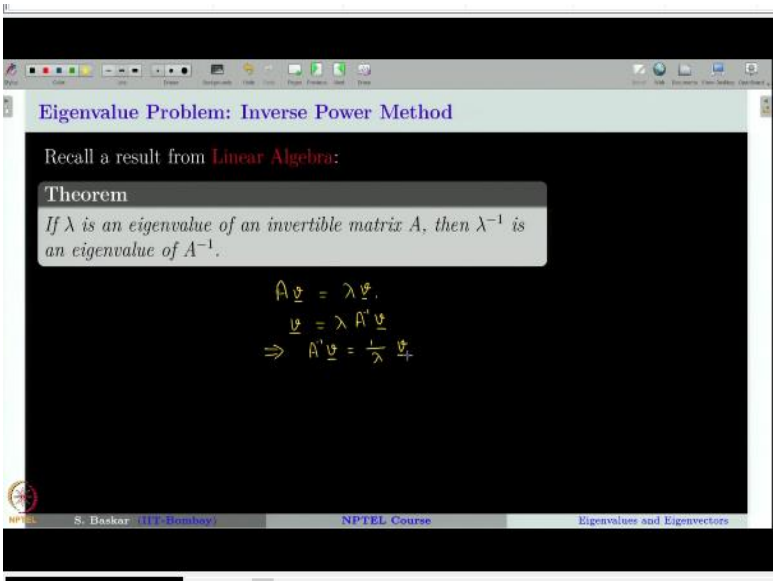
Numerical Analysis
Prof. Dr. S. Baskar
Department of Mathematics
Indian Institute of Technology-Bombay

Lecture-26
Eigenvalues and Eigenvectors: Power Method (Inverse and Shifted Methods)

Hi, we are studying numerical methods for computing eigenvalues and eigenvectors of a given matrix. In this we have introduced power method in the last class and also we learned convergence theorem for power method. In this class we will study a small variant of power method called inverse power method.

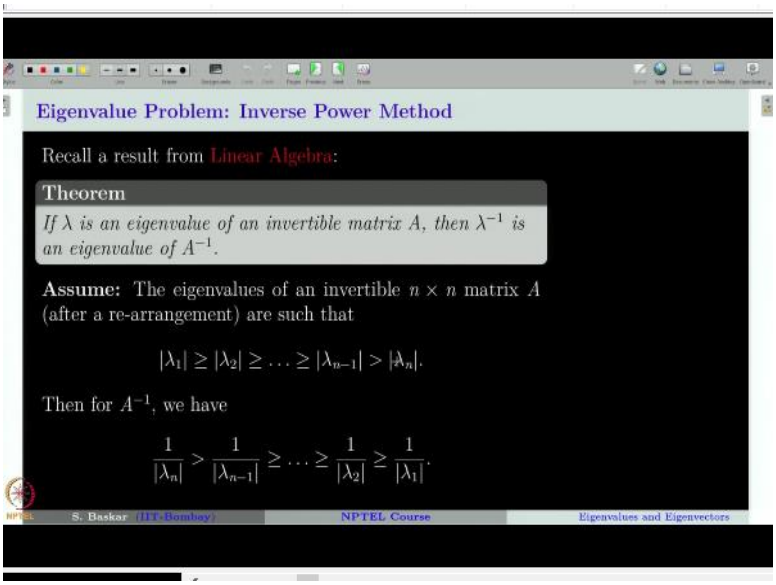
(Refer Slide Time: 00:48)



Let us first recall a result from linear algebra which states that if λ is an eigenvalue of an invertible matrix A then λ^{-1} is an eigenvalue of A^{-1}. This is not very difficult for us to see; you just take λ as the eigenvalue of a matrix A. Then you can write $A\,\boldsymbol{v} = \lambda\,\boldsymbol{v}$, where $\boldsymbol{v}$ is an eigenvector corresponding to the eigenvalue λ. Since A is invertible, you can write $\boldsymbol{v} = \lambda A^{-1}\boldsymbol{v}$.

That implies $A^{-1}\boldsymbol{v} = \frac{\mathbf{1}}{\lambda}\boldsymbol{v}$, here you can note that there is no change in the eigenvector; only the eigenvalue becomes the reciprocal of the eigenvalue of the matrix A. So, that is very important for us to note.

(Refer Slide Time: 02:06)

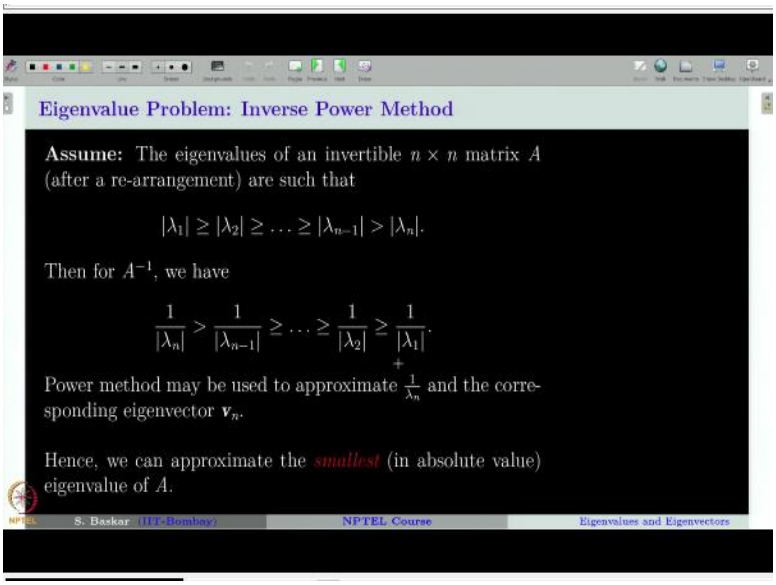


Using this theorem, we can in fact compute the smallest eigenvalue of a matrix *A*. When I say smallest eigenvalue it should be smallest in the absolute value. Say for instance, suppose we have a matrix *A* whose eigenvalues are rearranged such that $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{n-1}| > |\lambda_n|$.

Now you can see that if *A* has such eigenvalues then λ_n which is the smallest eigenvalue is unique. That is the only eigenvalue which is smallest in this sense. Now let us see how to compute this eigenvalue λ_n using power method. For that what we will do is instead of applying power method to *A* which will capture the dominant eigenvalue. Now we will apply the power method to A^{-1}.

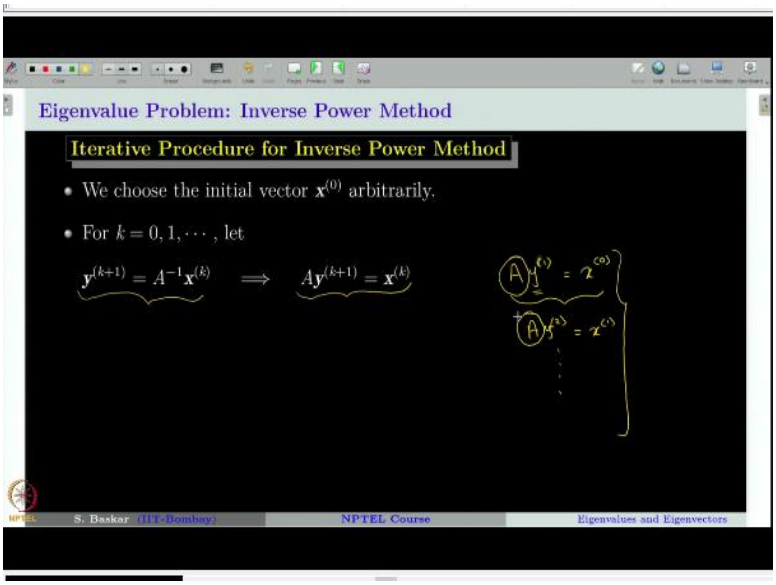
Why we do that? We have seen from the above theorem that if λ_n is the smallest eigenvalue of *A* then $\frac{1}{\lambda_n}$ will be the dominant eigenvalue of A^{-1}. That makes us to apply power method on A^{-1} in order to compute $\frac{1}{\lambda_n}$ and therefore we can obtain this smallest eigenvalue of *A*. That is the idea.

(Refer Slide Time: 03:52)



Remember, once we compute $\frac{1}{\lambda_n}$ the power method applied on A^{-1} will also capture and eigenvector corresponding to $\frac{1}{\lambda_n}$. That is the same as the eigenvector corresponding to λ_n also. So, in that way we can obtain the smallest eigenvalue of *A* and a corresponding eigenvector of the eigenvalue λ_n. So, that is the interesting part of power method. And this way of computing the smallest eigenvalue of a matrix is what we call as the inverse power method.

(Refer Slide Time: 04:36)



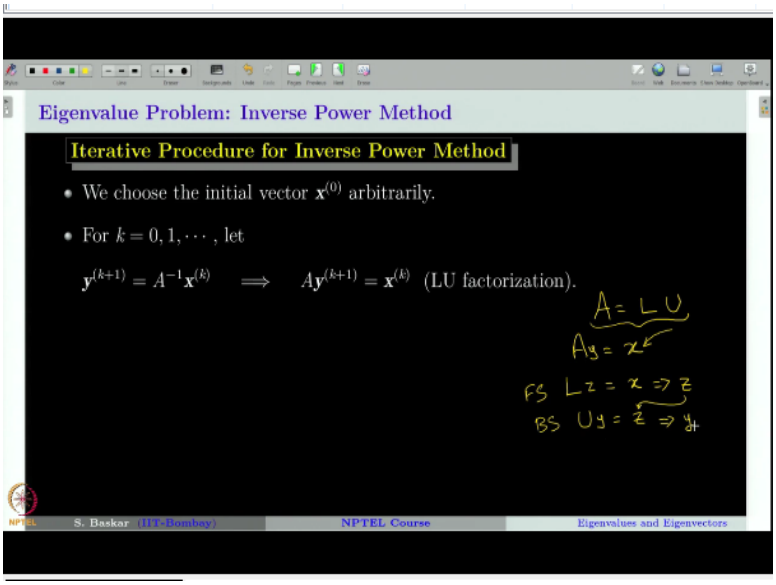
Now, let us give the iterative procedure for inverse power method, you can see that there is nothing much difference in inverse power method when compared to power method. The only difference is instead of applying the power method to the matrix *A* now we will be applying it to A^{-1}. That is for every iteration we will compute $\boldsymbol{y}^{(k+1)} = A^{-1}\boldsymbol{x}^{(k)}$. Remember, in power method we will define $\boldsymbol{y}^{(k+1)} = A\boldsymbol{x}^{(k)}$.

But here we will apply A^{-1} to get $\boldsymbol{y}^{(k+1)}$. This is not something which is straight forward for us because we are given the matrix A. Now in order to compute $\boldsymbol{y}^{(k+1)}$ we need to invert the matrix A which may be computationally costly. So, what we can do is you can write this equation as $A\,\boldsymbol{y}^{(k+1)} = \boldsymbol{x}^{(k)}$. Here you can observe that once you choose $\boldsymbol{x}^{(0)}$ you will plug in $\boldsymbol{x}^{(0)}$ on the right hand side which is a known vector.

And you will then solve this linear system to obtain the unknown $\boldsymbol{y}^{(k+1)}$. That is a major difference in inverse power method when compared to the power method. This complicated step is not there in power method because $\boldsymbol{y}^{(k+1)}$ is just defined as $A\boldsymbol{x}^{(k)}$, $\boldsymbol{x}^{(k)}$ is known to us and A is also given to us. Therefore, $\boldsymbol{y}^{(k+1)}$ can be explicitly computed from there, whereas in inverse power method $\boldsymbol{y}^{(k+1)}$ has to be computed by solving this linear system.

That is, we have to solve this linear system $\boldsymbol{y}^{(1)}$, once you have this then you will have $\boldsymbol{x}^{(1)}$ and that has to be then taken as the right hand side to get $\boldsymbol{y}^{(2)}$ and so on. So, if you see here the right hand side vectors will only change, whereas the coefficient matrix A will remain unchanged.

(Refer Slide Time: 07:13)

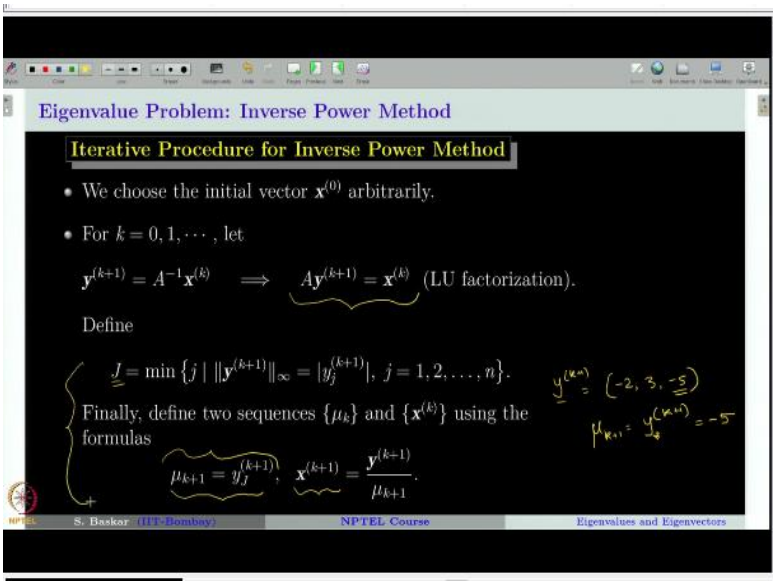


Therefore, what you can do is you go for a LU-factorization because once you do the LU-factorization $A = LU$ then the system $A\boldsymbol{y} = \boldsymbol{x}$ can be solved by first doing a forward substitution $L\boldsymbol{z} = \boldsymbol{x}$ and that gives you $\boldsymbol{z}$. Then we can plug in that into $U\boldsymbol{y} = \boldsymbol{z}$. So, you bring

that to the right hand side and that gives us the required vector $\boldsymbol{y}$. We have learned it in one of our previous classes where we have discussed LU-factorization; you just recall that.

So, the advantage of going for LU-factorization is that once you make the factorization L into U for the matrix *A* then at every iteration you just have to plug in the right hand side vector and do one forward substitution and one backward substitution you will get your $\boldsymbol{y}$. That will be computationally relatively cheaper. So, once you get this $\boldsymbol{y}^{(k+1)}$ rest of the algorithm remains the same as we did in the power method.

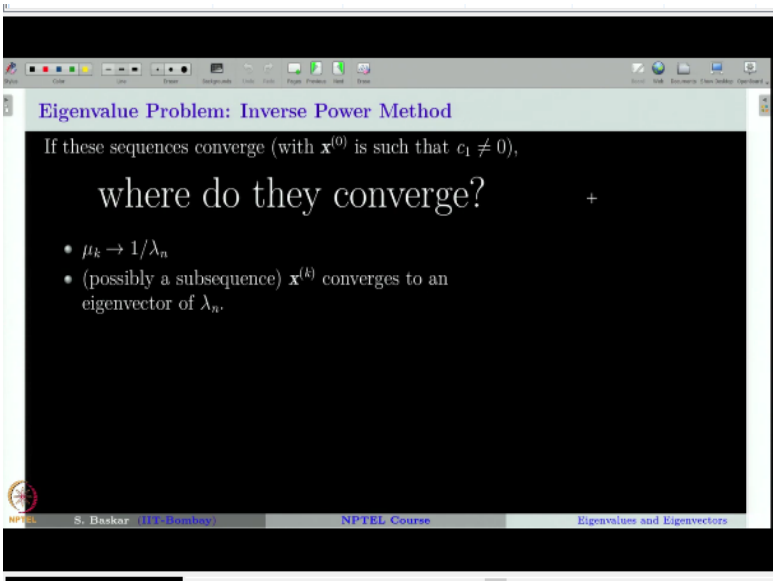
(Refer Slide Time: 08:32)



What you have to do? You take this vector $\boldsymbol{y}$ and find the maximum norm of that and in whichever index the maximum norm is achieved you take that value of the $\boldsymbol{y}$ coordinate and define that as μ_{k+1}. For instance, if $\boldsymbol{y}^{(k+1)}$ is equal to say (-2, 3, -5) then your J which is nothing but the minimum of all the J's at which the maximum norm is achieved. So, here it is achieved only at the third coordinate. Therefore, μ_{k+1} will be $\boldsymbol{y}_3^{(k+1)}$. In this particular example it is -5.

So, this is just an example. So, in that way only you have to choose μ_{k+1}. Once you have μ_{k+1}, you can find the $\boldsymbol{x}^{(k+1)}$very easily by taking $\boldsymbol{y}^{(k+1)}$ and divided by μ_{k+1} which will be a unit vector. So, we have seen all this when we were discussing power method and this is the iterative procedure for the inverse power method. The only difference and the complication is, in computing $\boldsymbol{y}^{(k+1)}$. Otherwise, all these steps remain the same as in the power method.

(Refer Slide Time: 10:14)

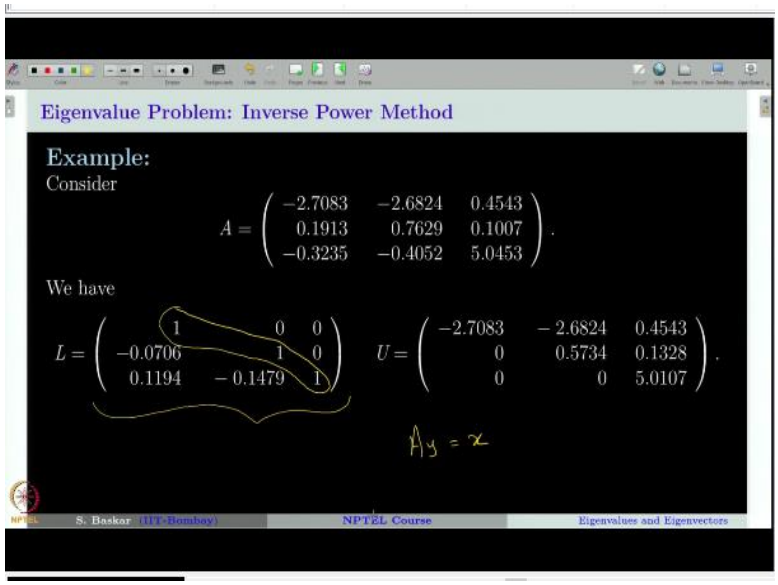


The next question is, if these sequences that is sequence μ_k and sequence $\boldsymbol{x}^{(k)}$; if both these sequences converge then where do they converge? That is the question. In a sense we have already answered this question, but we will make it more precise. Of course, we have to keep in mind that the initial guess $\boldsymbol{x}^{(0)}$ should be chosen in such a way that $c_1 \neq 0$. What is c_1? c_1 is nothing but the coefficient in the representation of $\boldsymbol{x}^{(0)}$ in terms of the eigenvectors.

$c_1 v_1 + \boldsymbol{c_2 v_2} + \cdots + c_n \boldsymbol{v}_n$. In that the coefficient c_1 should be non zero that is for the theoretical reason, but assume that everything went well with our choice of $\boldsymbol{x}^{(0)}$. Then our question is, where these two sequences will converge? The sequence μ_k will converge to the dominant eigenvalue of A^{-1} which is nothing but the reciprocal of the smallest eigenvalue of A.

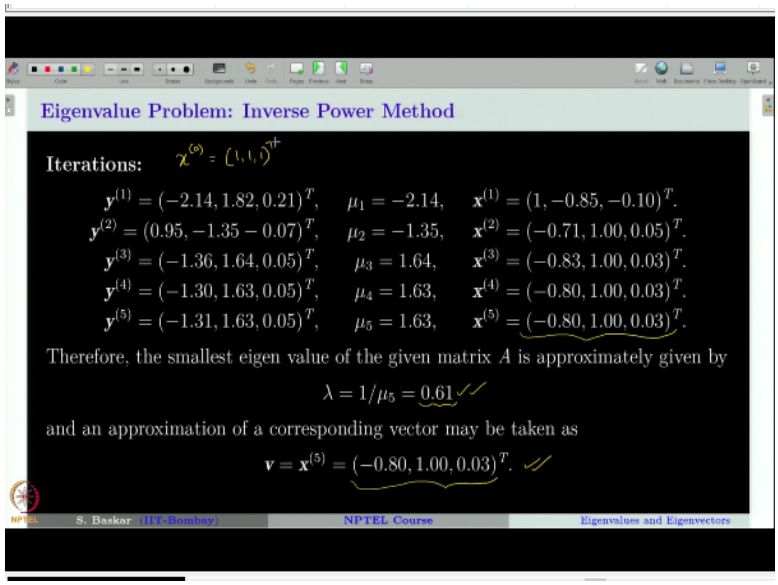
In our rearranged form it is λ_n; therefore μ_k will converge to $\frac{1}{\lambda_n}$. Then where does the sequence $\boldsymbol{x}^{(k)}$ converge? $\boldsymbol{x}^{(k)}$ will converge to an eigenvector of λ_n. Remember it will converge to an eigenvector of $\frac{1}{\lambda_n}$ which is of course the same as the eigenvector of λ_n also.

(Refer Slide Time: 12:00)



Let us see an example. Let us take the matrix *A* given like this, I am just giving you only 4 digits after the decimal place. As I told you, we prefer to first decompose the matrix *A* into L into U. I have just taken the Doolittle factorization you can just observe it here all the diagonal elements are 1. So, I have just taken the Doolittle factorization for *A*. So, once you have L and U then $A\boldsymbol{y} = \boldsymbol{x}$ can be done with a forward substitution and one backward substitution.

(Refer Slide Time: 12:54)



And at each iteration you can get the vector $\boldsymbol{y}$. Once you get the vector $\boldsymbol{y}$ then you will look for that coordinate of $\boldsymbol{y}$ at which the maximum norm is achieved. If it is achieved at multiple coordinates then you will take the one which is with the minimum coordinate. So, in that sense the maximum norm is achieved at the first coordinate of $\boldsymbol{y}^{(1)}$ and therefore μ_1 is given by -2.14.

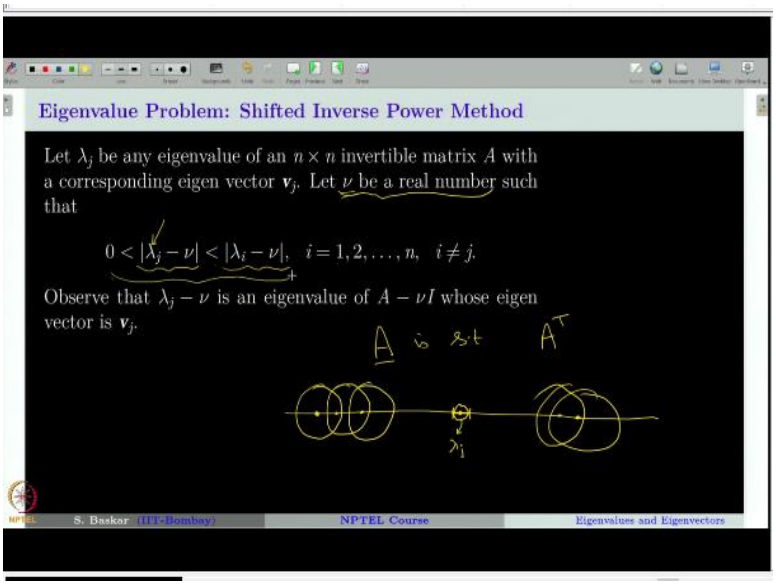
Remember it should be minus not plus, and once you get that then simply take $\boldsymbol{x}^{(1)} = \frac{\boldsymbol{y}^{(1)}}{\mu_1}$. So, that is your $\boldsymbol{x}^{(1)}$ and that is given by this and once you have $\boldsymbol{x}^{(1)}$ again you plug in $\boldsymbol{x}^{(1)}$ into $L\,U\,\boldsymbol{y}^{(1)} = \boldsymbol{x}^{(1)}$. A forward substitution and a backward substitution will give you $\boldsymbol{y}^{(2)}$, so you will get $\boldsymbol{y}^{(2)}$. From $\boldsymbol{y}^{(2)}$ again you will find the coordinate at which the maximum norm is achieved that is μ_2.

And then you divide $\boldsymbol{y}^{(2)}$ by μ_2 you will get $\boldsymbol{x}^{(2)}$. Now you repeat this process, I am just going ahead with the iteration; you can observe that μ is converging slowly to the number 1.63. Remember if μ is converging to 1.63 it means what it is actually converging to $\frac{1}{\lambda_3}$. What is λ_3? We have not given what are all the eigenvalues of *A*; therefore we do not know that is the smallest eigenvalue of *A*.

When I say smallest, it is the smallest value of all the eigenvalues in the absolute sense. Therefore, this is for us given as 1.63, when you take the reciprocal of that that gives you 0.61 and what is the corresponding eigenvector? For eigenvector you do not need to do anything, just comes as it is the eigenvector is taken as -0.81 and 0.03. So, this is the smallest eigenvalue of the matrix *A*.

And the corresponding eigenvector of this eigenvalue is given like this and this is how the inverse power method will go on. For this we have chosen the initial guess as $\boldsymbol{x}^{(0)} = (1, 1, 1)$, I am sorry I forgot to mention this.

(Refer Slide Time: 16:03)



The next question is, we learned how to capture the dominant eigenvalue of a matrix and the smallest eigenvalue of a matrix. Now is there a way to find some of the other eigenvalues of *A* also? That is the question. This can be done by shifting the matrix *A* appropriately and that results into shifted inverse power method. Let us try to understand what is shifted inverse power method?

Assume that we are interested in capturing an eigenvalue λ_j and one of its corresponding eigenvectors say $\boldsymbol{v}_j$. Now what you do is you choose some real number ν such that $|\lambda_j - \nu| < |\lambda_i - \nu|, i = 1,2, \cdots, n, i \neq j$. It means you try to find a nu such that when you take this quantity that will be the smallest value of all such quantities when applied to all other eigenvalues of the matrix A.

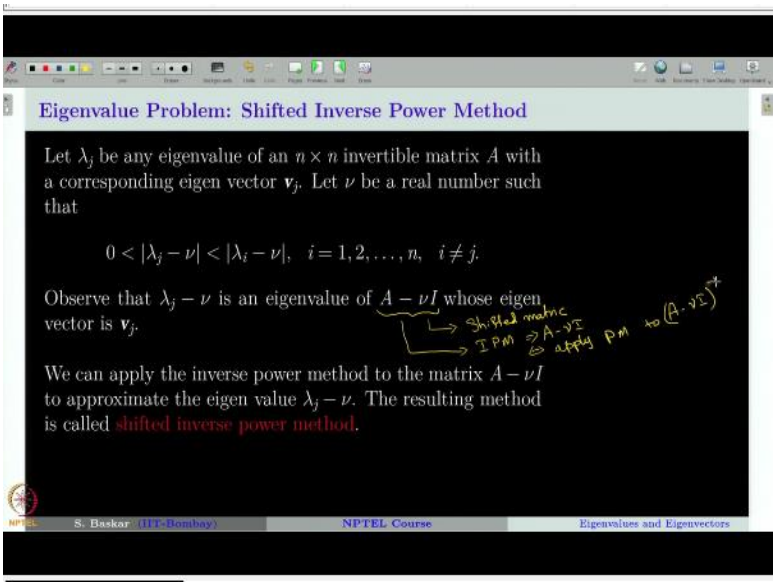
Practically, this is not possible to get because simply we do not know what is λ_j. However, in certain situations we may use Gerschgorin's disk to get an estimate of ν. For instance, suppose your matrix *A* is such that the Gerschgorin's disk are something like this disk 1, disk 2 and so on disk 3 and then disk 4 like that, disk 5 like that and then in between you have a small disc which is disjoint from here.

And you want to find this eigenvalue; let us call this as our λ_j. In such cases you know how to choose your ν. Say you take this interval and suppose this interval is very small then you have an idea of how to choose this ν. So, like this in certain restricted situations you may be able to

choose such ν′s using Gerschgorin's theorem. If such a favourable situation is not happening with the matrix A then you can also try with A^T.

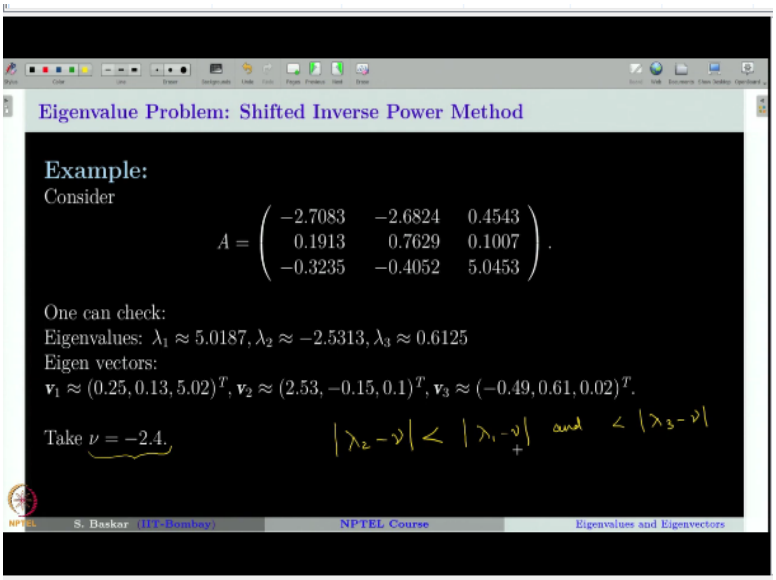
Sometimes that may give you a better information about how the eigenvalues are distributed. So, if such a favorable situation happens then one may go to use shifted inverse power method. Otherwise, it is not that easy for us to use this method in any practical situations. Assume that we got this ν somehow then how to compute λ_j is our question now. Once you have such a ν then probably by looking at this you can get an idea of how we are going to capture λ.

(Refer Slide Time: 19:38)



What you can do is, you apply the inverse power method to $A - \nu I$. So, this is called the shifted matrix and you apply inverse power method to this matrix. Inverse power method is applied to $A - \nu I$ that is equivalent to saying that apply power method to $(A - \nu I)^{-1}$. That is the idea.

(Refer Slide Time: 20:25)

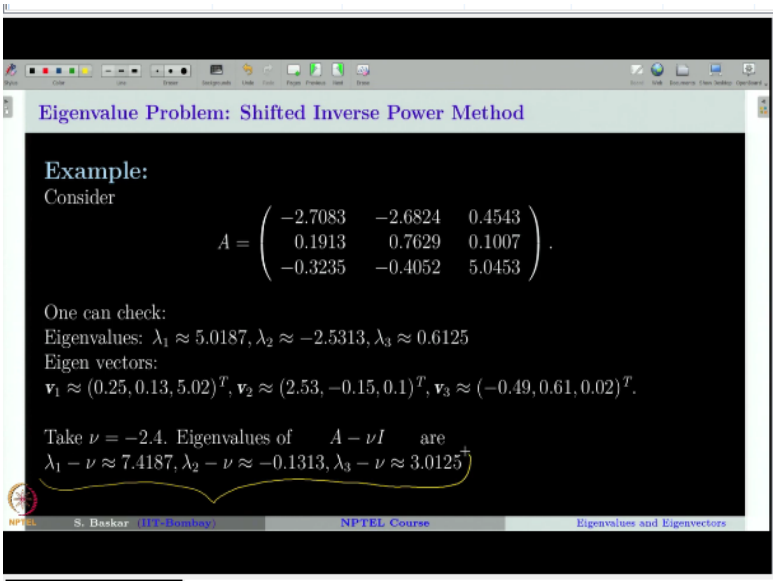


We do not need to give the procedure for this because it is just the same as what we have given in the inverse power method; only thing is instead of A^{-1} now we have to put $(A - \nu I)^{-1}$. That is the only difference. Let us consider this example. Now once you have this you will have to use an appropriate ν. As I told you this is generally not possible perhaps if you are very lucky you may get an idea of how to choose this nu from Gerschgorin's disk of either A or A^T.

I will leave it to you to think how to do that for some appropriate matrices; we will give you some such matrices in exercise problems. But here, to keep our discussion very simple; let us know what are all the eigenvalues and the corresponding eigenvectors of this matrix A. These are given by λ_1, λ_2 and λ_3 and the corresponding eigenvectors are given here. Now let us try to capture λ_2 and eigenvector of λ_2.

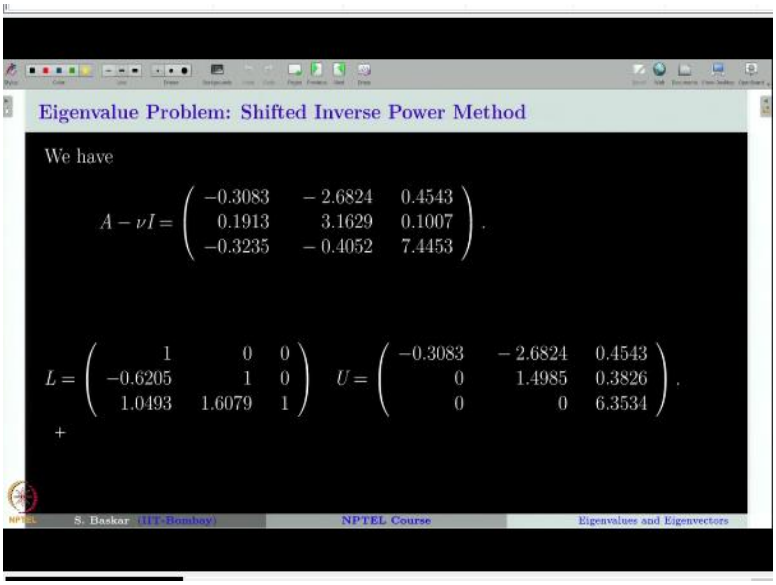
Let us see how this is going to happen. For that we have to choose and ν, let me choose ν as -2.4, you can choose any number in such a way that $|\lambda_2 - \nu| < |\lambda_1 - \nu|$ and this is strictly less than $|\lambda_3 - \nu|$ also. So, this choice of ν is going to work; therefore I chose like this. You can also choose any value.

(Refer Slide Time: 22:25)



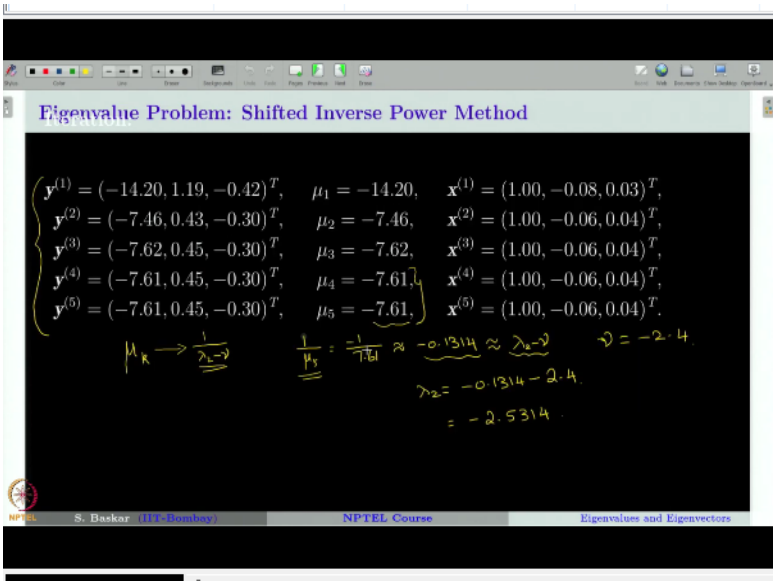
And now we have to apply the inverse power method for $A - \nu I$ that is $A + 2.4\, I$. You can clearly see that the eigenvalues of $A - \nu I$ are given like this.

(Refer Slide Time: 22:44)



Now we have to find the LU-factorization of $A - \nu I$. I have gone for the Doolittle factorizations, you can go for either Gaussian elimination or Crout factorization or even if the matrix is symmetric and positive definite you can go for Cholesky's factorization also. But I have gone here for Doolittle factorization. As per this the lower triangular and the upper triangular matrices are given like this.

(Refer Slide Time: 23:19)



And the iterations are given like this. I have stopped this iteration procedure at the 5th iteration because you can see that from the 4th iteration to 5th iteration there was no much improvement at least to the number of digits that we have shown here. That is why I have stopped up to here. If you want more accurate then you have to go for more iterations. That is the only need here.

But I am happy with this because it is just an illustration. So, I stop here. Therefore, you are iteration μ_k sequence is converging to remember it is going to converge to $\frac{1}{\lambda_2 - \nu}$. Remember we are now working with shifted matrix. Therefore, your μ_k is going to converge to $\frac{1}{\lambda_2 - \nu}$ and since μ_5 is given by -7.61, you can see that $\frac{1}{\mu_5}$ is something like of course -1 / 7.61.

That is in my calculation approximately equal to –0.1314. So, that is going to be approximately equal to $\lambda_2 - \nu$, because μ is going to converge to $\lambda_2 - \nu$. Therefore $\frac{1}{\mu_k}$ will converge to $\lambda_2 - \nu$. I have just stopped at the 5th iteration. Therefore, I am going to consider this value as an approximate value to $\lambda_2 - \nu$, but remember ν is taken as -2.4.

Therefore λ_2 = -0.1314 + ν that is nothing but -2.4 and that is equal to –2.5314. So, that is the λ that we are trying to obtain from the shifted inverse power method. Let us see what is the value of λ that we have taken previously. That is also approximately the same; it is -2.5313, what we got is -2.5314. So, we are pretty good in our approximation and the corresponding eigenvector captured by our shifted inverse power method is (1, -0.06, 0.04).

You can also see that it will be a scalar multiple of the corresponding eigenvector that we have chosen in our first line. So, this gives us an idea of how we may capture other eigenvalues of a given matrix. But there is a serious disadvantage in this approach that we do not know how to choose this scaling parameter ν. Apart from that the method is very interesting. With this we will finish our discussion of power methods. I thank you for your attention.