MA 214: Introduction to numerical analysis

Lecture 43

Shripad M. Garge.
IIT Bombay

(shripad@math.iitb.ac.in)

2021-2022

Iterative techniques

We now discuss iterative techniques to solve systems of linear equations.

Iterative techniques are seldom used for solving linear systems of small dimensions since the time required for sufficient accuracy exceeds that required for direct techniques.

We will study the Jacobi and the Gauß-Seidel iterative methods, classic methods that date to the late eighteenth century.

An iterative technique to solve the $n \times n$ linear system $Ax = b$ starts with an initial approximation $x^{(0)}$ to the solution x and generates a sequence of vectors $\{x^{(k)}\}_{k=0}^{\infty}$ that converges to x.

Jacobi method

We make an assumption that, in addition to the $\det(A)$ being nonzero, the diagonal entries of A are also nonzero.

Each equation in the system reads

$$\sum_j a_{ij}x_j = b_i \quad \text{or} \quad x_i = \frac{b_i - \sum_{j,j\neq i} a_{ij}x_j}{a_{ii}}.$$

This allows us to do the iterations.

Jacobi suggested in 1834 that we start with an initial vector $x^{(0)} = [x_1^{(0)}, \ldots, x_n^{(0)}]^t$ and define, for $k \geqslant 1$,

$$x_i^{(k)} = \frac{b_i - \sum_{j,j\neq i} a_{ij}x_j^{(k-1)}}{a_{ii}}.$$

Jacobi method

Let the vector $x = [x_1, \ldots, x_n]^t$ be the answer of the system and let us calculate the errors in this method:

$$x_i - x_i^{(k)} = \sum_{j, j \neq i} \frac{a_{ij}}{a_{ii}} \left(x_j - x_j^{(k-1)} \right).$$

Hence

$$|x_i - x_i^{(k)}| \leqslant \|x - x^{(k-1)}\|_\infty \left(\sum_{j, j \neq i} \left| \frac{a_{ij}}{a_{ii}} \right| \right)$$

and hence

$$\|x - x^{(k)}\|_\infty \leqslant \|x - x^{(k-1)}\|_\infty \left(\max_i \sum_{j, j \neq i} \left| \frac{a_{ij}}{a_{ii}} \right| \right).$$

Let $\mu = \max_i \sum_{j, j \neq i} \left| \frac{a_{ij}}{a_{ii}} \right|$.

Jacobi method

If $\mu < 1$, then convergence is guaranteed in Jacobi method. This condition translates to

$$\sum_{j, j \neq i} |a_{ij}| < |a_{ii}|.$$

Recall that a matrix satisfying this condition is called strictly diagonally dominant.

Consider the system:

$$\begin{pmatrix} 10 & -1 & 2 & 0 \\ -1 & 11 & -1 & 3 \\ 2 & -1 & 10 & -1 \\ 0 & 3 & -1 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 6 \\ 25 \\ -11 \\ 15 \end{pmatrix}.$$

We start with $x^{(0)} = (0, 0, 0, 0)^t$ and find approximations $x^{(k)}$.

Jacobi method

The results are

$$x^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad x^{(1)} = \begin{pmatrix} 0.6 \\ 2.2727 \\ -1.1 \\ 1.875 \end{pmatrix}, \ldots, x^{(10)} = \begin{pmatrix} 1.0001 \\ 1.9998 \\ -0.9998 \\ 0.9998 \end{pmatrix}.$$

The solution of the system is $(1, 2, -1, 1)^t$ and already in 10 steps we achieved a good approximation.

As has been the theme of the 19-th century mathematics, Gauß had a better method up his sleeve!

Gauss had mentioned his method in a private correspondence to his student Gerling in 1823, but it was published much later, in 1874, by Seidel, having worked it out on his own.

Gauss-Seidel method

The idea is that once we have improved the first component of $x^{(k)}$, we should use that while improving the second component of $x^{(k)}$, and so on.

The components of $x^{(k-1)}$ are used to compute all the components $x_i^{(k)}$ of $x^{(k)}$. But, for $i > 1$, the components $x_1^{(k)}, \ldots, x_{i-1}^{(k)}$ of $x^{(k)}$ have already been computed and are expected to be better approximations to the actual solutions $x_1, \ldots, x_{i-1}$ than are $x_1^{(k-1)}, \ldots, x_{i-1}^{(k-1)}$.

It seems reasonable, then, to compute $x_i^{(k)}$ using these most recently calculated values. We therefore use

$$x_i^{(k)} = \frac{1}{a_{ii}} \left[b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k)} - \sum_{j=i+1}^{n} a_{ij} x_j^{(k-1)} \right].$$

Gauss-Seidel method

We use this technique to solve the earlier example:

$$\begin{pmatrix} 10 & -1 & 2 & 0 \\ -1 & 11 & -1 & 3 \\ 2 & -1 & 10 & -1 \\ 0 & 3 & -1 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 6 \\ 25 \\ -11 \\ 15 \end{pmatrix}.$$

We start with $x^{(0)} = (0, 0, 0, 0)^t$ and find approximations $x^{(k)}$ using Gauß-Seidel method.

We get

$$x^{(0)} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad x^{(1)} = \begin{pmatrix} 0.6 \\ 2.3272 \\ -0.9873 \\ 0.8789 \end{pmatrix}, \ldots, x^{(5)} = \begin{pmatrix} 1.0001 \\ 2.0001 \\ -1.0000 \\ 1.0000 \end{pmatrix}.$$

Jacobi vis-á-vis Gauß-Seidel

k	0	1	2	3	4	5
$x_1^{(k)}$	0.0000	0.6000	1.0473	0.9326	1.0152	0.9890
$x_2^{(k)}$	0.0000	2.2727	1.7159	2.053	1.9537	2.0114
$x_3^{(k)}$	0.0000	−1.1000	−0.8052	−1.0493	−0.9681	−1.0103
$x_4^{(k)}$	0.0000	1.8750	0.8852	1.1309	0.9739	1.0214

k	0	1	2
$x_1^{(k)}$	0.0000	0.6000	1.030
$x_2^{(k)}$	0.0000	2.3272	2.037
$x_3^{(k)}$	0.0000	−0.9873	−1.014
$x_4^{(k)}$	0.0000	0.8789	0.9844

6	7	8	9	10
1.0032	0.9981	1.0006	0.9997	1.0001
1.9922	2.0023	1.9987	2.0004	1.9998
−0.9945	−1.0020	−0.9990	−1.0004	−0.9998
0.9944	1.0036	0.9989	1.0006	0.9998

3	4	5
1.0065	1.0009	1.0001
2.0036	2.0003	2.0000
−1.0025	−1.0003	−1.0000
0.9983	0.9999	1.0000

The above results appear to imply that the Gauß-Seidel method is superior to the Jacobi method.

This is almost always true, but there are linear systems for which the Jacobi method converges and the Gauß-Seidel method does not.

Such linear systems will correspond to matrices which are not strictly diagonally dominant, as the following result shows.

If A is strictly diagonally dominant, then for any choice of $x^{(0)}$, both the Jacobi and Gauß-Seidel methods give sequences $\{x^{(k)}\}_{k=0}^{\infty}$ that converge to the unique solution of $Ax = b$.

MA 214: Introduction to numerical analysis

Lecture 44

Shripad M. Garge.
IIT Bombay

(shripad@math.iitb.ac.in)

2021-2022

Residual vector

We have seen some direct and iterative methods of solving systems of linear equations, $Ax = b$.

If $\tilde{x}$ is an approximation to the solution of $Ax = b$ then the vector $r = b - A\tilde{x}$ is called the residual vector.

It seems intuitively reasonable that if $\|r\|$ is small, then $\|x - \tilde{x}\|$ would be small as well.

This is often the case, but certain systems, which occur frequently in practice, fail to have this property.

Residual vector

Consider the system

$$\begin{pmatrix} 1 & 2 \\ 1.0001 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3.0001 \end{pmatrix}$$

with solution $x = (1, 1)^t$.

We assume that $\tilde{x} = (3, -0.0001)^t$ is an approximation to x. The residual vector then is $r = b - A\tilde{x} = (0.0002, 0)^t$.

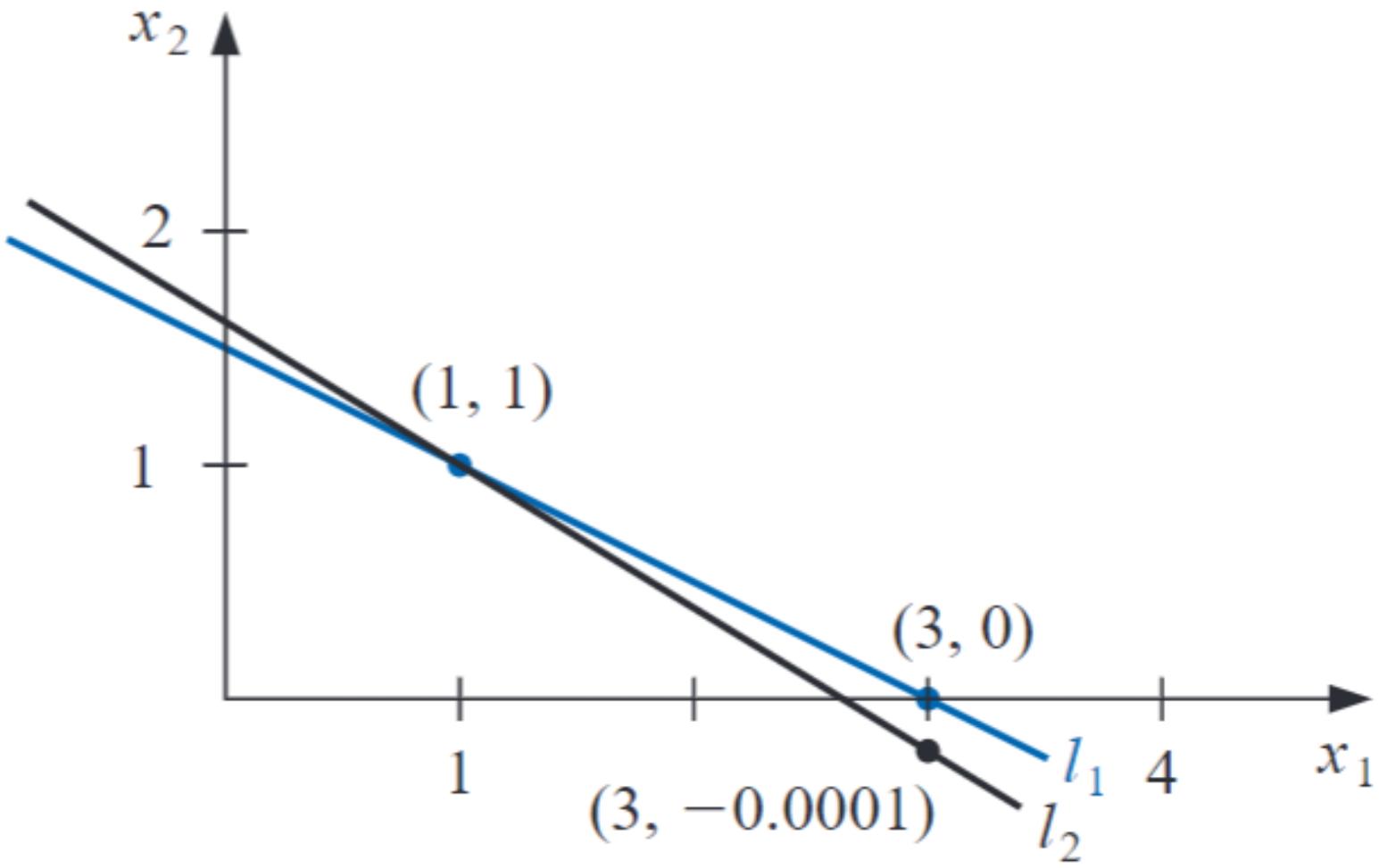
Hence $\|r\|_\infty = 0.0002$ but the approximation $\tilde{x} = (3, -0.0001)^t$ is evidently poor for the solution $x = (1, 1)^t$.

The solution to the system represents the intersection of the lines

$$l_1 : \quad x_1 + 2x_2 = 3 \qquad \text{and} \qquad l_2 : \quad 1.0001x_1 + 2x_2 = 3.0001$$

which are nearly parallel and the point $(3, -0.0001)^t$ lies on l_2.

Residual vector



Residual vector

The example was clearly constructed to show the difficulties that can - and, in fact, do - arise.

Had the lines not been nearly coincident, we would expect a small residual vector to imply a fairly accurate approximation.

In the general situation, we cannot rely on the geometry of the system to give an indication of when problems might occur.

We can, however, obtain this information by considering the norms of the matrix A and its inverse.

Condition number

Theorem

Let $\tilde{x}$ be an approximate solution to $Ax = b$ for a non-singular A and let $r = b - A\tilde{x}$ be the corresponding residual vector. Then

$$\|x - \tilde{x}\|_\infty \leqslant \|r\|_\infty \cdot \|A^{-1}\|_\infty.$$

Further, if $x \neq 0$ and $b \neq 0$ then

$$\frac{\|x - \tilde{x}\|_\infty}{\|x\|_\infty} \leqslant \|A\|_\infty \cdot \|A^{-1}\|_\infty \cdot \frac{\|r\|_\infty}{\|b\|_\infty}.$$

The condition number of a non-singular matrix A is

$$K(A) = \|A\|_\infty \cdot \|A^{-1}\|_\infty.$$

Condition number

With this definition, we rewrite the above theorem as

$$\|x - \tilde{x}\|_\infty \leqslant K(A) \cdot \frac{\|r\|_\infty}{\|A\|_\infty}$$

and

$$\frac{\|x - \tilde{x}\|_\infty}{\|x\|_\infty} \leqslant K(A) \cdot \frac{\|r\|_\infty}{\|b\|_\infty}.$$

We have

$$1 = \|I\|_\infty = \|AA^{-1}\|_\infty \leqslant \|A\|_\infty \|A^{-1}\|_\infty = K(A).$$

A non-singular matrix A is said to be well-conditioned if $K(A)$ is close to 1 and is called ill-conditioned if $K(A)$ is significantly greater than 1.

Condition number

For instance, the matrix we worked with in the above example is

$$A = \begin{pmatrix} 1 & 2 \\ 1.0001 & 2 \end{pmatrix} \quad \text{and} \quad A^{-1} = \begin{pmatrix} -10000 & 10000 \\ 5000.5 & -5000 \end{pmatrix}.$$

Then

$$\|A\|_\infty = \max\{|1| + |2|, |1.0001| + |2|\} = 3.0001$$

and, similarly, $\|A^{-1}\|_\infty = 20000$, hence the condition number is

$$K(A) = \|A\|_\infty \cdot \|A^{-1}\|_\infty = 3.0001 \cdot 20000 = 60002.$$

The size of the condition number for this example should certainly keep us from making hasty accuracy decisions based on the residual of an approximation.

Condition number without A^{-1}

Although the condition number of a matrix depends totally on the norms of the matrix and its inverse, the calculation of the inverse is subject to round-off error and is dependent on the accuracy with which the calculations are performed.

If the operations involve arithmetic with t digits of accuracy, the approximate condition number for the matrix A is the norm of A times the norm of the approximation A^{-1}, which is obtained using t-digit arithmetic.

In fact, this condition number also depends on the method used to calculate the inverse of A.

In addition, because of the number of calculations needed to compute the inverse, we need to be able to estimate the condition number without directly determining the inverse.

Condition number without A^{-1}

If we assume that the approximate solution to the linear system $Ax = b$ is being determined using t-digit arithmetic and Gaussian elimination then the residual vector r for the approximation $\tilde{x}$ has

$$\|r\|_\infty \approx 10^{-t}\|A\|_\infty \cdot \|\tilde{x}\|_\infty.$$

From this approximation, an estimate for the effective condition number in t-digit arithmetic can be obtained without the need to invert the matrix A.

One then computes the residual vector r in double precision, that is, in $2t$-digit arithmetic.

The approximation for the t-digit condition number $K(A)$ comes from consideration of the linear system $Ay = r$.

Condition number without A^{-1}

An approximate solution $\tilde{y}$ to $Ay = r$ satisfies

$$\tilde{y} \approx A^{-1}r = A^{-1}(b - A\tilde{x}) = A^{-1}b - \tilde{x} = x - \tilde{x}.$$

Thus, $\tilde{y}$ is an estimate of the error produced when $\tilde{x}$ approximates the solution x to the original system. We then have

$$\begin{aligned} \|\tilde{y}\| &\approx \|A^{-1}r\| \leqslant \|A^{-1}\| \cdot \|r\| \\ &\approx \|A^{-1}\| \cdot (10^{-t} \cdot \|A\| \cdot \|\tilde{x}\|) \\ &= 10^{-t} \cdot \|\tilde{x}\| \cdot K(A). \end{aligned}$$

This gives an approximation for the condition number involved with solving the system $Ax = b$ using Gaussian elimination and the t-digit arithmetic:

$$K(A) \approx \frac{\|\tilde{y}\|}{\|\tilde{x}\|} 10^t.$$

Example

Consider the linear system

$$\begin{pmatrix} 3.3330 & 15920 & -10.333 \\ 2.2220 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 15913 \\ 28.544 \\ 8.4254 \end{pmatrix}$$

which has the exact solution $(1, 1, 1)^t$.

Using Gaussian elimination and five-digit rounding arithmetic leads to the approximate solution

$$\tilde{x} = (1.2001, 0.99991, 0.92538)^t.$$

The residual vector corresponding to $\tilde{x}$ computed in double precision is

$$r = (-0.00518, 0.27412914, -0.186160367)^t, \qquad \|r\| = 0.27413.$$

Example, continued

We now need to compute $\tilde{y}$, the approximate solution to $Ay = r$.

We have

$$\tilde{y} = (-0.20008, 8.9987 \times 10^{-5}, 0.074607)^t$$

and then

$$K(A) \approx \frac{\|\tilde{y}\|}{\|\tilde{x}\|} 10^5 = \frac{0.20008}{1.2001} 10^5 = 16672.$$

To determine the exact condition number, we must determine A^{-1} first.

Example, continued

Using five-digit rounding arithmetic

$$A^{-1} = \begin{pmatrix} -1.1701 \times 10^{-4} & -1.4983 \times 10^{-1} & 8.5416 \times 10^{-1} \\ 6.2782 \times 10^{-5} & 1.2124 \times 10^{-4} & -3.0662 \times 10^{-4} \\ -8.6631 \times 10^{-5} & 1.3846 \times 10^{-1} & -1.9689 \times 10^{-1} \end{pmatrix}.$$

Then

$$\|A\| = 15934 \quad \text{and} \quad \|A^{-1}\| = 1.0041$$

and

$$K(A) = \|A\| \cdot \|A^{-1}\| = 15934 \times 1.0041 = 15999.$$

The estimate computed above is quite close to the actual value of $K(A)$ and requires considerably less computational effort.

Iterative refinement

In the above analysis, we used the estimate

$$\tilde{y} \approx x - \tilde{x}$$

where $\tilde{y}$ is the approximate solution to the system $Ay = r$.

In general, $\tilde{x} + \tilde{y}$ is a more accurate approximation to the solution of the linear system $Ax = b$ than the original approximation $\tilde{x}$.

The method using this assumption is called iterative refinement, or iterative improvement.

It consists of performing iterations on the system whose right-hand side is the residual vector for successive approximations until satisfactory accuracy results.

Iterative refinement

If the process is applied using t-digit arithmetic and if

$$K(A) \approx 10^q,$$

then after k iterations of iterative refinement the solution has approximately the smaller of t and $k(t-q)$ correct digits.

If the system is well-conditioned, one or two iterations will indicate that the solution is accurate.

There is the possibility of significant improvement on ill-conditioned systems unless the matrix A is so ill-conditioned that

$$K(A) > 10^t.$$

In that situation, increased precision should be used for the calculations.

MA 214: Introduction to numerical analysis

Lecture 45

Shripad M. Garge.
IIT Bombay

(shripad@math.iitb.ac.in)

2021-2022

Eigenvalue of a matrix

Let A be an $n \times n$ matrix with real entries.

An eigenvalue of A is a number $\lambda \in \mathbb{C}$ such that $Av = \lambda v$ for a non-zero vector v.

To determine the eigenvalues of A, we construct the characteristic polynomial $p(\lambda) = \det(A - \lambda I)$ and then determine its zeroes.

Finding the determinant of an $n \times n$ matrix is computationally expensive, and finding good approximations to the roots of $p(\lambda)$ is also difficult.

We will explore other means for approximating the eigenvalues of a matrix.

Gerschgorin theorem

Theorem

Let $A = [a_{ij}]$ be an $n \times n$ complex matrix and define

$$D_i = \left\{ z \in \mathbb{C} : |z - a_{ii}| \leqslant \sum_{j \neq i} |a_{ij}| \right\}.$$

Then all eigenvalues of A are contained in the union of all the disks D_i.

Moreover, the union of any k of the disks that do not intersect the remaining $(n-k)$ contains precisely k (counting multiplicities) of the eigenvalues.

Thus, if there are n disks not intersecting each other then each one contains exactly one eigenvalue!

Gerschgorin theorem

Consider the matrix

$$A = \begin{pmatrix} 4 & 1 & 1 \\ 0 & 2 & 1 \\ -2 & 0 & 9 \end{pmatrix}.$$

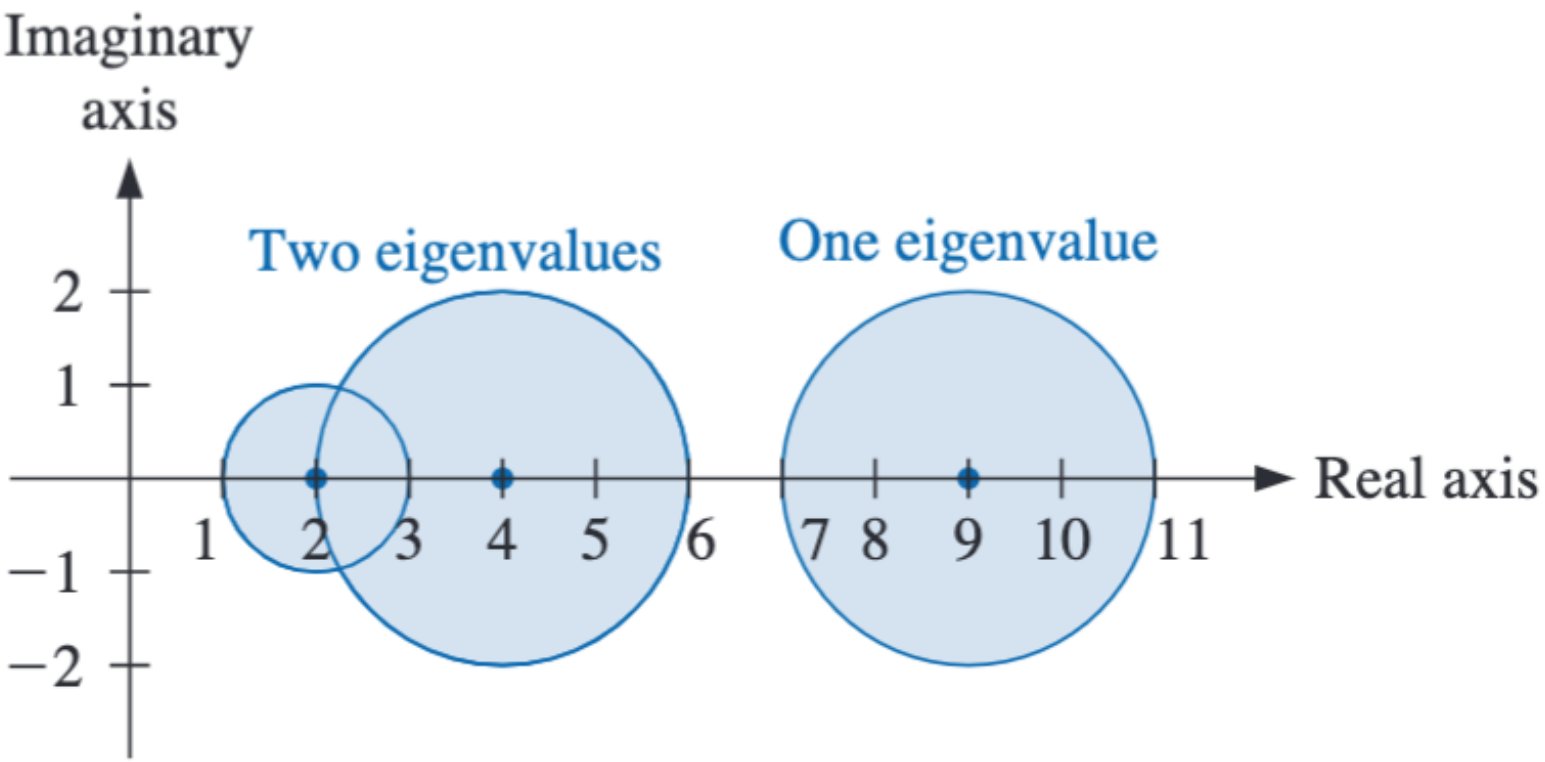
The disks in this case are

$$D_1 = \{z \in \mathbb{C} : |z - 4| \leqslant 2\}, \quad D_2 = \{z \in \mathbb{C} : |z - 2| \leqslant 1\},$$

$$\text{and} \quad D_3 = \{z \in \mathbb{C} : |z - 9| \leqslant 2\}.$$

Since D_1 and D_2 are disjoint from D_3, there are precisely two eigenvalues in $D_1 \cup D_2$ and exactly one eigenvalue in D_3.

Gerschgorin theorem



Gerschgorin theorem

If A is strictly diagonally row dominant, $\sum_{j \neq i} |a_{ij}| < |a_{ii}|$ for every i, then A is invertible.

This is because 0 cannot belong to any D_i.

What would happen if A is strictly diagonally column dominant?

Consider A^T. ..... ☺

Now when we need to find the eigenvalues, many techniques for their approximation are iterative.

Determining regions in which they lie is the first step for finding the approximation, because it provides us with an initial approximations.

Power method

Now that we have some control on locating the eigenvalues of A let us go about finding them. We study the "power method".

Let A be an $n \times n$ complex matrix.

We assume that its eigenvalues satisfy $|\lambda_1| > |\lambda_2| \geqslant \cdots \geqslant |\lambda_n|$. Here λ_1 is called the dominant eigenvalue of A.

We also assume that A has n linearly independent eigenvectors, $\{v_1, \ldots, v_n\}$, $Av_i = \lambda_i v_i$.

Choose a nonzero $z \in V$. The method consists of computing $A^k z$ and taking the limit

$$\lim_{k \to \infty} A^k z.$$

Here, if $z = \sum \alpha_i v_i$ then

$$\begin{aligned} A^k z &= \lambda_1^k \alpha_1 v_1 + \lambda_2^k \alpha_2 v_2 + \cdots + \lambda_n^k \alpha_n v_n \\ &= \lambda_1^k \left[\alpha_1 v_1 + \left(\frac{\lambda_2}{\lambda_1}\right)^k \alpha_2 v_2 + \cdots + \left(\frac{\lambda_n}{\lambda_1}\right)^k \alpha_n v_n \right] \end{aligned}$$

Thus, as $k \to \infty$, $A^k z \to \lambda_1^k \alpha_1 v_1$, an eigenvector for λ_1.

Of course, when $0 \neq z$ is chosen at random the coefficient α_1 is "likely" to be non-zero. One can modify the method if $\alpha_1 = 0$.

If A is invertible and if

$$|\lambda_1| \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_{n-1}| > |\lambda_n| > 0,$$

then use the power method for A^{-1}.

Power method

If the dominant eigenvalue λ_1 has the property that $|\lambda_1| < 1$ then $\lambda_1^k \to 0$.

On the other hand, if $|\lambda_1| > 1$ then $\lambda_1^k \to \infty$.

To take care of these two possibilities, we scale $A^k(z)$ so that the limit, $\lim_{k\to\infty} A^k(z)$, is finite and non-zero.

The scaling begins by choosing z to be a vector $z^{(0)}$ of unit norm, that is, $\|z^{(0)}\|_\infty = 1$ and choosing a component $z_{p_0}^{(0)}$ of $z^{(0)}$ with $|z_{p_0}^{(0)}| = 1$.

Let $w^{(1)} = Az^{(0)}$ and define $\mu^{(1)} = w_{p_0}^{(1)}$. Then

$$\mu^{(1)} = w_{p_0}^{(1)} = \frac{w_{p_0}^{(1)}}{z_{p_0}^{(0)}}.$$

We know that the vector $z^{(0)}$ is assumed to be of the form

$$z^{(0)} = \alpha_1 v_1 + \cdots + \alpha_n v_n.$$

Taking the p_0-component

$$z^{(0)}_{p_0} = \alpha_1 (v_1)_{p_0} + \cdots + \alpha_n (v_n)_{p_0}.$$

Similarly,

$$\mu^{(1)} = \frac{w^{(1)}_{p_0}}{z^{(0)}_{p_0}} = \frac{\lambda_1 \alpha_1 (v_1)_{p_0} + \cdots + \lambda_n \alpha_n (v_n)_{p_0}}{\alpha_1 (v_1)_{p_0} + \cdots + \alpha_n (v_n)_{p_0}}$$

$$= \lambda_1 \frac{\alpha_1 (v_1)_{p_0} + \cdots + (\lambda_n/\lambda_1) \alpha_n (v_n)_{p_0}}{\alpha_1 (v_1)_{p_0} + \cdots + \alpha_n (v_n)_{p_0}}.$$

Power method

We choose p_1 to be the least integer with $|w_{p_1}^{(1)}| = \|w^{(1)}\|$ and define $z^{(1)}$ by

$$z^{(1)} = \frac{1}{w_{p_1}^{(1)}} w^{(1)} = \frac{1}{w_{p_1}^{(1)}} Az^{(0)}.$$

This is a vector with $\|z^{(1)}\|_\infty = 1$.

We then repeat the process.

Assume that a vector $z^{(m-1)}$ of ∞-norm one is obtained in the $(m-1)$-th step and p_{m-1} is the least integer with $|z_{p_{m-1}}^{(m-1)}| = 1$.

Then we define $w^{(m)} = Az^{(m-1)}$ and

$$\mu^{(m)} = w_{p_{m-1}}^{(m)} = \lambda_1 \left(\frac{\alpha_1 (v_1)_{p_{m-1}} + \cdots + (\lambda_n/\lambda_1)^m \alpha_n (v_n)_{p_{m-1}}}{\alpha_1 (v_1)_{p_{m-1}} + \cdots + (\lambda_n/\lambda_1)^{m-1} \alpha_n (v_n)_{p_{m-1}}} \right).$$

Power method

Further, we define $z^{(m)} = w^{(m)}/w_{p_m}^{(m)}$, so that $\|z^{(m)}\|_\infty = 1$.

We now observe that $\lim_{m\to\infty} \mu^{(m)} = \lambda_1$.

Moreover, the sequence $z^{(m)}$ converges to a vector of norm one which turns out to be an eigenvector for λ_1.

Let us consider the matrix

$$A = \begin{pmatrix} -2 & -3 \\ 6 & 7 \end{pmatrix}$$

with eigenvalues $\lambda_1 = 4$ and $\lambda_2 = 1$.

If we start with $z^{(0)} = (1,1)^t$ then $Az^{(0)} = (-5,13)^t$, $A^2z^{(0)} = (-29,61)^t$, $\ldots$, $A^6z^{(0)} = (-8189,16381)^t$.

Power method

The approximations to the dominant eigenvalue are

$$\mu^{(1)} = \frac{61}{13} = 4.6923, \ldots, \mu^{(5)} = \frac{16381}{4093} = 4.0022.$$

Taking the norm one vector parallel to $A^6 z^{(0)}$ gives us

$$\begin{pmatrix} -0.4999 \\ 1 \end{pmatrix}$$

which is almost an eigenvector for A with eigenvalue 4.

How do we find other eigenvalues?

Suppose that A is a matrix and that you have found an eigenvector v for A with the eigenvalue λ, sufficiently accurately.

We then use the Gram-Schmidt orthonormalisation to write an ONB for V starting with v, say $\{v, w_2, w_3, \ldots, w_n\}$.

We write the matrix A with respect to this basis, it has the form

$$\begin{pmatrix} \lambda & b_{12} & \ldots & b_{1n} \\ 0 & b_{22} & \ldots & b_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & b_{n2} & \ldots & b_{nn} \end{pmatrix}.$$

The eigenvalues of A are λ and the eigenvalues of the bottom-left $(n-1) \times (n-1)$ matrix $[b_{ij}]_{i,j \geqslant 2}$.

What next?

This completes our syllabus.

Thank you!