector x, what is the effect of the multiplication of A matrix of p, multiplication of a vector by A matrix? So, here is

.

a simple example where you we have A matrix consisting of orally means 2 by 2 matrix and a vector x consisting of 2 elements.

Now, this is shown in the here in the figure. The x vector is shown like this. Now, when we pre multiple x by A, what we get is a vector 5, 2. So, you will see that the original vector got rotated to a new vector and not only that, its magnitude also got increased. In some cases, it may so happen that the matrix will rotate the vector and the magnitude will get reduced. So, if you want to rotate any vector x, it is a good idea to re multiplied by a suitable matrix A, so that the vector gets rotated. So, in our course where we talk about different optimization methods, we will require this concept that the rotation of a direction by a appropriate pre multiplication of a matrix. So, this is an important concept.

(Refer Slide Time: 18:09)



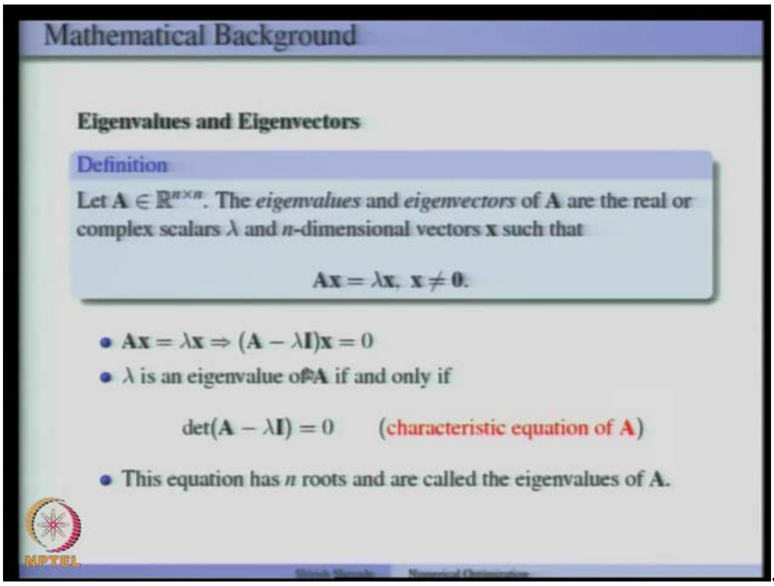
Now, there could be some cases that the pre multiplication of A vector bio matrix Aa will result in the vector in the same direction. So, I have considered the same matrix as we considered in the last case, but I have considered the different vector. So, this vector is x is equal to 2 and 1. Now, if you compute A x, you will see that the x is at 4. That is nothing but 4 times the original vector x. so, what we have seen is that instead of the usual rotation of a vector the u vector is along the same direction as the original vector, only thing is that it got expanded or it got stretched.

So, it may happen sometimes that you could get a vector in exactly the opposite direction. So, minus of 2 1; that is also a possibility. So, it is not always necessary that it

.

will always point in the same direction. It could be a negative of that direction. But, the point is that the original vector gets sometimes stretched or sunk. So, this will lead to an important concept of Eigen values and Eigen vectors.

(Refer Slide Time: 19:30)



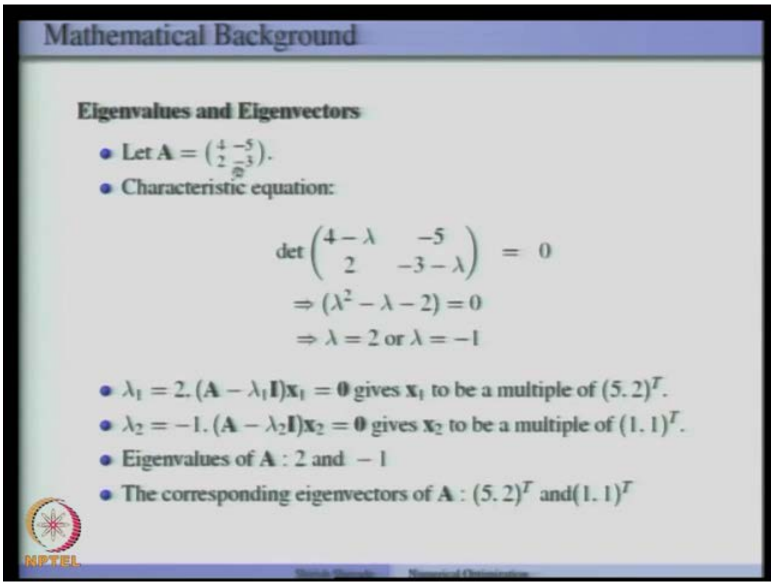
So, suppose that we are given a square matrix n by n. Then, the Eigen values and Eigen vectors of A are the real or complex scalars lambda and n dimensional vectors x such that A x is equal to lambda x. So, we saw in the previous case that x equal to 4 x. So, in this case, x is an Eigen vector and Eigen vector of A and 4 is the corresponding Eigen value. So, if this whole is x equal to lambda x where x not equal to 0, then lambda is called the Eigen value of A and the corresponding Eigen vector is x. Now, if you rewrite this expression, we can write it as A minus lambda I into x equal to 0. So, it means that the matrix m minus lambda I should lose its rank, which will not be a full rank matrix.

So, that means that determinant of that matrix has to be 0. So, determinant of A minus lambda I has to be 0. So, lambda is an Eigen value of A only determinant of A minus lambda is equal to 0. So, this is called the characteristic equation of A. The quantity on the left side is a characteristic polynomial of the matrix A, of the matrix A. So, this polynomials in lambda and if the matrix A is of five is n by n, then this is the polynomial of degree n. Then, by fundamental theorem of algebra, this equation has n roots. Now, remember that these roots, all the roots need not be real roots. Some of them could be

.

complex roots. So, this n roots are called Eigen values. Solving this equation, we will get the Eigen values of A.

(Refer Slide Time: 21:30)

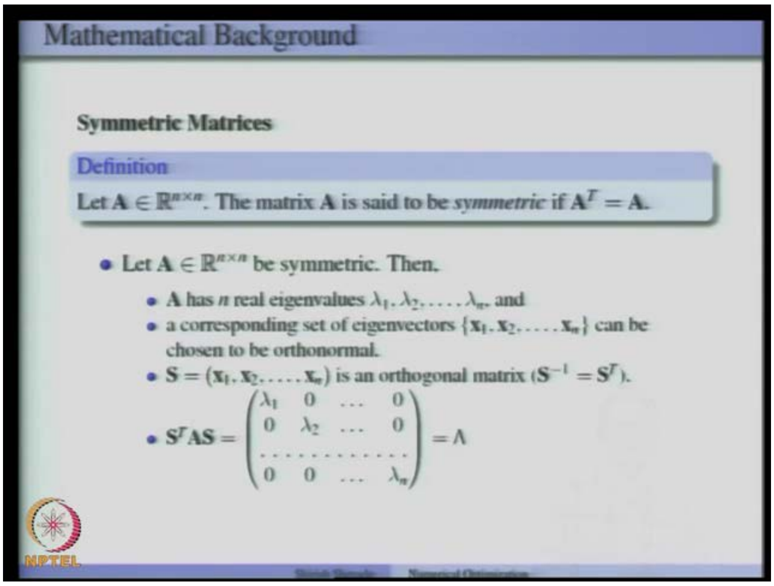


So, here is the simple example. Let us consider a matrix 2 by 2 matrix. So, if you write its characteristic equation, so determinant of A minus lambda I, lambda gets subtracted from the diagonal entries. That gives, that is the 2 degree polynomial or a quadratic in lambda. It has 2 roots, lambda equal to 2 or lambda equal to minus 1. So, these 2 are the Eigen values of this matrix A.

Now, how do we get the corresponding Eigen vectors? So, we substitute this 2 in this equation. So, A minus lambda 1 I into x 1 should be 0. So, if you solve this, you get x 1 to be a multiple of 5 comma 2. Similarly, by putting lambda 2 equal to minus 1 in this equation, we get x 2 to be a multiple of 1, 1. So, this matrix A has 2 Eigen values 2 and minus 1, which are the roots of the characteristic equation. The corresponding Eigen vectors are 5 and 2 and 1 1. Remember that I have said here that x 1 is the multiple of 5, 2. So, any x 1 in this direction is an Eigen vector.

.

(Refer Slide Time: 22:49)



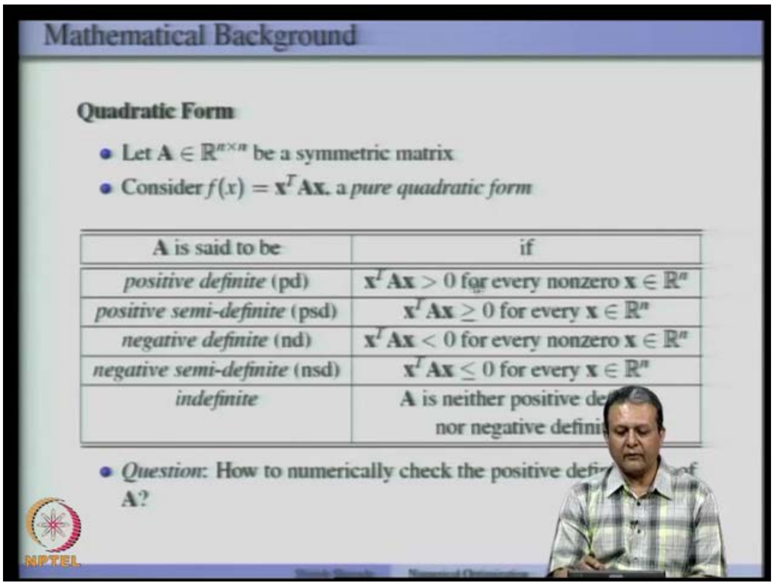
So, in this course, we will come across symmetric matrices, which have some special properties. So, a matrix is symmetric if the transpose of the matrix is same as the original matrix. So, remember that we are talking about you know n by n matrices. Now, one can show that a symmetric matrix has real Eigen values. So, these are lambda 1 to lambda n. Then, corresponding to the Eigen values, I can choose a basis, which is ortho normal.

So, if the Eigen values are distinct, then I can show that the Eigen vectors can be the Eigen vectors have to be orthogonal to each other. Then, if the Eigen values coincide, some side of Eigen values coincide, then corresponding to them, one can choose an orthogonal set of basis. So, once we have orthogonal set, if basis for the space, then one can use procedure like this procedure to get a ortho normal basis from this.

Now, if you have a matrix S, where the Eigen vectors of x 1 2 Eigen vectors of the matrix A x 1 to x n are arranged column wise, then you can verify that S inverse is nothing but s transpose. This is because it is a orthogonal matrix. So, S transpose S can be written like a diagonal matrix, where the diagonal can tend the Eigen values of the matrix A. Let us call this matrix as lambda.

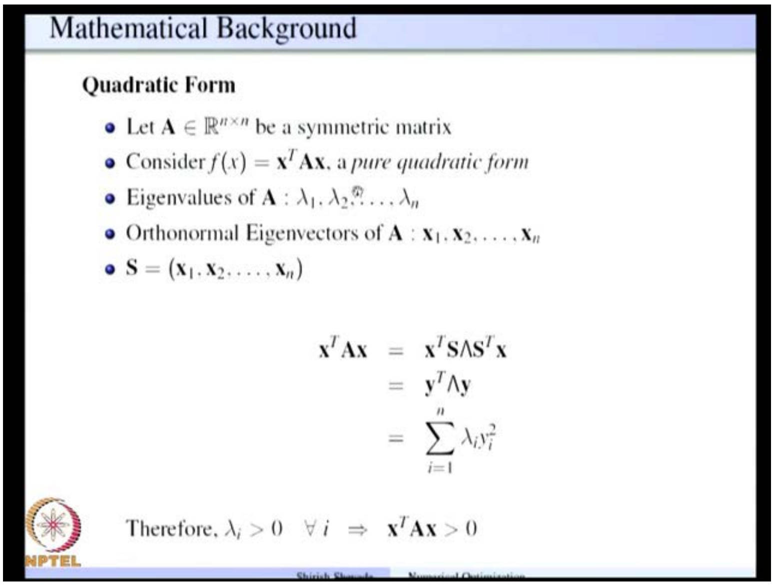
.

(Refer Slide Time: 24:46)



Now, let us look at quadratic form. Consider a symmetric matrix A. The quadratic form A, pure quadratic form can be written as f of x is x transpose x. Now, here is a definition is says that A is said to be positive definite if x transpose x is greater than 0. That means that quadratic form has to have positive value for every non zero x in R n. Similarly, we have definitions for positive semi definite matrices negative definite and so on. Now, the question is that how do we check whether a matrix is a positive definite because this definition is very difficult to verify. It says that for every non zero x in R n, this should hold. So, it is very difficult to verify.

(Refer Slide Time: 25:34)

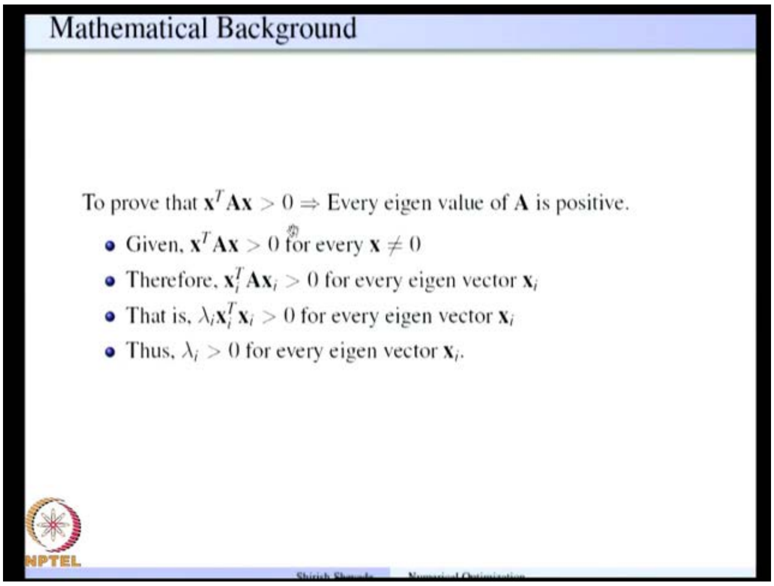


.

So, in the pure quadratic form, where f x is x transpose A x, let the Eigen values of A be lambda 1 to lambda n. Let us consider the corresponding ortho normal Eigen vectors of A, which are x 1 to x n. Since, x 1 to x n are Eigen vectors are non zero, so note that A is a symmetric matrix. Now, if we arrange the Eigen vectors in the matrix S in the form n columns, then the matrix S will look something like this, where the n Eigen vectors of A are arranged.

Now, if you consider the quadratic form x transpose A x and if you write A as A lambda A transpose as we saw earlier, then x transpose x can be written x transpose S lambda S transpose x. Now, let us define y to be S transpose x. Since, the x s are deign vectors, they are non zero. So, transpose x is a non zero vector. We have x transpose x written as y transpose lambda y. You expand it further. We get that sigma lambda y i square. So, this matrix lambda is a diagonal matrix and on its diagonal is the Eigen values lambda 1 to lambda n of the original matrix A. So, the quadratic form is simplified to sigma lambda y r square. If lambdas are positive, then we have x transpose x to be greater than 0 because y square is always a positive quantity since y is obtained using S transpose x. Now, whenever we have lambda is greater than 0, x transpose x is greater than 0. Now, it is the converse to this.

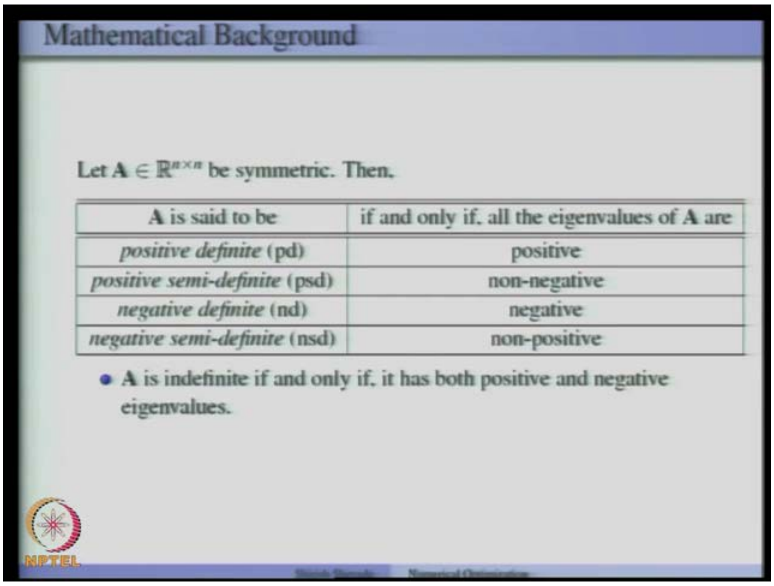
(Refer Slide Time: 27:55)



So, we show that x transpose x is greater than 0 implies that every Eigen value of A is positive. Now, note that we are given external source x is greater than 0 for every x non zero in the n dimensional space. Since, that whole for A, every x in the n dimensional space, which is non zero can be written as that x r transpose x i is better than 0 for every Eigen vector of x i. That can be written.

Since, x i is Eigen vector and the corresponding Eigen value is lambda A, so we can write x r transpose x greater than 0 implies that lambda i into x i transpose x i is greater than 0 for every Eigen vector x i. Since, x i transpose x i is non zero, lambda has to be greater than 0 for every Eigen vector x i. So, this shows that matrix A, which is symmetric is positive definite if and only if, all its Eigen values are strictly positive. Now, this result can be extended to positive semi definite matrices, where we say that symmetric matrix A is positive semi definite if and only if, all its Eigen values are non limitable.

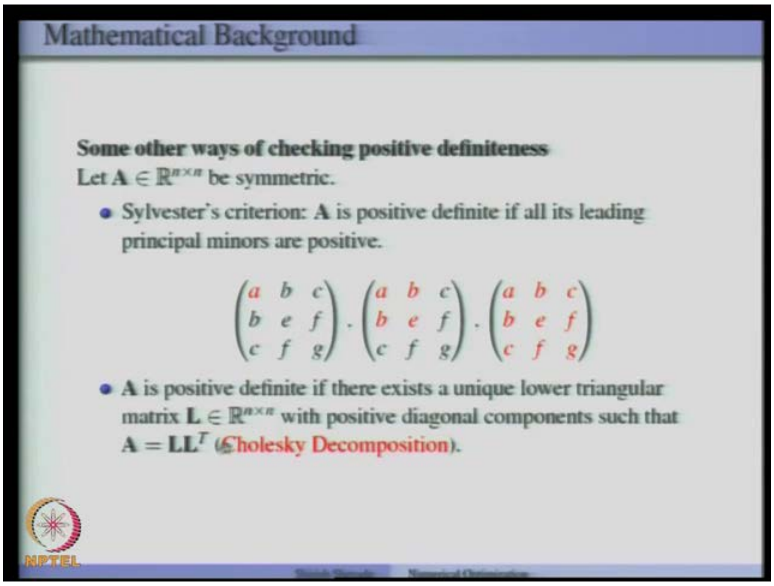
(Refer Slide Time: 29:24)



So, this is the same result what we have written earlier. But, it is said to be positive definitive. x transpose x is greater than 0 can be equivalently written as A is positive definite if and only if, all the Eigen values of A are positive. Similarly, one can write the other result in terms of Eigen values of A. Now, A is indefinite if and only if, it has both positive and negative Eigen values.

.

(Refer Slide Time: 29:52)



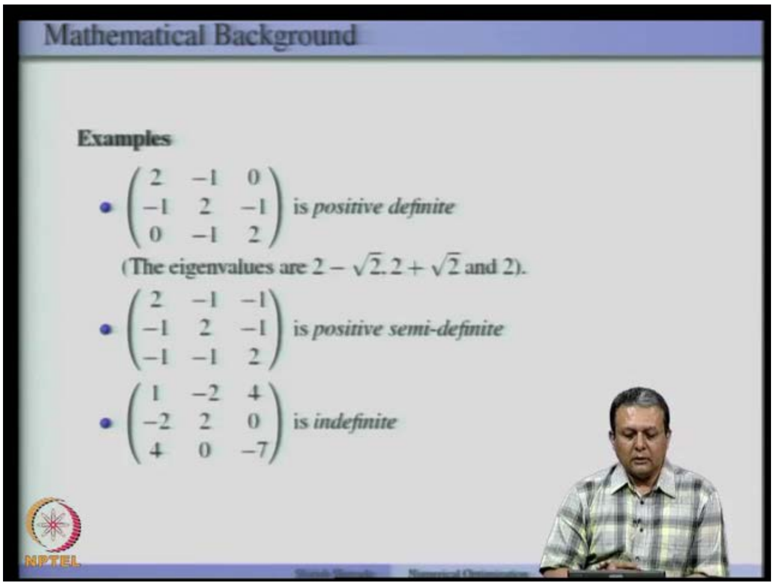
Now, apart from checking the Eigen values of matrix side to find out whether matrix is positive definite or not, there are some other ways. So, again we start with a symmetric matrix. So, there is a criterion called Sylvester's criteria. He says that A is positive definite if all the leading principle minus are positive. So, let us consider 3 by 3 matrix. So, what it says is that you take the leading matrices and take the principle minor. Then, you take the determinant of those leading matrices.

So, leading L by L matrix A, it is it is A L by L matrix. So, its determinant is A that has to be positive. Then, you take a leading 2 by 2 matrix, which consists of a, b, b, e. So, determinant of these should be positive and then the determinant of the third leading matrix that also has to be positive. So, if all the determinants are positive, then according to Sylvester's criteria, the matrix is positive definite.

There is another important criteria, which says that if if there exist a unique lower triangular matrix L, which is again of the size n by n such that it has all positive diagonal components, then A is equal to L L transpose. Then, A is said to be positive definite. So, this decomposition of the matrix A into L L transpose, this is called cholesky decomposition. It is a very useful concept; especially while solving sustains of equations. So, remember that this L matrix is unique lower triangular matrix with positive diagonal components A is decomposed has L A is equal to L L transpose.

.

(Refer Slide Time: 31:50)

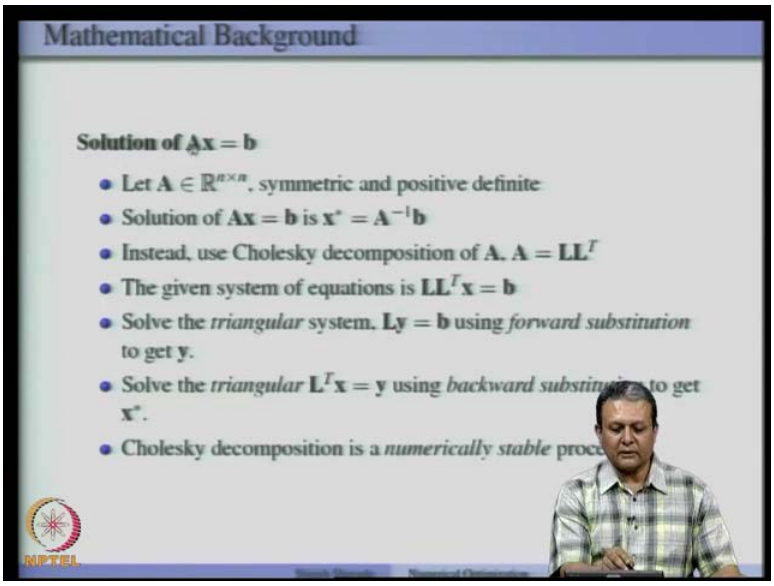


Now, here are some examples. So, this matrix is positive definite. One can also find its Eigen values that are known to be 2 minus root 2, 2 plus root 2 and 2, which are clearly positive. So, this matrix is positive definite. One can also look at the Sylvester's criteria. So, if you take 2 is positive, then, that determinant of this 2 by 2 matrix is positive and the determinant of 3 by 3 matrixes is also positive.

Now, you look at this matrix. So, I have just changed this 0 to minus 1 here. Now, if you look at the Eigen vales of this matrix, there are 3 Eigen values. One of them is 0. So, this makes this matrix a positive semi definite matrix where because the other 2 Eigen matrix are Eigen values are positive. Now, here is another example where the matrix is indefinite. So, it has some positive Eigen values and some negative Eigen values.

.

(Refer Slide Time: 32:53)



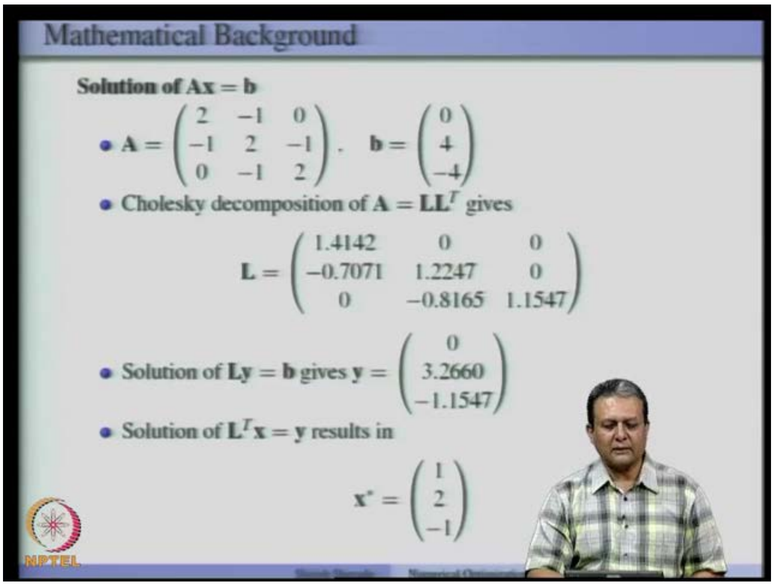
Now, so in this course we will require to solve this system of equations x is equal to b. So, let us assume that they are consistent. The solution exists. So, there is 1 way to solve. So, let us assume that A is symmetric positive definite matrix. Therefore, what one can do is that one can simply take the inverse of A. Since, it is positive definite, all the Eigen values are positive determinant is non zero. So, one can take the inverse of this matrix. So, x star is equal to A inverse b, but this low separation is numerically non trustable operation.

So, instead of that, what one can do is that one can use cholesky decomposition of A. So, suppose that the cholesky decomposition of A is L L transpose. Then, the given system of equation can be written has L L transpose x is equal to b. Now, this system of equations can be solved in 2 steps. So, first solve the triangular system L y equal to b. Remember that L is a lower triangular matrix. This system of equations can be easily solved using forward substitution to get y.

Once you get y, then L can solve a transpose x is equal to y to get is in backward substitution to get x star. This cholesky decomposition is a numerically stable operation and that is always preferred. So, whenever we want to find out something like A inverse b, so one can treat it has the system of equation x is equal to b and then assuming if the matrix is symmetric positive definite. Then, one can do the cholesky decomposition of that and then solve it.

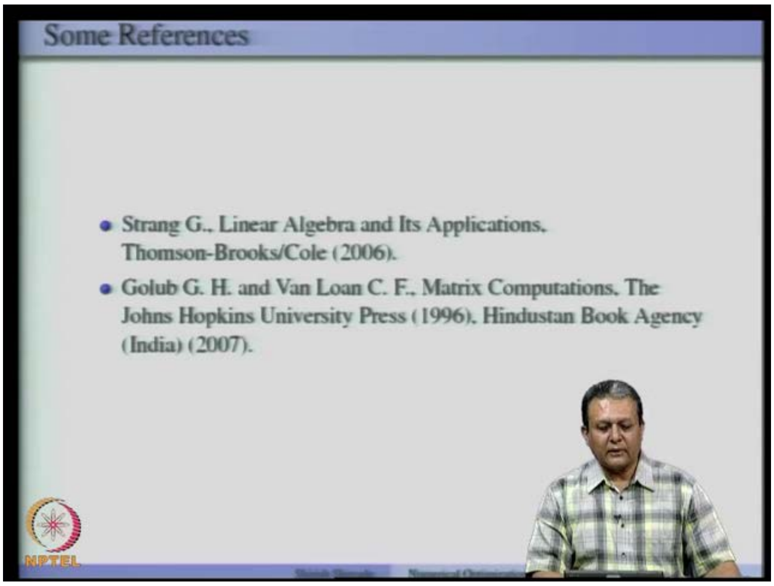
.

(Refer Slide Time: 34:43)



So, here is the simple example, which is worked out. So, I have symmetric matrix, which is positive definite and b is given vector. So, x equal to b is the system that we want to solve. So, we do the cholesky decomposition of the corresponding L matrix is like this. Now, you will see that all the entries are in the diagonal are positive. Now, we solve L y. The system L y is equal to b. Since, L is lower triangular, it is very easy to solve this system of equations. After having obtained L y, one can solve L transpose x is equal to y, which gives us x star is 1, 2, minus 1. So, you can check that this is indeed the solution of the given A x is equal to b.

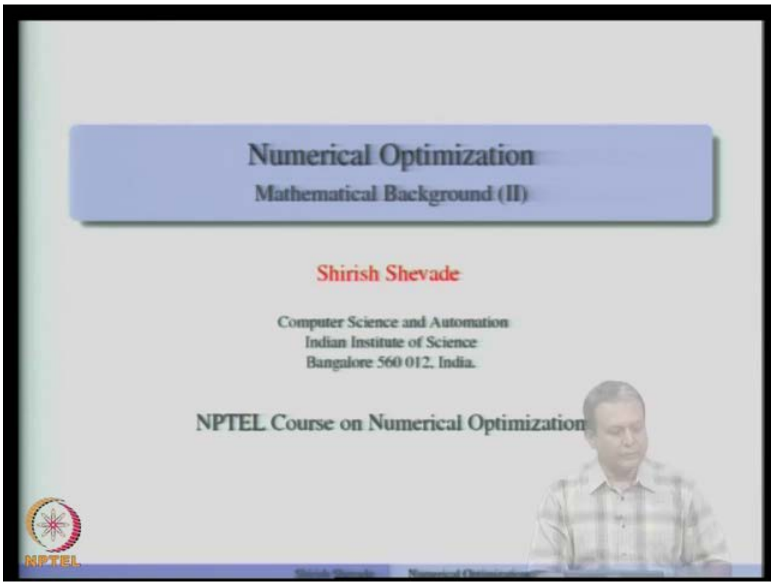
(Refer Slide Time: 35:30)



.

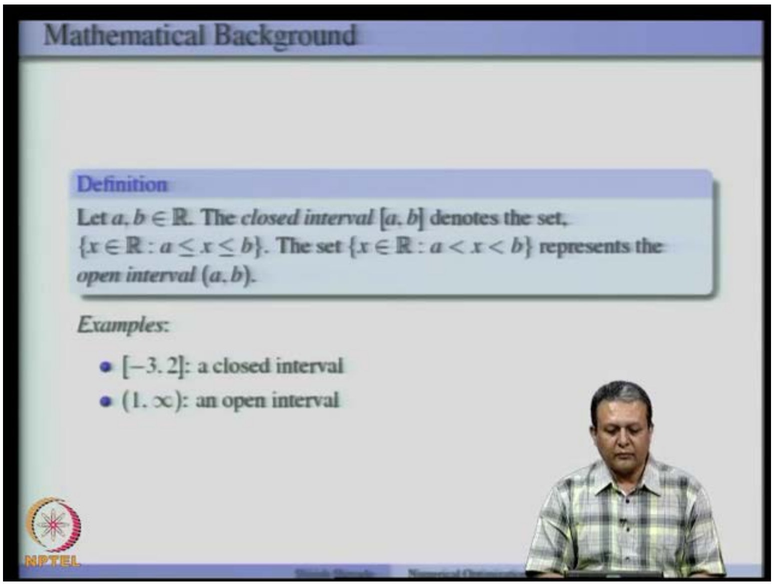
Now, here are some references for linear algebra. So, there is a good book by Strang and also there is a book on matrix competition by Golub and Van Loan. So, you will see some details about cholesky decomposition and other decompositions, other types of decomposition of symmetric positive definite matrix in this book. So, we will now look at some other background, which will be needed for this course.

(Refer Slide Time: 35:58)



We will see this now.

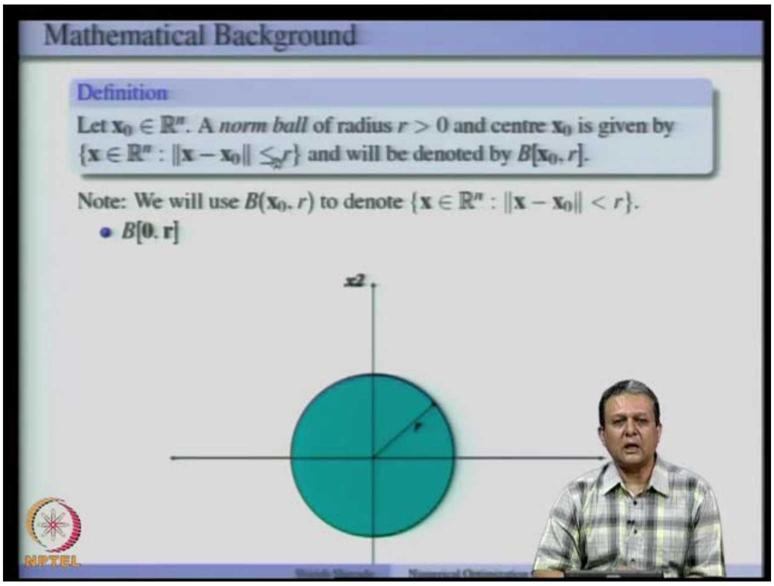
(Refer Slide Time: 36:02)



.

So, this is mainly from the calculus and analysis view point. So, suppose a and b are real numbers. Then, the close interval a, b, which is been denoted in the square bracket, a comma b is a set real numbers as that a is less than r equal to x less than r equal to b. The set a less than x less than b is will be called and open interval. So, here is an example of a close interval minus 3 comma 2. So, it can solve real numbers within the range minus 3 to 2 and inclusive of minus 3 and 2. Here is an example of an open interval, which contains all the elements in the range 1 to infinity and 1 and infinity not included.

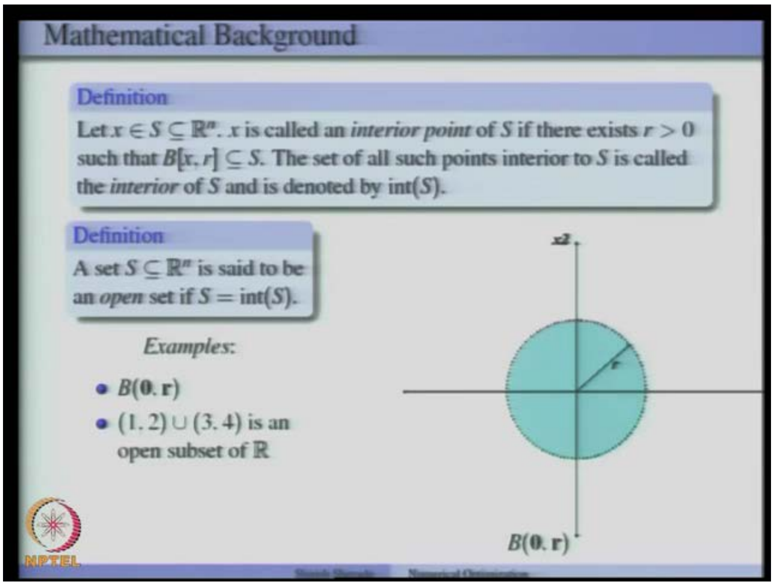
(Refer Slide Time: 36:58)



Now, we define a norm ball. So, a norm ball of radius r greater than 0 and center x naught, where x naught is the point in R n is a set of points, which are at the distance at most r from naught. So, this norm defines the distance. So, we collect all those points, which are at a distance at most x r from x naught. We will denote it by a closed ball because we are taking the points on the boundary also, which we will denote it by B open square bracket x naught comma r, close the bracket. Similarly, we define an open ball, where we can consider all the points whose distance from x naught is less than r. So, I have shown here a ball of centered around origin. So, it contains all the points on the boundary as well as in the interior. So, this is an example of a closed ball and the norm defines the distance.

.

(Refer Slide Time: 38:18)

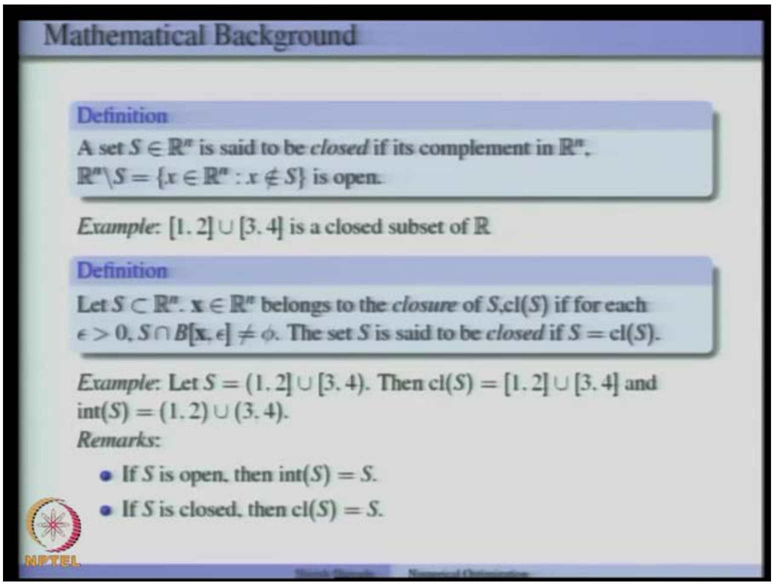


Now, let us define what is called interior point. So, let x be any point of the set S. So, it is called the interior point of S, if there exists some r greater than 0 such that the close ball of radius r around x is contained in x. So, if there exists 1 r such that this close ball lies in the set S, then x is called a interior point. Now, the set of all such points, which are interior to S is called the interior of S.

Now, if you look at this, so this is a open ball centered at 0 and radius r. now, the boundary is shown here by a dotted line. Now, the interior point r are the points which are in which are inside this boundary. So, interior point of this side is shown by a shaded region here or colored region here. Now, the set is said to be open set if this set is same as the interior of the set. So, this open ball is a open set. Another example of a open set is say union of the intervals 1, 2 and 3, 4. So, this is an open set.

.

(Refer Slide Time: 39:45)



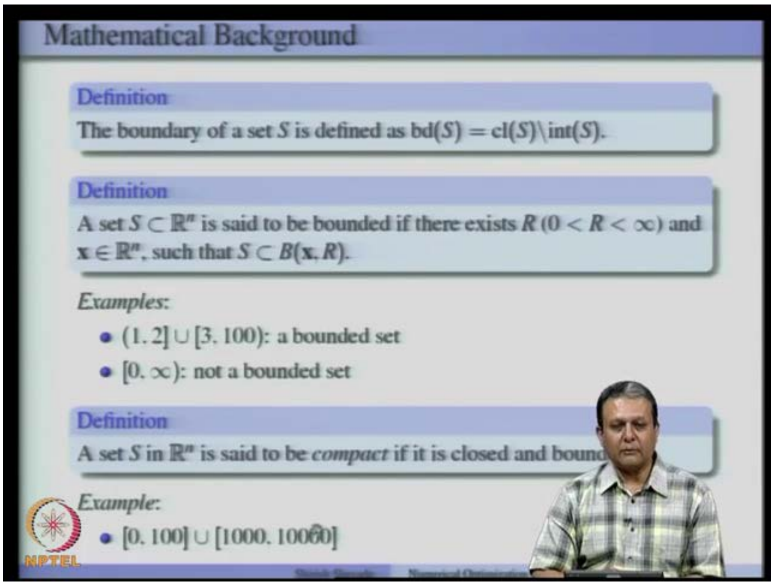
So, one can also define close set. So, a set S in R n is said to be closed if its complement in R n is open. So, the complement of R n is all those elements in R n, which are not in complements of S in R n is all those elements in R n, which are not in S. So, this union of these closed intervals is also close set. A close interval is a close sub set of R. Now, we define the closure of a set.

So, the closure of a set is defined in such a way that if x is a point in if x is a point in R n and is the sub set of R n, then x belongs to the closure of S. We will denote it has c l of S if for each epsilon is greater than 0, the intersection of S with the close ball of radius epsilon around x is not not empty. The set is closed if and only if S is equal to closure of x.

So, if you take this example, then the closure of set is union of the 2 closed intervals and the interior of this set S is the union of the 2 open intervals. Now, the set is open if its interior of that set is equal to S and it is closed if closure of S is S. Remember that we are working with a set of real number, sub sets of real numbers. So, most of the results are easy to follow.

.

(Refer Slide Time: 41:27)

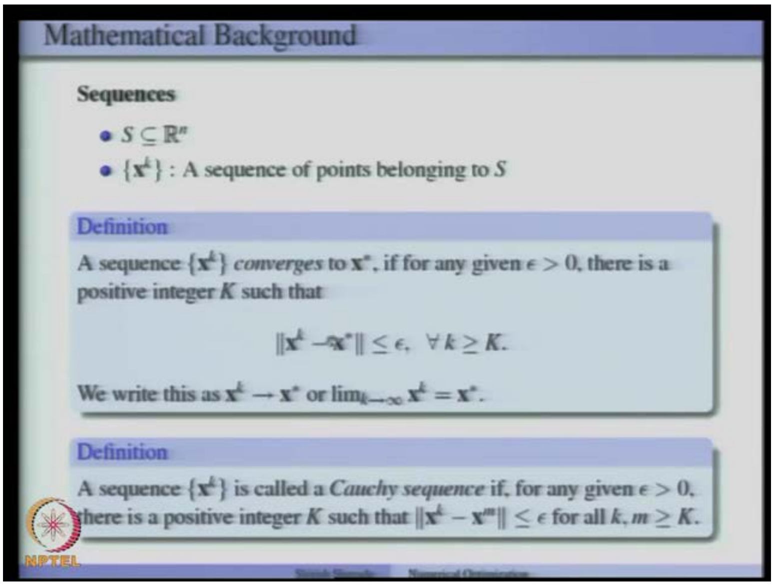


Now, the boundary of the set is defined as the closure of S minus interior of S. so, if we take this example S, so the closure of S is like this and the interior is this. So, closure of S minus interior of S minus interior of S will consist of 4 points 1, 2, 3 and 4. So, the others will others will be others will form the boundary. Now, a set is said to be bounded of it can be put in a ball of finite non zero radius.

So, here is the example of a bounded set. So, this set can be put in a ball of finite radius. On the other hand, this this interval is not a bounded set. It cannot be put in a ball of finite radius. Then, in R n, we need one more important concept, which is called a compact set. So, a set S in R n, S is said to be compact if it is closed and bounded. So, here is an example of a compact set, where we have a closed set. Also, it can be put in a ball of finite radius.

.

(Refer Slide Time: 42:38)

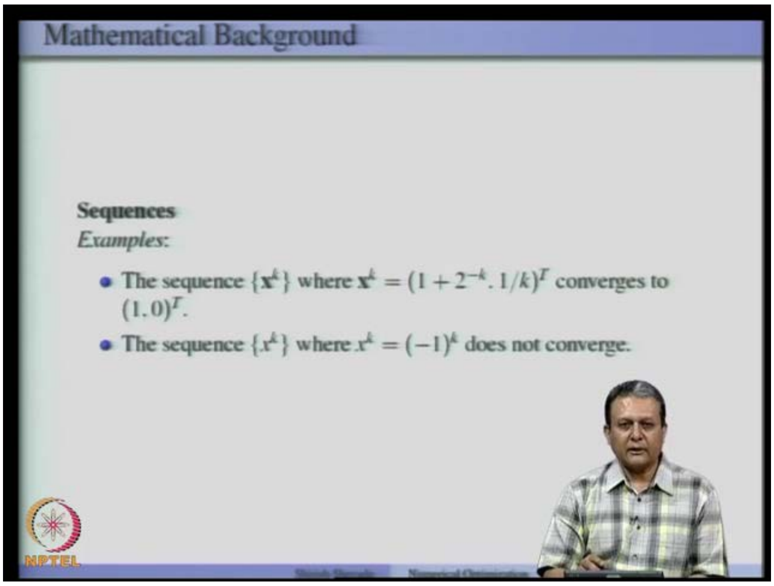


We move on to the definition of sequences. So, let be S be a sub set of R n and x can be a sequence of points belonging to the set S. So, the sequence x k converts to x star, if for any given epsilon is greater than 0, there exist of positive number K, which could be which could be large such that the distance between x k and x star is less than or equal to epsilon for all the elements of the sequence index whose index is greater than k.

So, the sequence, we will see that the points in the sequence the points in the sequence are ridicules to x star there at the distance of the most epsilon. So, we write this as x k tends to x star or limit has k tends to infinity x k equal to x star. Now, a sequence is called a Cauchy sequence if, for any given epsilon, there is a positive integer so that if you take any 2 points of that sequence beyond K, then the distance between them is at the most epsilon.

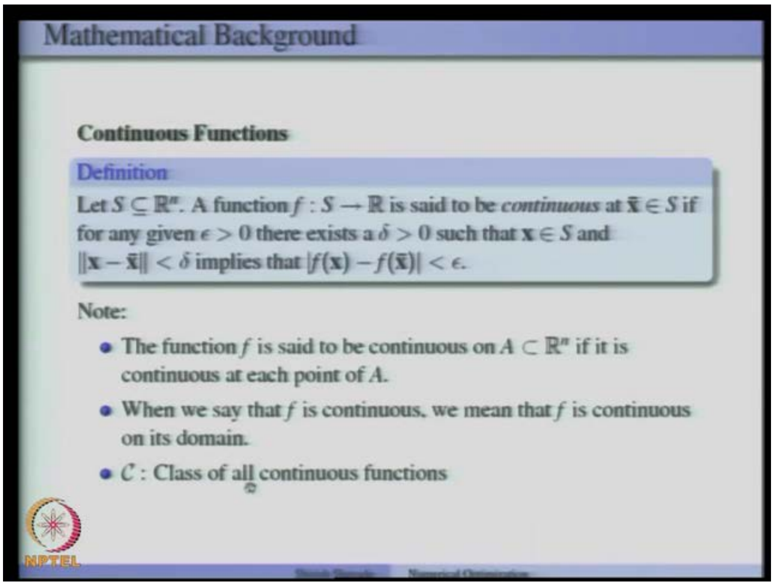
.

(Refer Slide Time: 43:51)



So, we have a sequence, which is given here. So, you will see that this sequence converges to 1 comma 0. On the other hand, if you take a sequence, which condense points minus 1, 1, minus 1, 1, so this does not convert to a point.

(Refer Slide Time: 44:14)



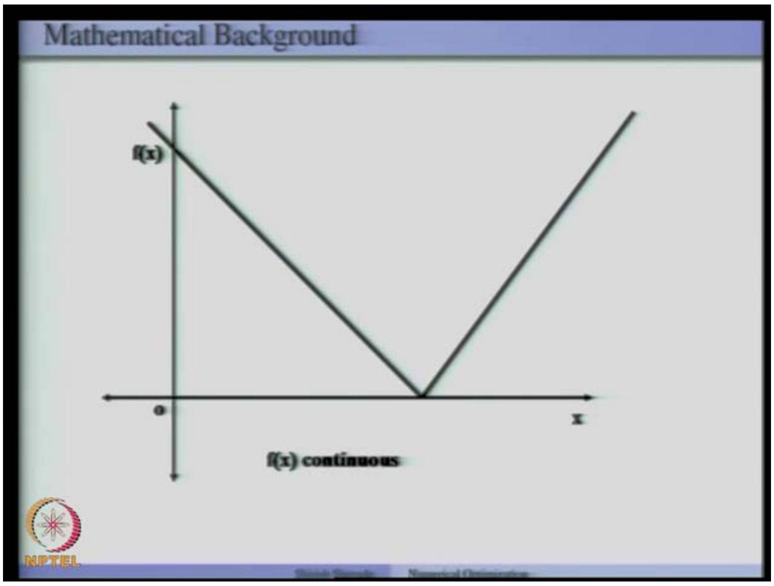
Now, we will look at the definitions of continuous functions. So, let S be a sub set R n and function define from S to R. Now, such a function is continuous at x bar. If given any epsilon greater than 0, there exists some data such that if x belongs to S. Then, the

.

distance of x from x bar is less than delta implies that the function value has the difference in the function value at those 2 points is at the most epsilon.

So, for any epsilon, there exists some delta shows that the distance between x and x bar is less than delta. It implies the difference between the function value is at the most epsilon. Now, we say that the function f is continuous on A, if it is continuous at each point of A. When we say that f is continuous, we mean that it is continuous on its entire domain. So, let C denote, script C denote the class of all functions.

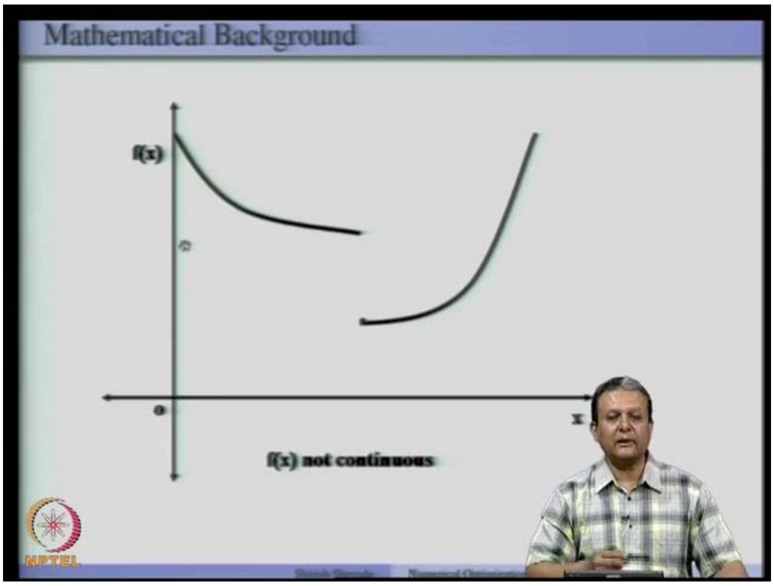
(Refer Slide Time: 45:20)



Now, here is an example of a continuous function. So, you will see that you take any epsilon greater than 0. Given any epsilon greater than 0, there exists some delta, which is greater than 0 such that if you take any 2 points in the domain x and x bar, then the distance between them is at the most delta. That implies the distance between the f s will be at the most epsilon. So, if you take any point x bar, then the distance between x and any x and x bar, if it is less than delta; that implies that the function value the differences at the most epsilon.

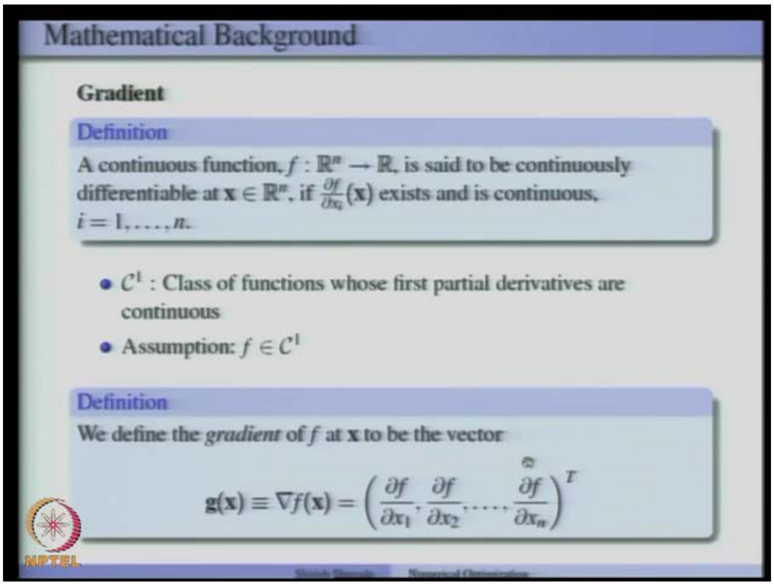
.

(Refer Slide Time: 46:05)



Now, here is the definition. Here is an example of not continuous function. So, you will see that a function, which is not continuous, has these breaks. You can verify that the definition of continuity does not hold here.

(Refer Slide Time: 46:25)

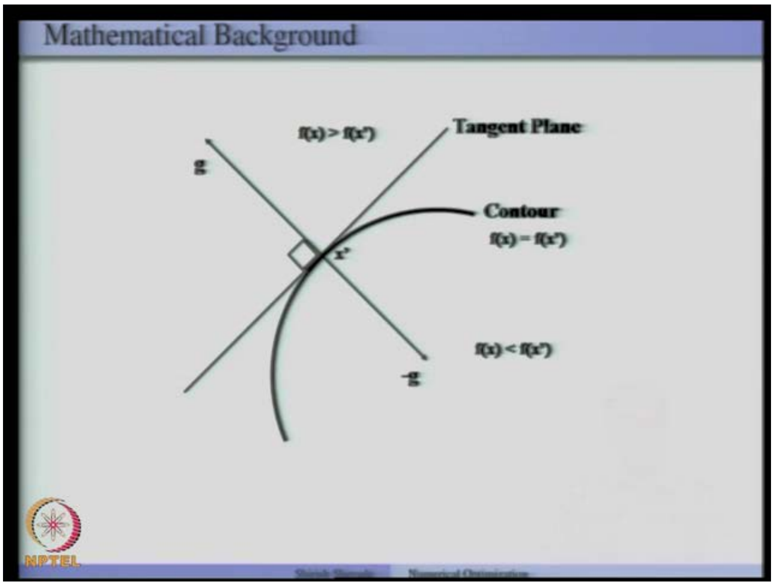


Now, we look at another important concept, which is a key concept. It will be used quite often in our course. That is called the gradient. So, suppose that we have a continuous function from R n to R. That is said to be continuously differentiable at x if the partial of f at with respect to the individual x size at x exists and it is continuous. So, let us denote

.

the class of function whose first partial derivative to our continuous price script C 1 and it is assumed that f belongs to C 1. Then, we define the gradient of f at x to be the vector. So, we will denote the gradient by g x, which is same has gradient of f of x. so, you take the partial derivative of f with respect to x 1 with x respect to x 2 and so on. That is a gradient is the column vector.

(Refer Slide Time: 47:27)

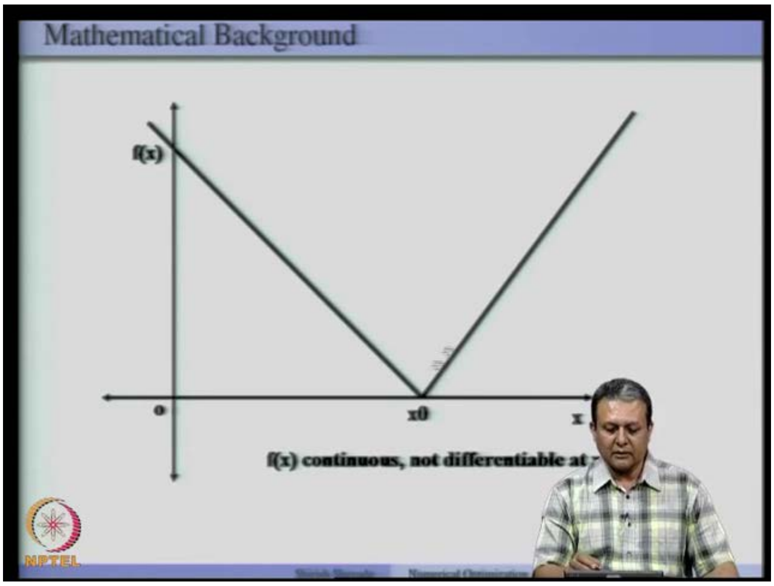


Now, here is the interpretation of gradient. So, we have this function passing through x dash. So, the equation of this function is f x equal to f of x dash. So, this is our contour of this function f of x. Now, at x dash, we will draw a tangent line. So, this is the tangent line at x dash. So, a tangent line also can be thought of has the approximation of the function f linear, a fine approximation of a function f at x dash. Now, we take a perpendicular to this tangent.

Now, on side of this tangent, f of x is greater than f of x dash. On the other side f of x is less than f of x dash. So, you will see that there is a gradient always points towards the direction, where the function value increases and the negative of the gradient points in the direction, where function value decreases. So, this concept will be useful when we talk about the minimum of a function.

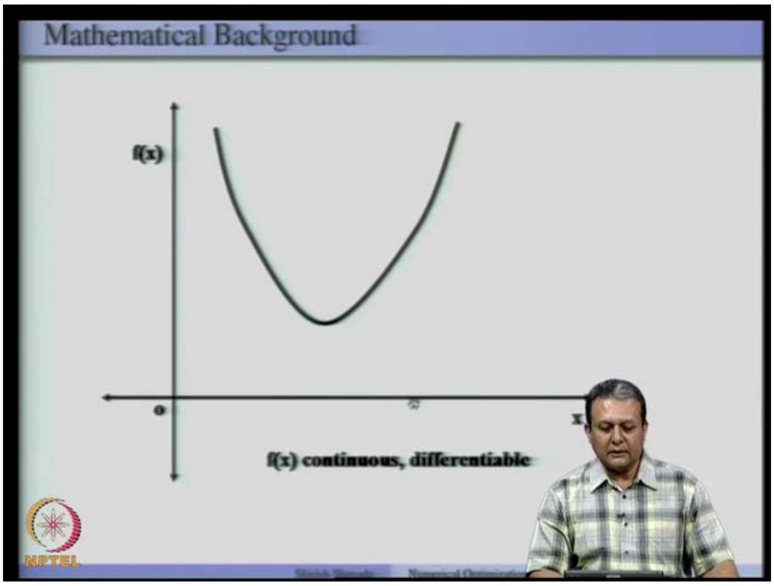
.

(Refer Slide Time: 48:40)



Now, here is a function where which is continuous. But, it is not differentiable at x naught, while you see that this function is continuous at every other point except at x naught.

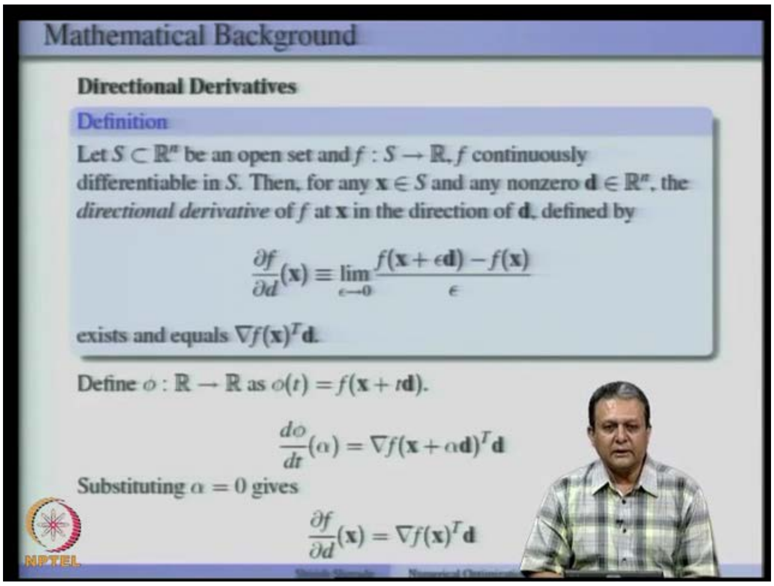
(Refer Slide Time: 48:50)



Here is the example of a function, which is continuous as well as differentiable.

.

(Refer Slide Time: 48:59)

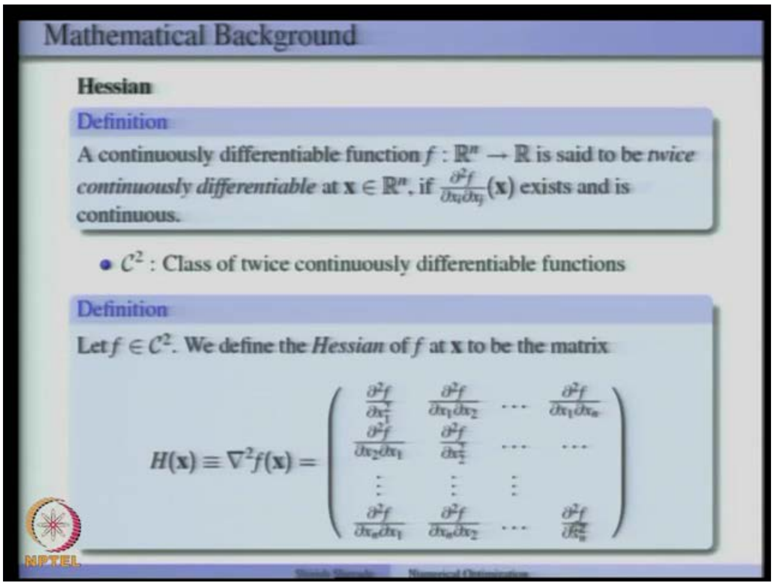


We will also need the concept of directional derivatives. So, let S be a subset of R n. We have an open set and f will be a function from S to R. That is continuously differentiable. Then, for any x and any non zero d, the directional derivative of f in a direction d is defined as partial of f with respect to d validated at x. So, this is nothing but limit has epsilon tends to 0, f of x plus epsilon d minus f of x by epsilon. So, you make a small movement epsilon along the direction d from x. In the limit has epsilon tends to 0, how does this behave?

So, this is called the directional derivative of f at x along the direction d. This equals gradient of x transpose d or g, g of x transfers d. It is easy to see that suppose that if you define a new function phi from R to R, which is parameterized using d phi t is nothing but f of x plus t d, then delta f by delta d at alpha equal to 0 will give us gradient of f of x transpose d. That is the directional derivative of f at x along the direction d.

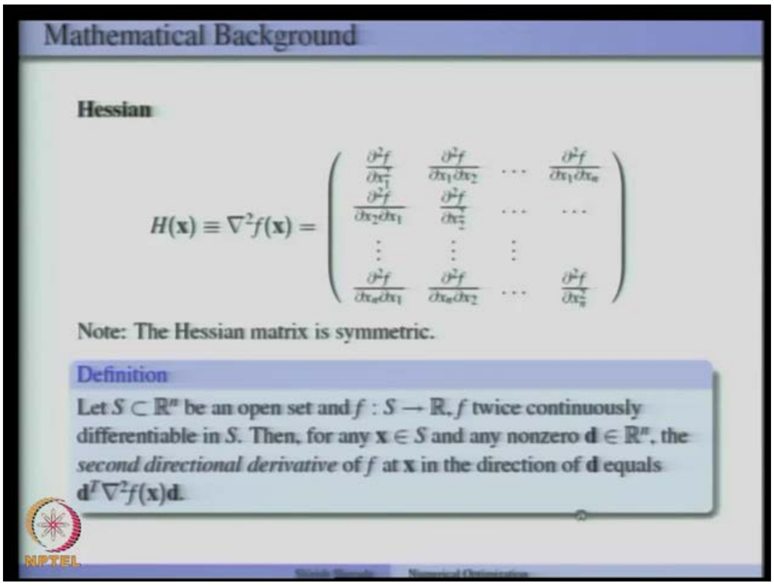
.

(Refer Slide Time: 50:25)



We now move to the second order derivatives. So, the function which is continuously differentiable is said to be twice continuously differentiable, if the secondary derivative exist with respect to every pair x i and x a. So, let C 2 denote the class of twice continuously differentiable functions. Let f be f belongs to C 2, and then we define the hessian of f at x to be the matrix containing all this secondary vectors. So, you will see that this is again n by n matrix.

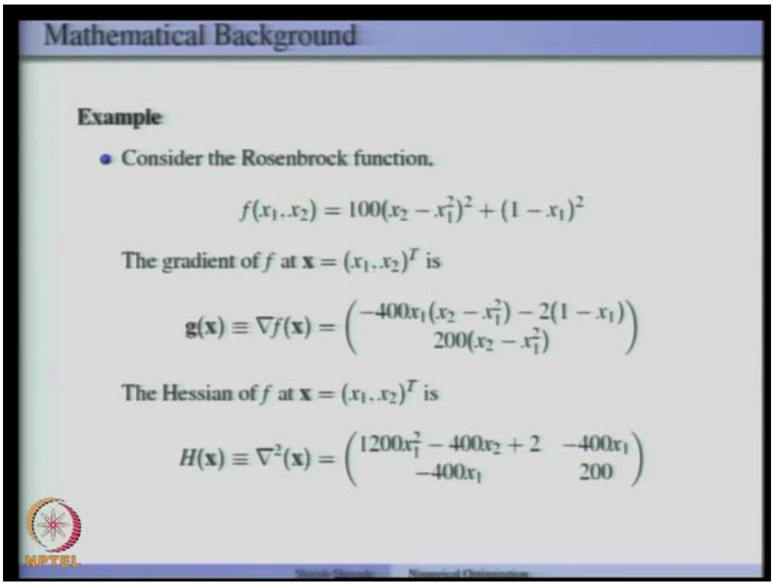
(Refer Slide Time: 51:03)



.

You will also see that this is the symmetric matrix. So, this is the important thing. So, that is why, we studied some of the properties of the hessian symmetric matrices when we reviewed a linear algebra background. Now, similar to the first directional derivative, one can have a second directional derivative and that derivative that derivative transpose to be d transpose brad square f x into d.

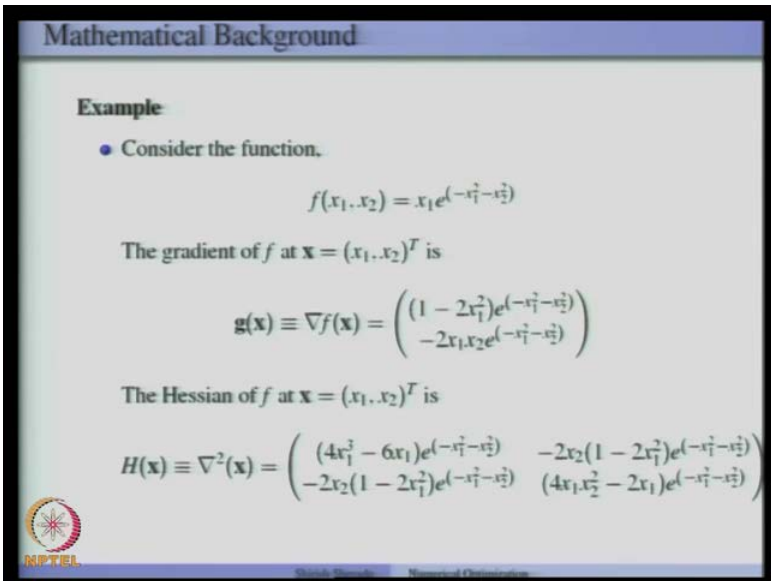
(Refer Slide Time: 51:33)



Now, here are some examples. So, we say we have seen the Rosenbrock function in the first lecture. That function is like this. So, its gradient at any point x 1, x 2 turns out to be this. The hessian turns out to be this matrix. Now, you will see that the hessian is a symmetric matrix. So, that is very important.

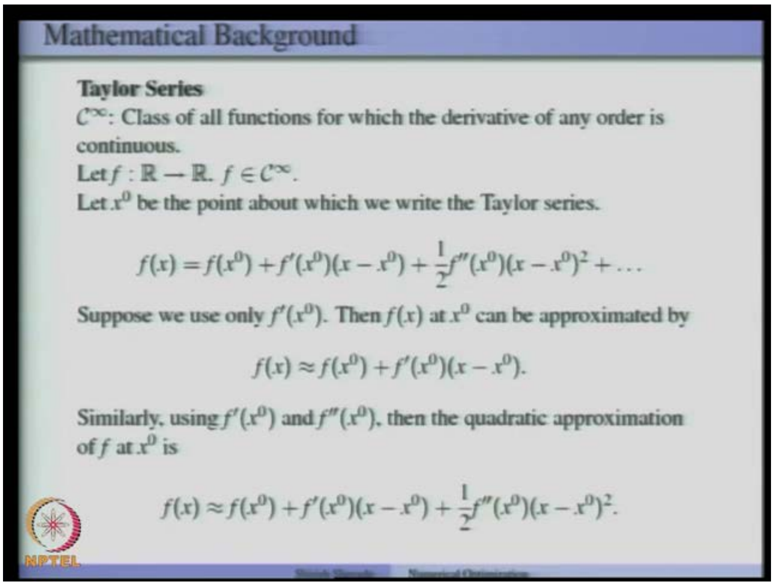
.

(Refer Slide Time: 52:01)



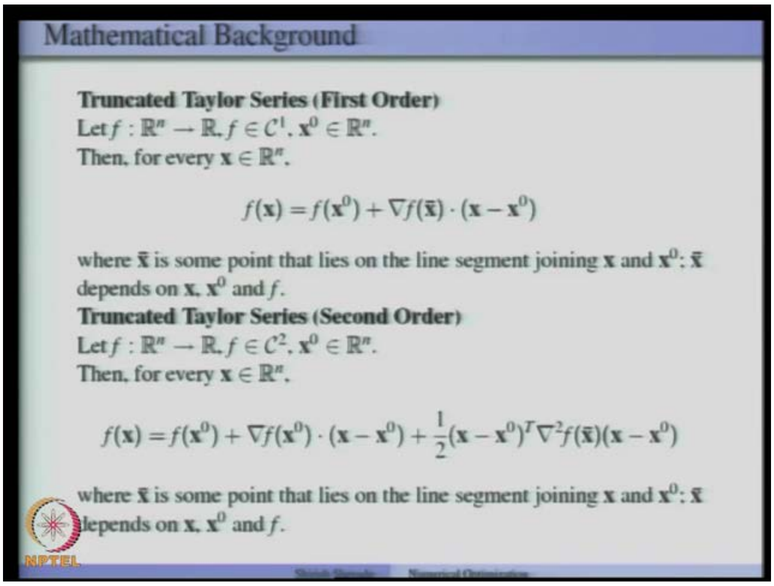
Let us look at another function, which is f of x 1 x 2 is x 1 into x e to the power minus x 1 square minus x 2 square. So, here are the gradients and the hessian a gradient vector and the hessian of this function evaluated at any arbitrary x 1, x 2.

(Refer Slide Time: 52:23)



So, we will now need a concept of Taylor series. So, suppose that we consider a class of functions, which are whose derivative of any order is continuous. So, let f be a function from R to R. It belongs to C infinity. Then, we can write the Taylor series of f around x naught in this form. So, it is an infinite series. Now, suppose that we decide to use only f dash x naught. Then, the function value, the function f of x at x naught can be a fine function using only the first 2 terms. So, we do not use the remaining terms. Similarly, the quadratic approximation of f has x naught can be written like this by ignoring the remaining terms.

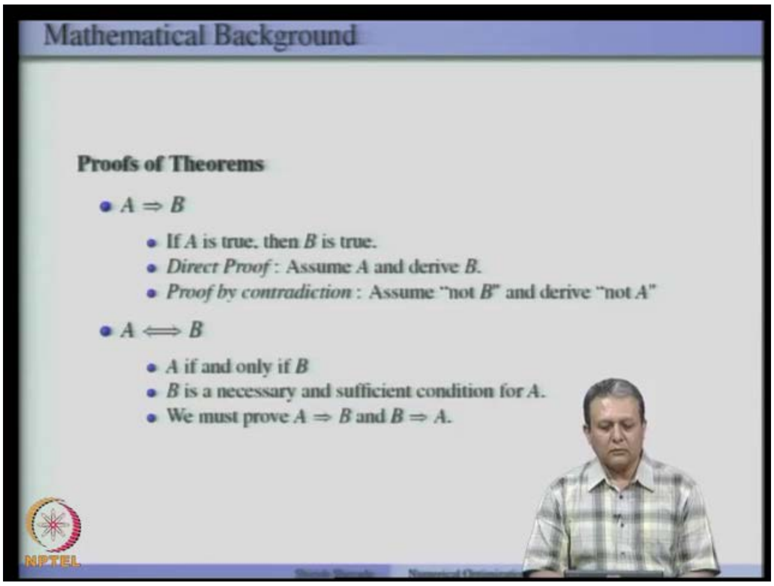
(Refer Slide Time: 53:09)



Now, we will see what the truncated Taylor series are. So, truncated Taylor series restricted basically to the functions were the f belongs to C 1 and it belongs to C 2. So, this is the truncated Taylor series of first order. So, f of x around x naught can be written has f of x equal to f of x naught plus gradient of f of x bar into x minus x naught, where x bar is the point on the line segments joining x and x naught. It depends on f x and x naught. It depends on f x and x naught. Now, these ideas can be extended to the second order truncated Taylor series. So, we will need this truncated Taylor series when we approximate functions f at different points.

.

(Refer Slide Time: 53:59)

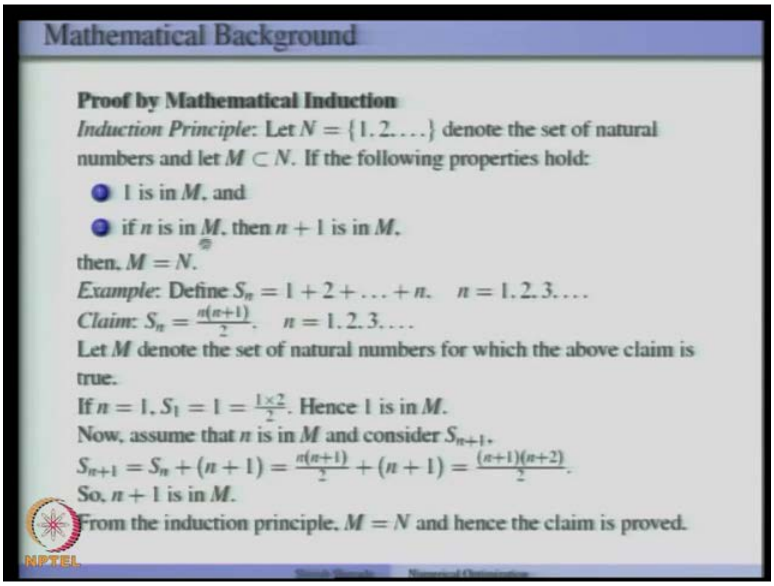


Now, we will look at some important things about the proofs of theorem. So, in this course, we will come across situations where we have to show that given that A is true, B is true. For example, if p is an even number, then p square will be an even number. So, this is the 1 way implication. Now, that can be proved by assuming that A is true and then one proves that B is also true. So, there is other way, which is called the proof by contradiction, where we assume that B is not true and then show that A is not true. So, we will use either of these ways to prove this implication.

Now, when it comes to 2 way implication, where A implies B and B implies A, so which means that A is true if and only if B is true. So, one can think of it has B is a necessary and sufficient condition for a. Now, the important point to be noted here is that we must prove that A implies B and B implies A. A common example of this is that if p is the prime number, then 1 and p are the, if p is the prime number then 1 and p are the only factors of p. It is the other way also that if 1 and p are the only factors of p, then p is a prime number. So, we can write it as p is the prime number, p is the prime number if and only if 1 and p are the only factors of p. So, this is the 2 way implication. So, while proving it, we have to make sure that we prove both A implies B, and B implies A.

.

(Refer Slide Time: 55:53)



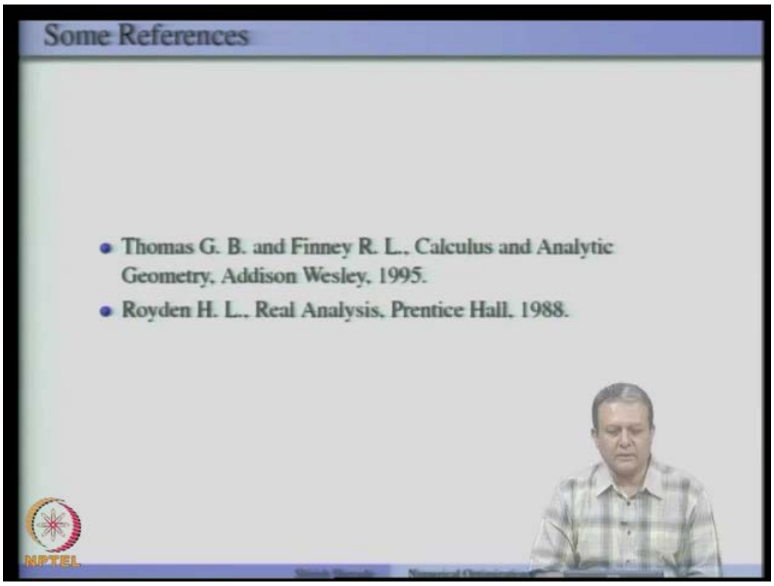
So, here is another interesting concept that will be used sometime in this course. That is called the proof by mathematical induction. So, this proof is based on induction principle. So, let N be a set of natural numbers and M be a sub set of N. Now, if suppose the following properties hold that 1 is in M that means the number 1 is in M, the set M; and if n is in M, then n plus 1 is also in M. Then, the two sets are equal; M equal to N. So, we can conclude that by induction principle, this holds M equal to N.

Now, here is an example. So, let us define a sequence S, define some S n to be some of first n positive integers, so n going from 1 to 3 onwards. Now, you must, must have seen this result earlier that the sum of the n positive integers, the first n positive integers is n into n plus 1 by 2, where n goes from 1 to 3. Now, we want to prove this by principle of mathematical induction. So, let M denote the set of natural numbers for which this claim is true.

Now, if n equal to 1, then S 1 equal to 1 and that is nothing but 1 into 2 by 2. Therefore, it belongs to M. So, the first property is satisfied. Now, assume that n is in M and consider the sum n plus 1. So, S n plus 1 is nothing but you take S n and add n plus 1 to it. That can be written as n plus 1 into n plus 2 by 2. So, this satisfies this property that if n is in M, n plus 1 is also in M. Therefore, from the induction principle, M equal to n. Hence, the claim is proved. So, this is the way to prove the claims using mathematical induction. So, we will need this somewhere during this course.

.

(Refer Slide Time: 58:03)



Now, there are some references for this. So, there is a book by Thomas and Finney on calculus. There is a book by Royden real analysis. So, one can look at this to get some more details about calculus in a real analysis. So, in the next class onwards, we will start looking at one dimensional optimization problem and then convex sets on its functions and so on. So, this mathematical background should be enough for this course. You can always look at the references for more details.

Thank you.

.