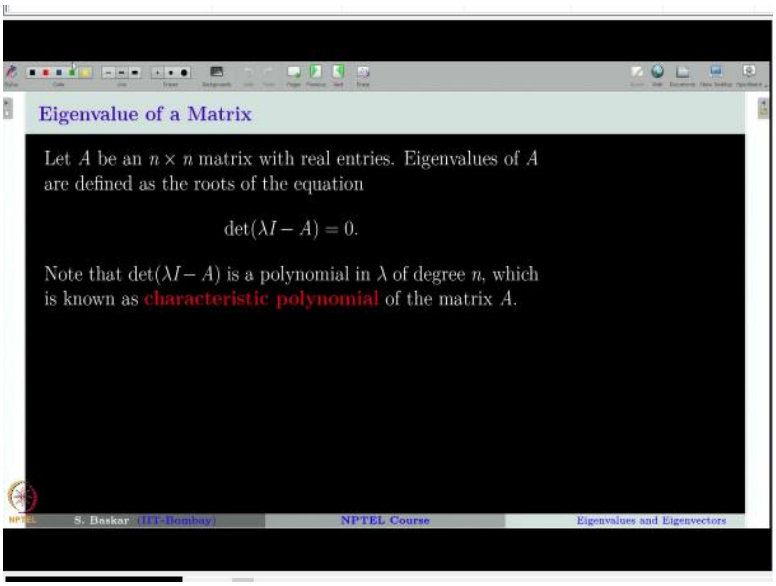
Numerical Analysis
Prof. Dr. S. Baskar
Department of Mathematics
Indian Institute of Technology-Bombay

Lecture-23
Eigenvalues and Eigenvectors: Power Method (Construction)

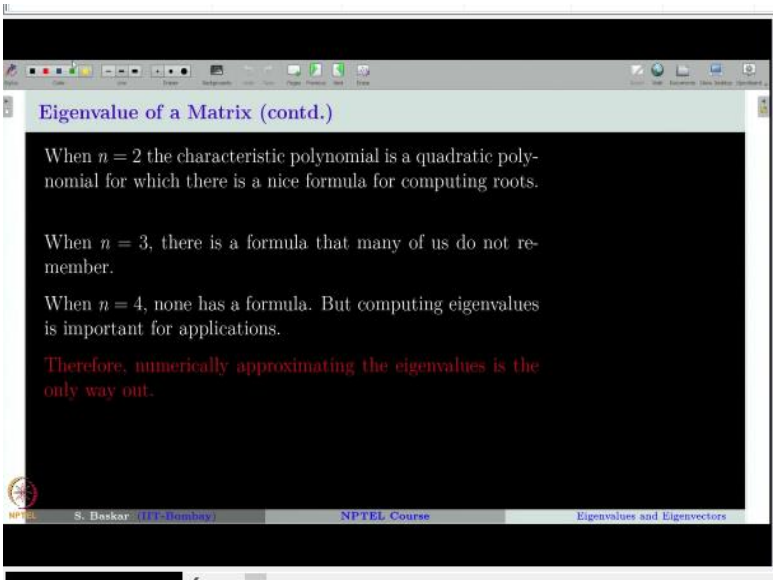
Hi, today we will start a new topic; this is on methods for computing eigenvalues and eigenvectors. In this lecture we will derive a method called power method. Before getting into the method let us quickly recall what is mean by eigenvalues and eigenvectors of a matrix.

(Refer Slide Time: 00:42)



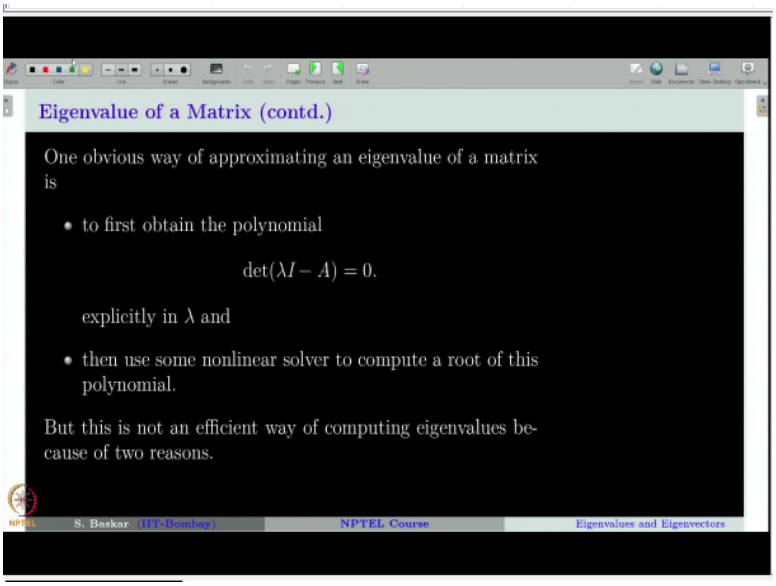
We are given a $n \times n$ matrix of course with real entries. Eigenvalues of A are defined as the root of the equation $det(\lambda I - A) = 0$. Note that $det(\lambda I - A)$ is a polynomial in λ of degree n and it is called the characteristic polynomial of the matrix A.

(Refer Slide Time: 01:14)



When $n = 2$ the characteristic polynomial is a quadratic polynomial and we know how to get the roots of this polynomial. When $n = 3$, well with little difficulties we can still somehow compute the roots of the characteristic polynomial. For $n = 4$ and for $n > 4$ it is rather difficult to compute the exact roots of the polynomial. We do not have any formula to compute the roots when it comes to fourth degree or higher degree polynomials. Therefore, one has to go for some numerical approximation in order to get the eigenvalues of a matrix.

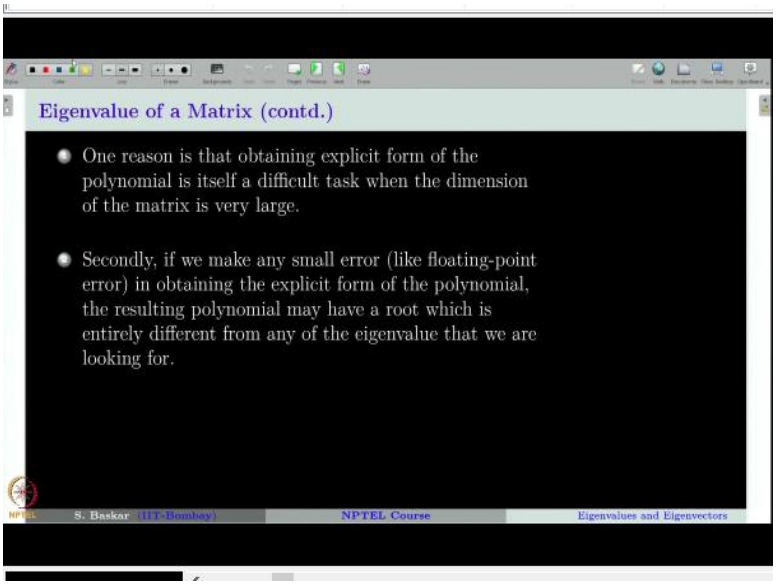
(Refer Slide Time: 02:10)



One obvious way of approximating eigenvalues of matrix is to first write the polynomial explicitly in λ and then use some non-linear solvers which we will be introducing in the next chapter. That is polynomial equations or non-linear equations we have some iteration methods to obtain approximate roots of non-linear equations. We can use one of such methods to

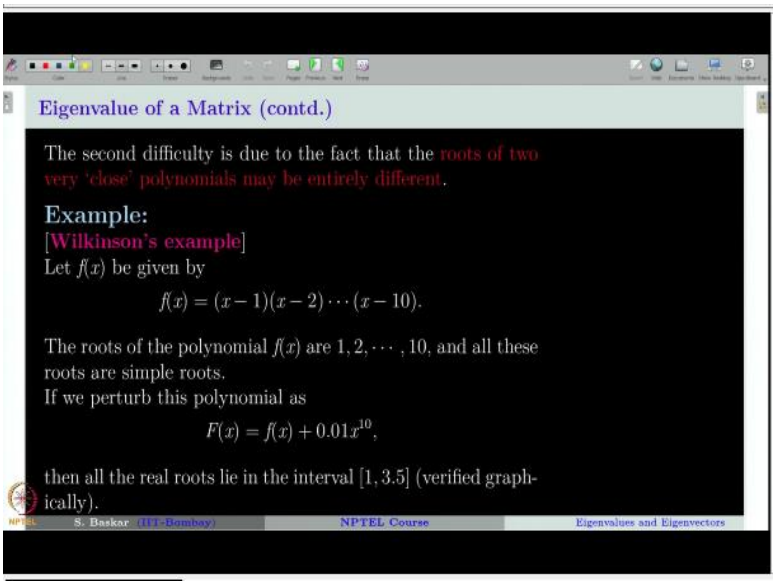
approximate the roots of a polynomial equation also. But this is not so good because of 2 reasons.

(Refer Slide Time: 02:56)



One reason is that obtaining explicit form of the polynomial is itself a difficult task especially when the dimension of the matrix is very large and second thing is polynomials are very sensitive to errors. It means if you make a small error in the coefficients of the polynomials then the corresponding root may be drastically different.

(Refer Slide Time: 03:24)

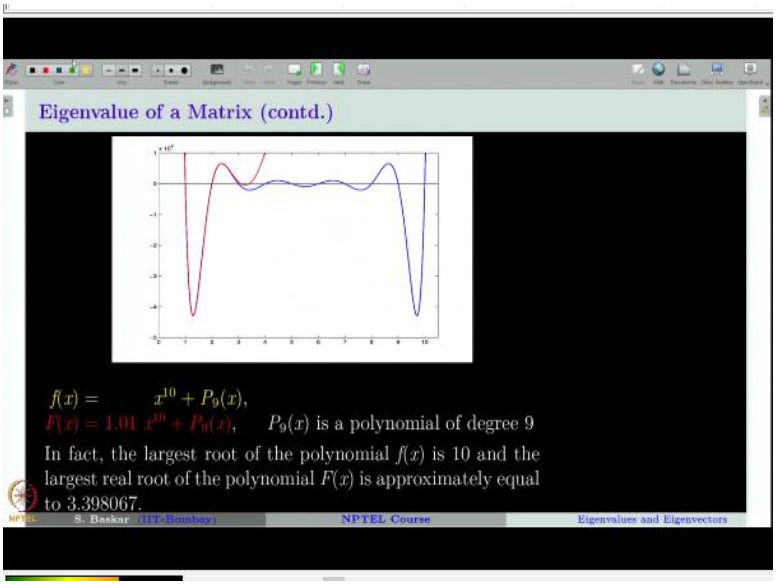


Let us have an example to see how dangerous it is to deal with polynomials when we tend to make errors. Let us consider a function $f(x) = (x-1)(x-2)\cdots(x-10)$ you can see that it is a polynomial of degree 10 with roots as 1, 2, 3 up to 10. Now what we will do is we will

make a slight perturbation in the coefficient of x^{10} terms in the polynomial $f(x)$ and we will call this new perturbed polynomial as $F(x)$.

One can see that the real roots of the polynomial equation $f(x) = 0$ that is $F(x) =$ will lie in the interval 1 to 3.5; well you can verify this graphically.

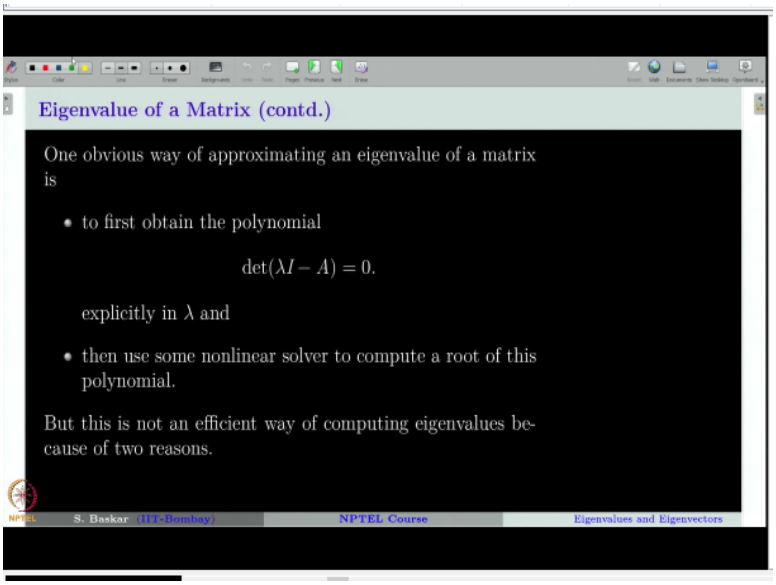
(Refer Slide Time: 04:27)



In this graph we have plotted both the polynomials $f(x)$ as well as $F(x)$, $f(x)$ is plotted in blue solid line, you can see that the point of intersection of the graph with the line $x = 0$ or the roots of this polynomial and similarly the graph of the function $F(x)$ is shown in red solid line. Again, you can see what all the roots of this polynomial equation are at least roughly. And you can see that after around 3.4 the graph of the function $F(x)$ tends to increase and it never again comes back to intersect the x-axis.

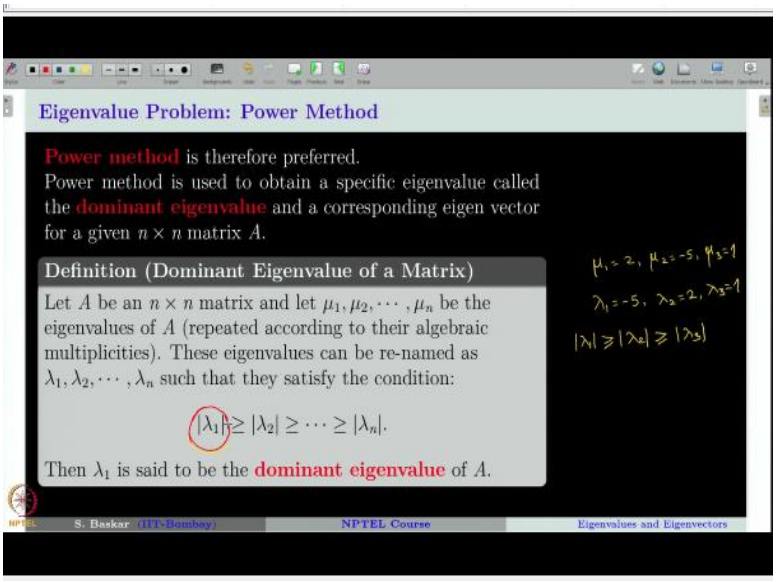
That shows $F(x) = 0$ has roots between 0 to around 3.4, whereas $f(x)$ has roots 1, 2, 3 up to 10. So, what is the difference between $f(x)$ and $F(x)$? There is only one percent error in the coefficient of x^{10}. Otherwise, these 2 polynomials are same. So, a small perturbation therefore in any of the coefficients of a polynomial can make the roots entirely different from the original one.

(Refer Slide Time: 06:15)



Therefore, constructing the polynomial explicitly from the expression $det(\lambda I - A) = 0$ especially on a computer may tend to make some rounding error and that in turn will give you a polynomial in λ which may have entirely different roots than what actually we wanted as eigenvalues of the matrix A. Therefore, it is not a good idea for us to go for solving the polynomial equations in order to get eigenvalues of the matrix at least computationally.

(Refer Slide Time: 06:52)

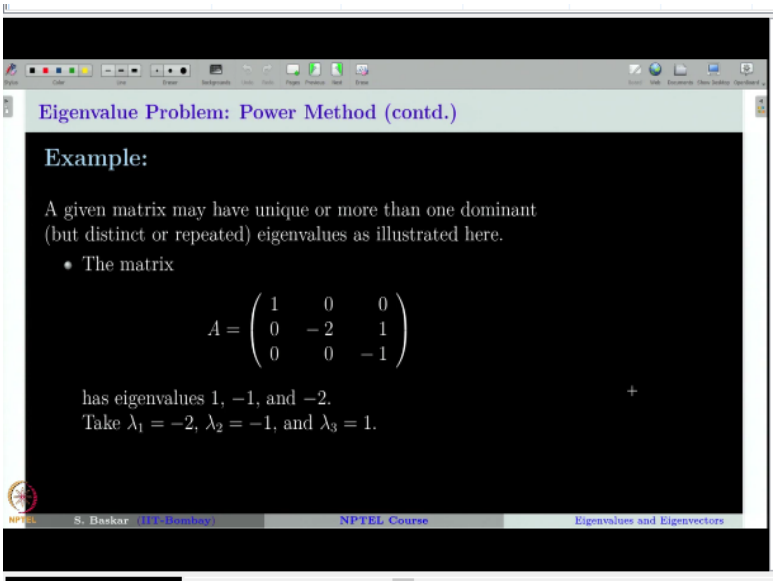


An alternate idea is what we are going to learn in this class and that is the power method. Let us see how to obtain eigenvalues without solving the characteristic polynomial. Power method is actually used to obtain a specific eigenvalue of a given matrix called the dominant eigenvalue and it also approximates corresponding eigenvector of the dominant eigenvalue of a given matrix A. First let us understand what is mean by dominant eigenvalue?

You are given a matrix A which is a $n \times n$ matrix and let $\mu_1, \mu_2, \cdots, \mu_n$ be the eigenvalues of A, they are repeated according to their algebraic multiplicity. Now what you do is first you arrange them in a way that after rearranging the eigenvalues will satisfy this sequence of inequalities. That is suppose you have $\mu_1 = 2$ and $\mu_2 = -5$ and $\mu_3 = 1$ then what you do is you name $\lambda_1 = -5$ and $\lambda_2 = 2$ and $\lambda_3 = 1$.

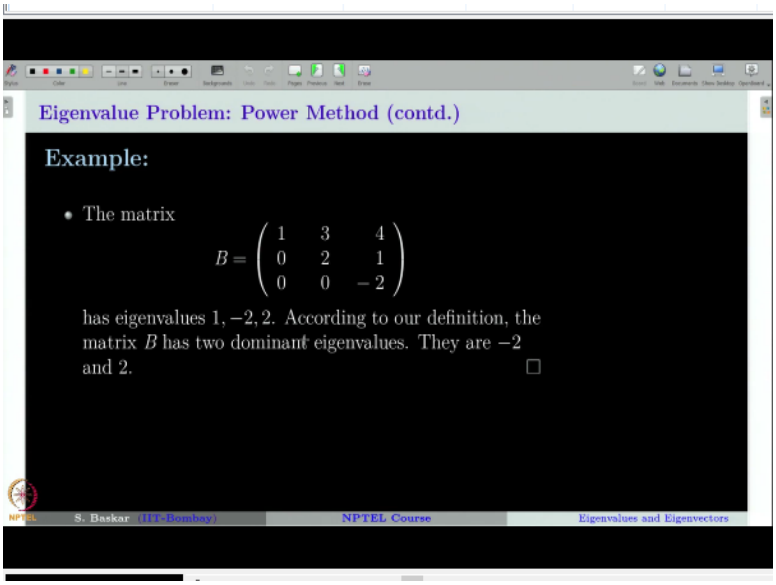
So, that $|\lambda_1| \geq |\lambda_2| \geq |\lambda_3|$. It is just a rearrangement there is nothing new in this and then we call λ_1 as the dominant eigenvalue of the matrix A.

(Refer Slide Time: 08:53)



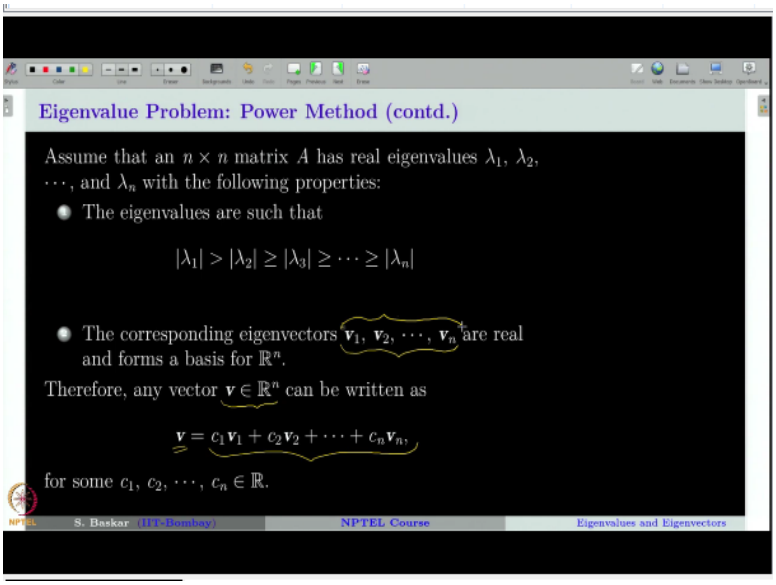
Let us see few examples let us take this matrix A = 1, 0, 0, 0, -2, 1 and 0, 0 -1, you can see that the eigenvalues are 1, -1 and -2. Therefore, you take λ_1 = -2, λ_2 is equal to you can take either -1 or +1, λ_3 = 1 and you can see that λ_1 is the dominant eigenvalue of this matrix and you can also see that the dominant eigenvalue of the matrix A is unique.

(Refer Slide Time: 09:44)



Let us see another example. Here the eigenvalues of the matrix *B* are 1, -2 and 2. You can see that although *B* has distinct eigenvalues, its dominant eigenvalue is not unique because both -2 and 2 will come as dominant eigenvalues.

(Refer Slide Time: 10:11)

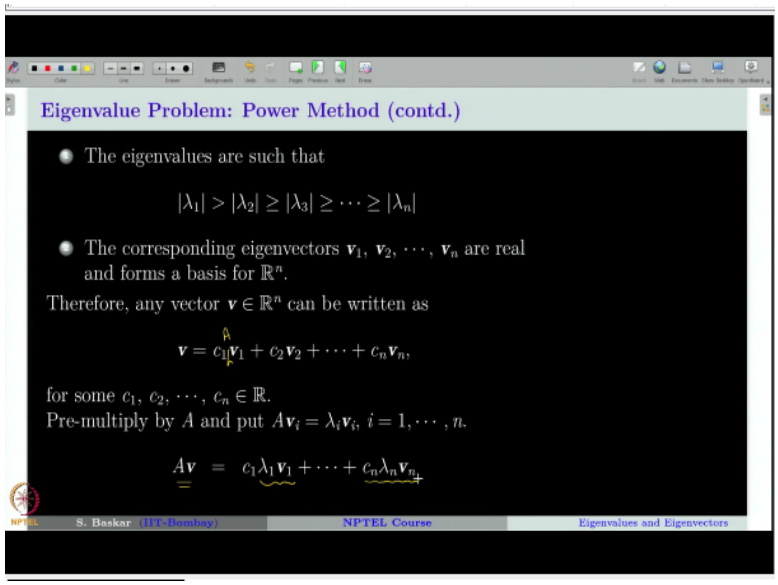


Now let us see the conditions under which we can apply power method. Power method demands that the dominant eigenvalue should be unique; it means $|\lambda_1| > |\lambda_2|$ and then other thing goes as usually. Therefore, the dominant eigenvalue should be unique. If you recall the matrix *A* comes under this condition whereas matrix *B* will not come into this condition; although it has distinct eigenvalues, that is interesting here.

The next condition is that the corresponding eigenvectors which we will denote by $\boldsymbol{v}_1, \boldsymbol{v}_2$ up to $\boldsymbol{v}_n$ are real and forms a basis for $\mathbb{R}^n$. That is all these vectors should be linearly independent.

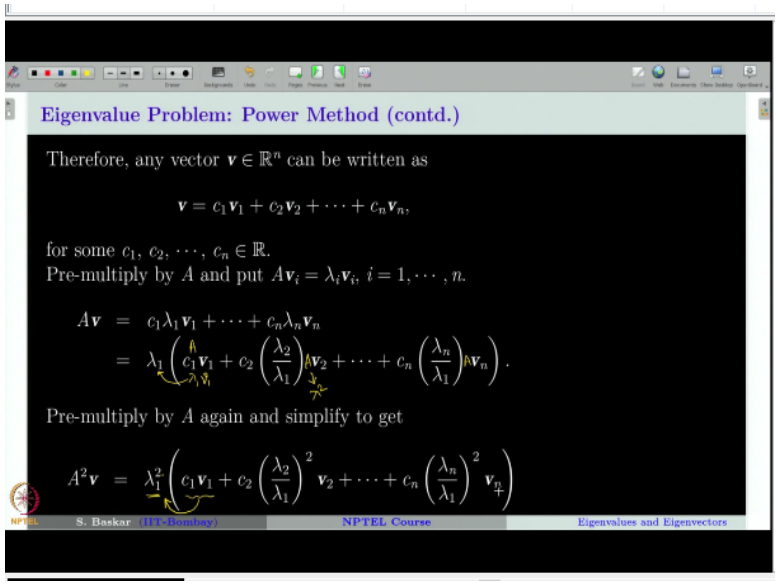
This is also a condition under which you can apply the power method. Now once you assume this condition that is the eigenvectors forms a basis it means what you are given any vector $\boldsymbol{v} \in \mathbb{R}^n$ that vector $\boldsymbol{v}$ can be written as a linear combination of the eigenvectors. That is what we mean by saying that this set of eigenvectors form a basis.

(Refer Slide Time: 11:47)



Now once you have this representation of a given vector then what you do is, you pre multiply this equation by A, which is also done on the hand side and note that all these $\boldsymbol{v}_i$'s are eigenvectors. Therefore, $A\boldsymbol{v_i} = \boldsymbol{\lambda_i v_i}$. So, you use this in the above equation to get $A\boldsymbol{v} = \boldsymbol{c_1 A v_1}$ will become $\lambda\, \boldsymbol{v_1}$ and so on. You have till $c_n\lambda_n\boldsymbol{v_n}$.

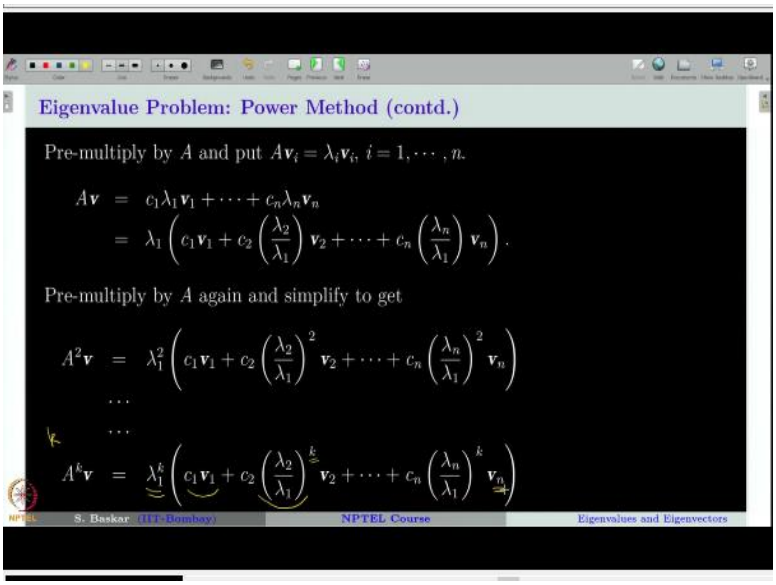
(Refer Slide Time: 12:39)



Now what you do is, you just take this λ_1 outside and write this expression as $\lambda\_1(c_1\boldsymbol{v_1} + \boldsymbol{c_2}\left(\frac{\lambda_2}{\lambda_1}\right)\boldsymbol{v_2} + \cdots + \boldsymbol{c_n}\left(\frac{\lambda_n}{\lambda_1}\right)\boldsymbol{v_n}$. Now what you do is, pre multiply this equation by *A* again that will give you again all this, $A\boldsymbol{v_1}$ will give $\lambda_1\boldsymbol{v_1}$, $A\boldsymbol{v_2}$ will give $\lambda_2\boldsymbol{v_2}$ and so on $A\boldsymbol{v_n}$ will give $\lambda_n\boldsymbol{v_n}$. Again, remove one λ_1 from here and write it as λ_1^2.

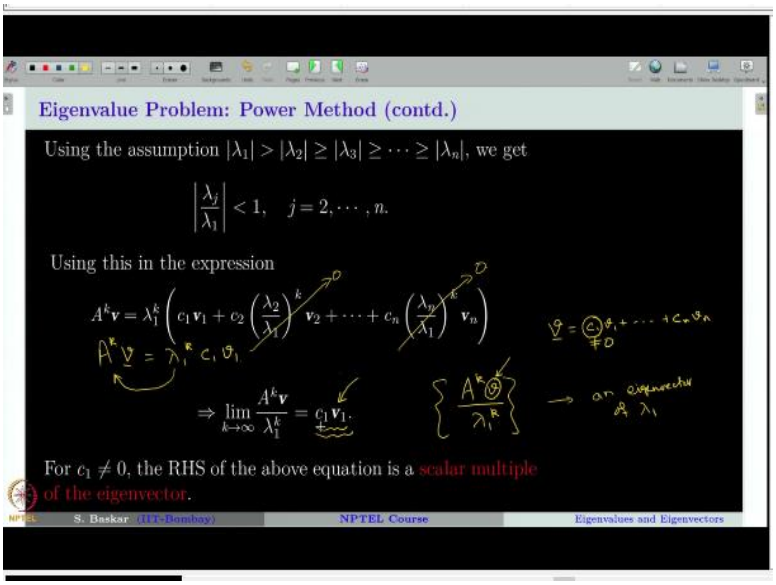
And therefore, you will have the first term as $c_1\boldsymbol{v_1}$ and the second term is c_2. Now already you have λ_2, you are getting one more λ_2 from here and one more λ_1 in the denominator will come because you are pulling one λ_1 outside. Therefore, it is $\left(\frac{\lambda_2}{\lambda_1}\right)^2\boldsymbol{v_2}$+ so on till $c_n\left(\frac{\lambda_n}{\lambda_1}\right)^2$.

(Refer Slide Time: 14:06)



And now you keep on doing this that is you keep on pre multiplying *A* and every time you remove λ_1 outside and write the expression accordingly. At any *k* stage you have $\lambda^k\boldsymbol{v}$ is equal to you would have pulled *k* times λ_1 outside. Therefore, you have λ_1^k and then $c_1\boldsymbol{v_1}$ and at the same time you will also have λ_2^k and since you have pulled λ_1^k out, therefore you have $\left(\frac{\lambda_2}{\lambda_1}\right)^k\boldsymbol{v_2}$ and so on till $c_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\boldsymbol{v_n}$.

(Refer Slide Time: 14:57)



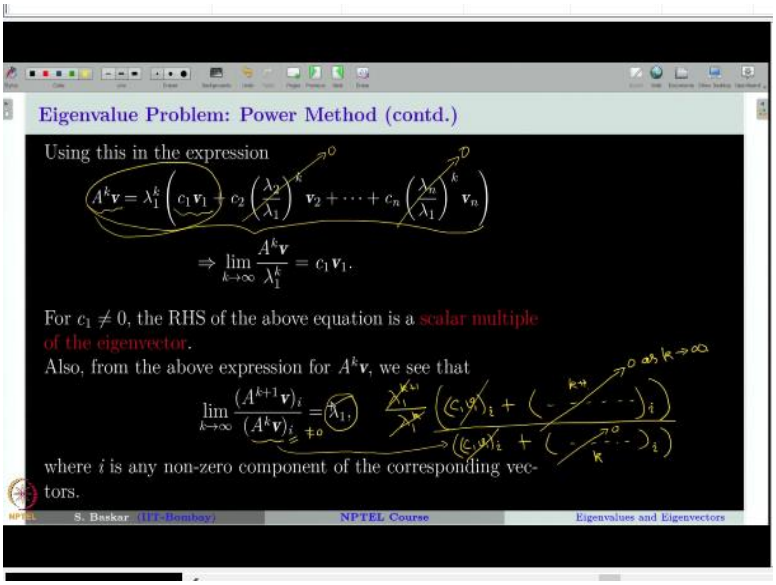
So, we have this expression. Now you recall that we assumed that the dominant eigenvalue of A is unique. Therefore, we have $\frac{|\lambda_j|}{|\lambda_1|} < 1$. Now let us use this condition in the expression that we got previously that is by pre multiplying A, k times to $\boldsymbol{v}$ we obtained this expression. Now you see each of this term is going to be something less than 1 when you take the modulus on both sides.

That shows that as $k \to \infty$ these terms go to 0. So, what we will be left out with is that this is going to 0, this is going to 0 and therefore we are left out with $A^k \boldsymbol{v} = \boldsymbol{\lambda_1^k c_1 v_1}$. Therefore, limit $\lim_{k\to\infty} \frac{A^k \boldsymbol{v}}{\lambda_1^k}$ will be scalar multiple of the eigenvector $\boldsymbol{v_1}$ corresponding to λ_1. That shows that the sequence $\{\frac{A^k \boldsymbol{v}}{\lambda_1^k}\}$ converges to an eigenvector of λ_1. That is what we understand.

Remember this $\boldsymbol{v}$ is any vector in $\mathbb{R}^n$, we have not specified any condition on this. Only thing is you can observe that in order to get an eigenvector for λ_1 you should have this $\boldsymbol{v}$ in such a way that in its representation. That is if you recall we started with $\boldsymbol{v = c_1 v_1 + \cdots + c_n v_n}$. This scalar c_1 should be non zero. Why it is so because otherwise you will just have a 0 vector here which is not an eigenvector of λ_1.

So, in order to have a non-zero vector here we need to have the c_1 to be non zero. That is very important. So, you can see that we have obtained a sequence which may converge to an eigenvector of the dominant eigenvalue λ_1. That is what we have understood from this simple calculation.

(Refer Slide Time: 18:01)



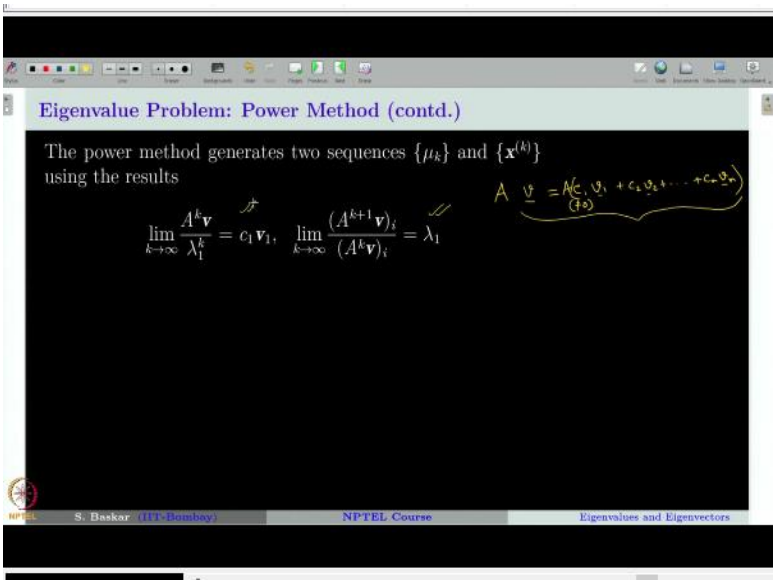
Now let us try to construct a sequence that converges to the dominant eigenvalue λ_1. How are we going to do this? Again look at this expression as *k* tends to infinity all these terms are going to go to 0 and you only have this term. Now what you do is you take one of the components of this vector; remember this is a vector and similarly the right hand side is also a vector.

You just pick up one component say $A^k\boldsymbol{v}$ is a vector, you just take a component say *i*th component which is not equal to 0. Now what you do? The same component you take from $A^{k+1}\boldsymbol{v}$ also. Then you can see that this term is nothing but $\frac{\lambda_1^{k+1}}{\lambda_1^k}(c_1v_1)_i$. This is a vector and its *i*th component plus all the other terms which are finally going to give us a *n*-dimensional vector.

And you take the *i*th component of that vector also, you can see that this entire thing goes to 0 as *k* tends to infinity and therefore you land up with this and this is for $k + 1$ and similarly you have for the denominator also, what is that? It is also c_1v_1. The first term is the same plus it has this vector which is obtained at the *k*th step and its *i*th component. That does not matter, if this will also go to 0 and what remains is going to be equal, therefore they get canceled.

Here also λ_1^k will get canceled with λ_1^k here and you will be left out with just λ_1. So, that is how we are getting the sequence for approximating the dominant eigenvalue of the matrix *A*.

(Refer Slide Time: 20:27)

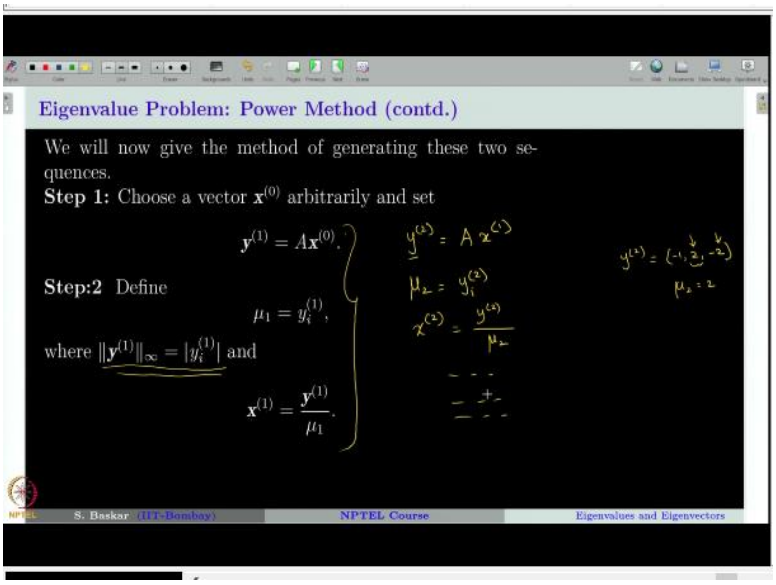


So, what we did so far; we just started with some arbitrary vector $\boldsymbol{v}$, we wrote that vector as $\boldsymbol{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n}$ and of course the first scalar should not be equal to 0. This is just a construction; we are not defining it as a method of course when we are going for a power method it means we do not know what are the eigenvalues and what are the eigenvectors.

That is why we are going for an approximation procedure, but this is theoretically what we have to impose as a condition. However, in practice we will not know this therefore you just have to start with some arbitrary vector and just go ahead with the method. So, once you have this vector chosen then you keep on multiplying *A* with it and every time on the right hand side you pull λ_1 out.

In that way we have constructed 2 sequences; one sequence converges to the dominant eigenvalue and another sequence that we constructed converges to corresponding eigenvector of the dominant eigenvalue if at all these sequences converge. If they converge, they converge to what we actually want. However, the question is will they converge? That is the question.

(Refer Slide Time: 22:11)



Let us write this method in a systematic way. So that, it can be implemented on a computer. Let me explain you the computational procedure first and then we will write the algorithm in a nice systematic way. To compute the dominant eigenvalue and a corresponding eigenvector using power method what we have to do is; first choose an arbitrary vector $\boldsymbol{x}^{(0)}$. Theoretically that $\boldsymbol{x}^{(0)}$ should satisfy certain conditions.

That we will list later; one condition we have already seen that its representation in terms of the eigenvectors should have its first term that is c_1 to be non zero. So, that is something which is expected but practically we do not know. So, therefore practical implementation in that sense is a blind implementation. You just have to take an arbitrary vector and then what you do is you find a vector $\boldsymbol{y}^{(1)}$ which is given by $A\boldsymbol{x}^{(0)}$.

Remember we have to keep on pre multiplying A with the vector that we have chosen. That is the only idea in power method; we are doing the same thing in a rather algorithmic way. So, first choose the vector $\boldsymbol{x}^{(0)}$ and multiply A with $\boldsymbol{x}^{(0)}$, call that as $\boldsymbol{y}^{(1)}$. Then what you do is you define a scalar μ_1 which is nothing but the coordinate of $\boldsymbol{y}$ at which the maximum norm is achieved.

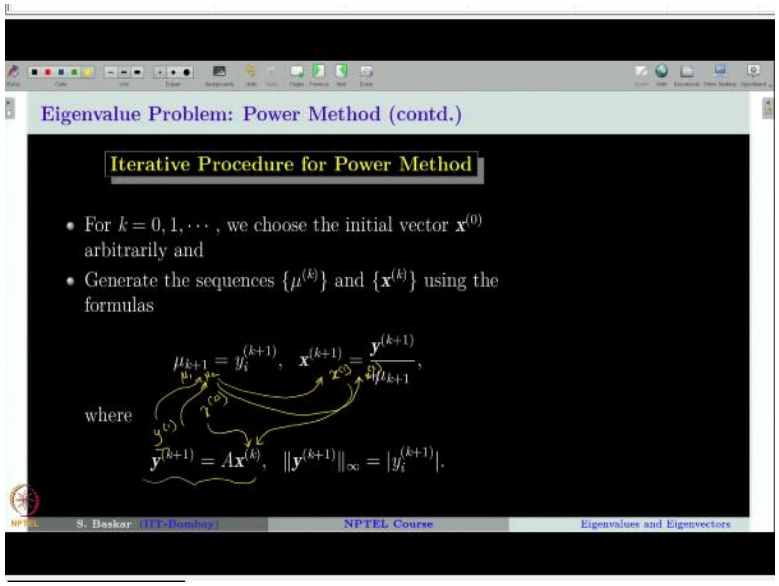
Suppose $\boldsymbol{y}^{(1)}$ is say (-1, 5 and 0) then $\mu_1 = 5$ or suppose $\boldsymbol{y}^{(1)}$ = (2, -5 and 1) then μ_1 is -5 it is not +5, because when you are finding the maximum that will take modulus, but you pick only the index and take the value as μ_1 not the absolute value. This is something which often students

make mistake, they take modulus of $\boldsymbol{y}_i^{(1)}$, you should not take that and now what you do is define a vector $x^{(1)} = \frac{\boldsymbol{y}^{(1)}}{\mu_1}$.

Now this completes one typical iteration. Once you get $\boldsymbol{x}^{(1)}$ now again you go to find $\boldsymbol{y}^{(2)}$ which is equal to $A\,\boldsymbol{x}^{(1)}$. Once you get this then find that coordinate of $\boldsymbol{y}^{(2)}$ at which the maximum norm is achieved. To have a precise choice of this we will take the minimum index at which the maximum is achieved. That is suppose if you have $\boldsymbol{y}^{(2)}$ as say (-1, 2, -2) then the maximum norm is achieved both at the second coordinate as well as at the third coordinate.

So, we will choose the second coordinate. That is, we will take μ_2 as 2, so $\mu_2 = \boldsymbol{y}_i^{(2)}$, where $\boldsymbol{y}_i$ is the coordinate with minimum index at which the maximum norm is achieved. Once you have this you will go for computing $x^{(2)}$. That is nothing but $\frac{\boldsymbol{y}^{(2)}}{\mu_2}$. Like that you will keep on going. Once you get $\boldsymbol{x}^{(2)}$ again you will find $\boldsymbol{y}^{(2)}$, μ_3 and then $\boldsymbol{x}^{(3)}$ and it goes on. In that way you have an iterative sequence.

(Refer Slide Time: 26:33)

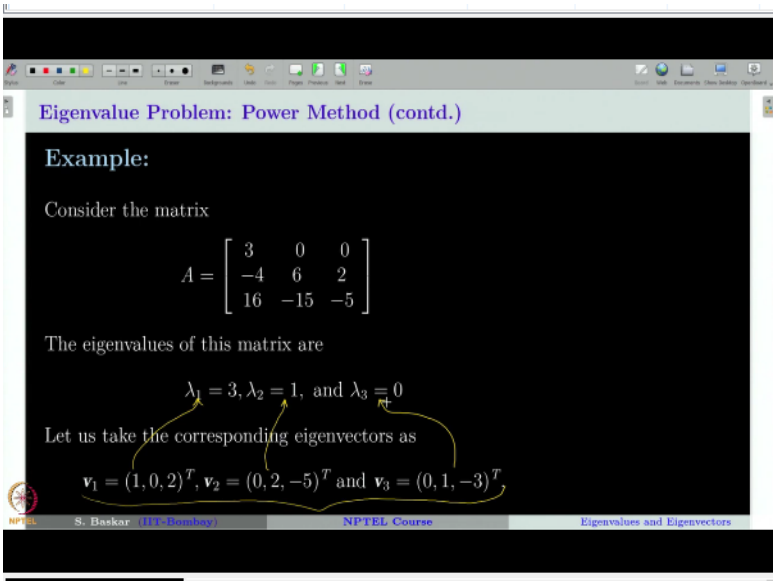


Let us put this iterative sequence in a little better way. So, you are given $\boldsymbol{x}^{(0)}$ which is arbitrarily chosen then for each k we define 2 sequences; one is a sequence of real numbers $\{\mu_k\}$ and another is a sequence of vectors $\{\boldsymbol{x}^{(k)}\}$; they are given by $\mu_{k+1} = y_i^{(k+1)}$ and $\boldsymbol{x}^{(k+1)} = \frac{\boldsymbol{y}^{(k+1)}}{\mu_{k+1}}$. What is $\boldsymbol{y}^{(k+1)}$? $\boldsymbol{y}^{(k+1)}$ is nothing but $A\boldsymbol{x}^{(k)}$. So, you start with $\boldsymbol{x}^{(0)}$, you plug in here; you get $\boldsymbol{y}^{(1)}$

.

And once you have $\boldsymbol{y}^{(1)}$ you go to find μ_1 and once you have μ_1 you go to find $\boldsymbol{x}^{(1)}$ and then again once you have $\boldsymbol{x}^{(1)}$ you again plug in here get $\boldsymbol{y}^{(2)}$; once you have $\boldsymbol{y}^{(2)}$ you plug in here to get μ_2 and once you have μ_2 you plug in here to get $\boldsymbol{x}^{(2)}$ and so on. So, that is a very clear algorithm that we got and this is called the power method. So the outcome of the power method is a pair of sequences. One is a sequence of real number and another one is a sequence of vectors.

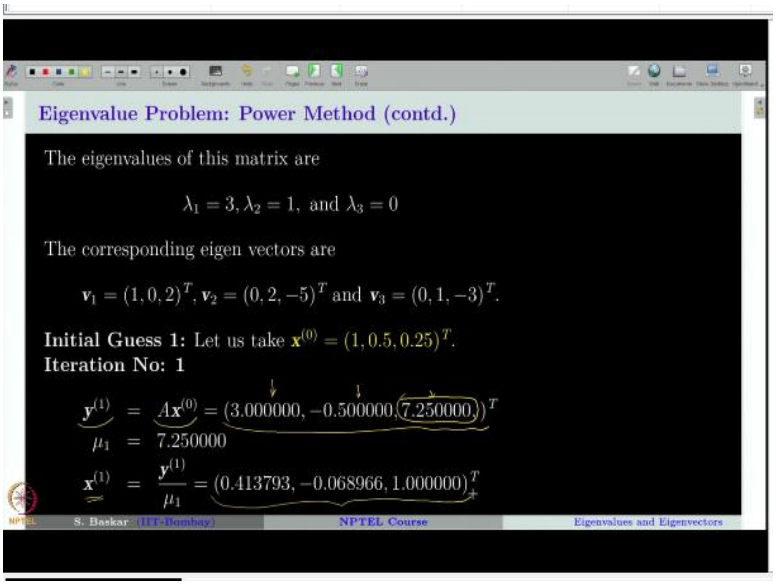
(Refer Slide Time: 28:07)



Let us take an example. Let us consider this matrix *A* just for the information the eigenvalues of *A* are given like this. You can see that *A* has a dominant eigenvalue 3 which is the unique dominant eigenvalue for *A*. For information we will also see what are the eigenvectors that we are considering here. The eigenvectors are $\boldsymbol{v}_1$ = (1, 0, 2) which corresponds to λ_1, $\boldsymbol{v}_2$ corresponds to λ_2 and $\boldsymbol{v}_3$ is taken like this which is an eigenvector of λ_3.

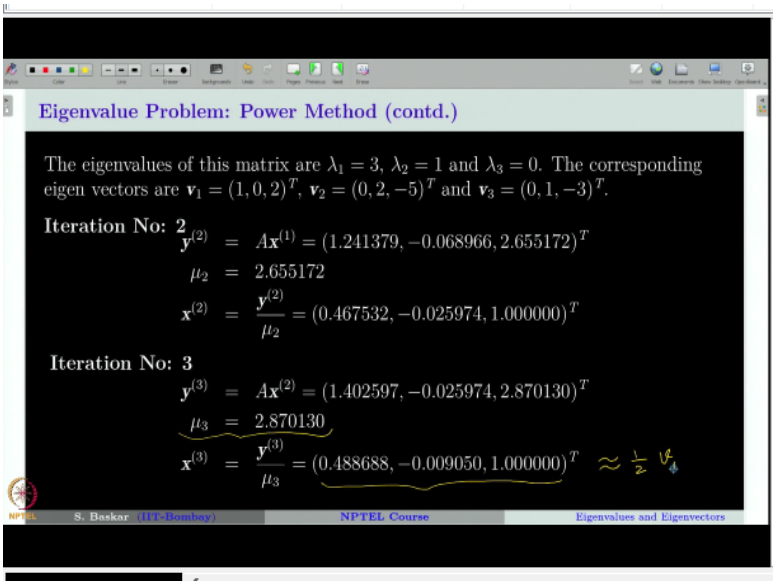
Now this is just for the information, we do not need this information in order to work with the power method, it is just for our information, whereas to run the power method we just need this initial guess.

(Refer Slide Time: 29:05)



Remember as far as the implementation is concerned, we can simply take this $\boldsymbol{x}^{(0)}$ as a arbitrary vector. Once you have this you will calculate $\boldsymbol{y}^{(1)}$ which is given by $A\boldsymbol{x}^{(0)}$ and that happens to be this vector. Now from here you will find the maximum of the absolute values of this coordinates and that is achieved in the third coordinate in this vector and therefore μ_1 is taken as 7.25. Once you have μ_1 you will compute $\boldsymbol{x}^{(1)}$ which is $\boldsymbol{y}^{(1)}$ divided by μ_1 and that is given by this vector.

(Refer Slide Time: 29:59)

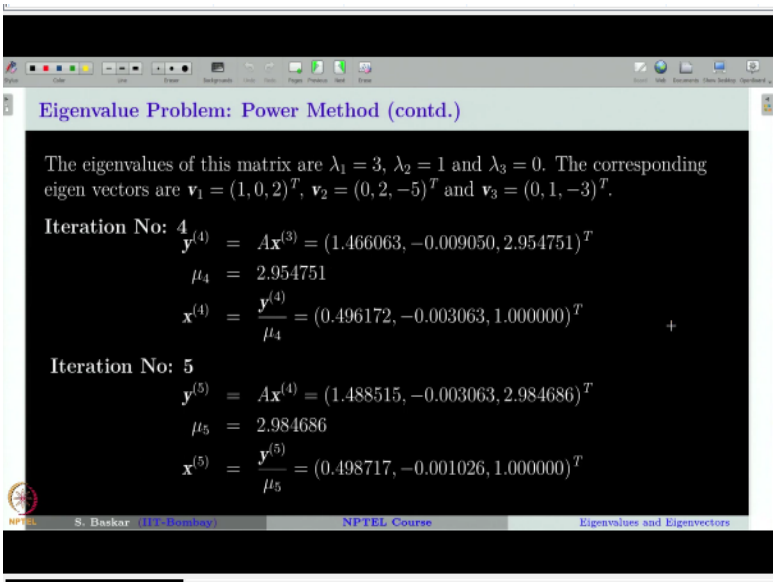


Now let us go to compute the second iteration. Here you will take the $\boldsymbol{x}^{(1)}$ which was computed in iteration number 1, plug in here and you will get $\boldsymbol{y}^{(2)}$ as $A\boldsymbol{x}^{(1)}$. Again, you see which coordinate is achieving the maximum norm that is coming again in the third coordinate.

Therefore μ_2 is taken as this value and then $\boldsymbol{x}^{(2)}$. You just observe how the $\{\mu\}$ is going and also observe how the $\{\boldsymbol{x}^{(k)}\}$ is going on.

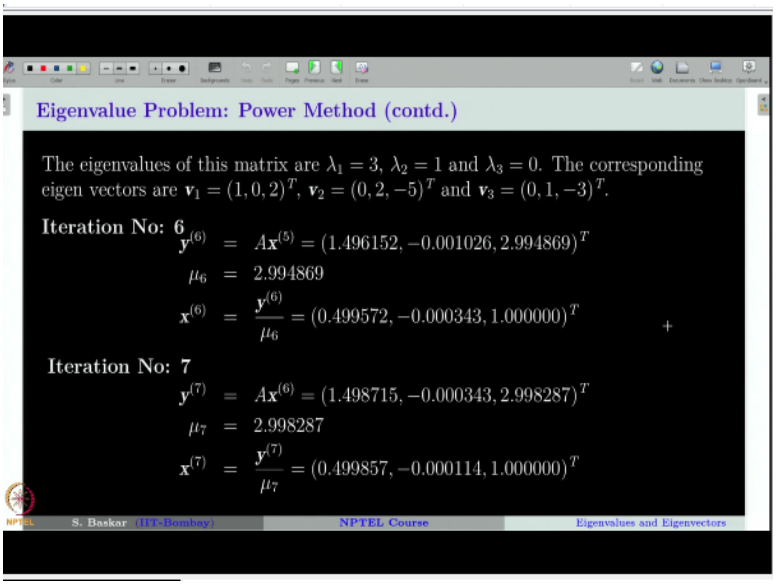
You can see that in the first iteration μ was 7.25, in the second iteration it jumped actually pretty close to 3, we are expecting μ_2 converges to λ_1 and $\{\boldsymbol{x}\}$ is expected to converge to a scalar multiple of $\boldsymbol{v}_1$. Let us see how the next iteration goes. The next iteration gives us μ_3 equal to its little more closer to 3 therefore μ is going very well, $\boldsymbol{x}^{(3)}$ is coming like this. You can observe that $\boldsymbol{x}^{(3)}$ is actually going closer and closer to $\frac{1}{2}\boldsymbol{v}_1$.

(Refer Slide Time: 31:31)



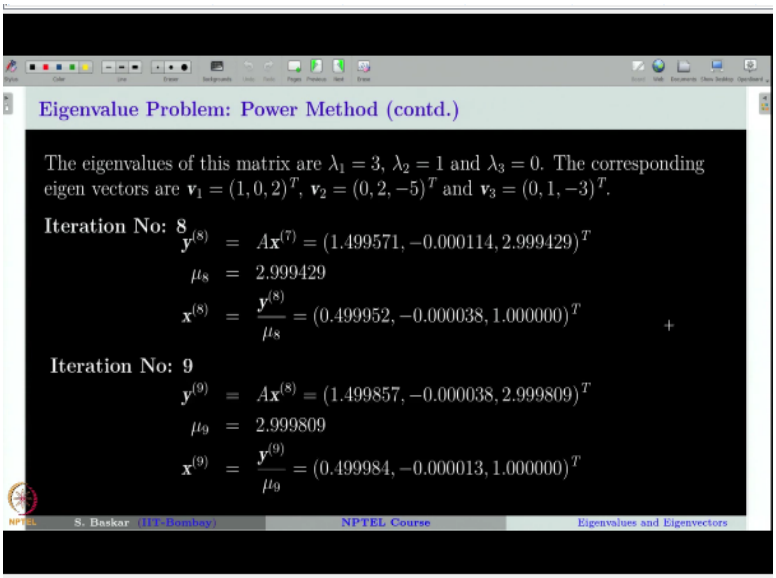
Let us go ahead $\boldsymbol{x}^{(4)}$ pretty close to 3 and this is also pretty close to $\frac{1}{2}\boldsymbol{v}_1$.

(Refer Slide Time: 31:41)

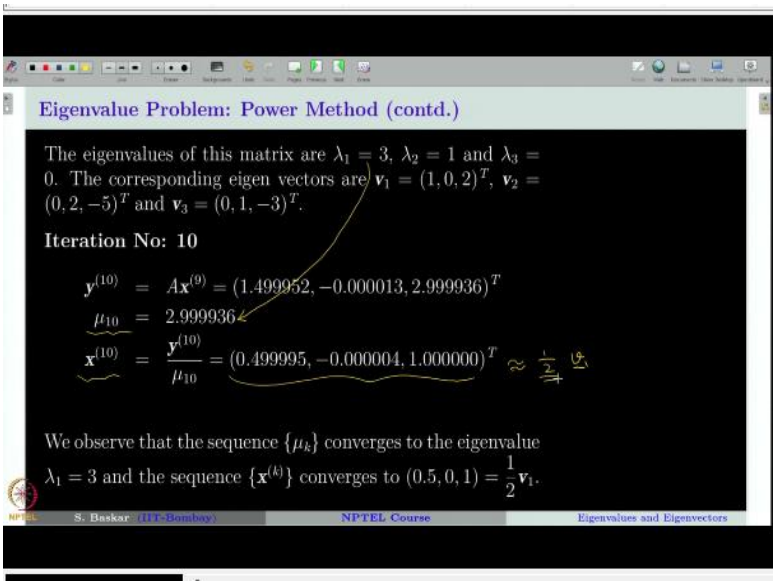


$\boldsymbol{x}^{(5)}$, similarly you can go on computing $\boldsymbol{x}^{(6)}, \boldsymbol{x}^{(7)}, \boldsymbol{x}^{(8)}, \boldsymbol{x}^{(9)}, \boldsymbol{x}^{(10)}$ and so on.

(Refer Slide Time: 31:48)



(Refer Slide Time: 31:49)



You can keep on going, but I have stopped my computation at the iteration number 10. You can see that your μ is pretty close to the dominant eigenvalue of the matrix A and $\boldsymbol{x}^{(10)}$ is pretty close to $\frac{1}{2}\boldsymbol{v}_1$. So, that is what we could observe here. Now the question is when does the power method converge and if it is converging will we know that the {μ} will converge to the dominant eigenvalue.

And the $\{\boldsymbol{x}^{(k)}\}$ will converge to a scalar multiple of $\boldsymbol{v}_1$, we can also see what is that scalar multiple. Here it happens to be $\frac{1}{2}$; you can also understand what is that scalar multiple and so

on. For that we have to understand the convergence theorem of power method which we will do in the next class. Thank you for your attention.