Solving Linear Systems of Equations

Kostas Kokkotas

October 29, 2019

Eberhard Karls Univerity of Tübingen

Introduction

The type of problems that we have to solve are:

- Solve the system: $\mathbf{A} \cdot \mathbf{x} = \mathbf{B}$, where

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & \cdots & a_{1N} \\ a_{12} & & & a_{2N} \\ \vdots & & & \vdots \\ a_{1N} & \cdots & \cdots & a_{NN} \end{pmatrix} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \quad \mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}$$

- To find $\mathbf{A}^{-1}$ (inverse of a matrix)
- The determinant of a matrix $\mathbf{A}$
- The eigenvalues and eigenvectors of a matrix $\mathbf{A}$

Gauss Method i

Let assume that we have to solve the linear system $\mathbf{A} \cdot \mathbf{x} = \mathbf{B}$ with $\det(\mathbf{A}) \neq 0$.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \ldots + a_{1N}x_N &= b_1 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \ldots + a_{2N}x_N &= b_2 \\ &\vdots \\ a_{N1}x_1 + a_{N2}x_2 + a_{N3}x_3 + \ldots + a_{NN}x_N &= b_N \end{aligned} \tag{1}$$

We will try to transform it into an **upper-triangular** linear system. The steps are the following:

Gauss Method ii

STEP 1

Multiply the 1st equation with a_{21}/a_{11} and substract it from the 2nd.

Similarly, multiply the 1st equation with a_{31}/a_{11} and substract it from the 3rd etc

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1N}x_N &= b_1 \\ 0 + a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \cdots + a_{2N}^{(1)}x_N &= b_2^{(1)} \\ \vdots \quad & \vdots \\ 0 + a_{N2}^{(1)}x_2 + a_{N3}^{(1)}x_3 + \cdots + a_{NN}^{(1)}x_N &= b_N^{(1)} \end{aligned} \tag{2}$$

where, for example, we set:

$$a_{22}^{(1)} = \frac{a_{12}a_{21}}{a_{11}} - a_{22}$$

.

Gauss Method iii

STEP 2

The 1st row and the 1st column remain unchanged and we multiply the 2nd equation with $a_{32}^{(1)}/a_{22}^{(1)}$ and substract it from the 3rd and so on.

Thus after $n-1$ operations we get

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \cdots + a_{1N}x_N & = & b_1 \\ 0 \quad + a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \cdots + a_{2N}^{(1)}x_N & = & b_2^{(1)} \\ 0 \qquad + 0 \quad + a_{33}^{(2)}x_3 + \cdots + a_{3N}^{(2)}x_N & = & b_3^{(2)} \\ \qquad\qquad\qquad\qquad \vdots \quad \vdots & & \\ 0 \qquad + 0 \quad + a_{N3}^{(2)}x_3 + \cdots + a_{NN}^{(2)}x_N & = & b_N^{(2)} \end{array} \tag{3}$$

Where for example we have set:

$$a_{33}^{(2)} = \frac{a_{32}^{(1)}\, a_{33}^{(1)}}{a_{22}^{(1)}} - a_{33}^{(1)}$$

.

Gauss Method iv

STEP $N-1$

After $N-1$ steps we get the following **upper-triangular** system:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \quad &\cdots \quad + a_{1\,N-1}x_{N-1} + a_{1N}x_N &&= b_1 \\
a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \quad &\cdots \quad + a_{2\,N-1}^{(1)}x_{N-1} + a_{2N}^{(1)}x_N &&= b_2^{(1)} \\
a_{33}^{(2)}x_3 + \quad &\cdots \quad + a_{3\,N-1}^{(2)}x_{N-1} + a_{3N}^{(2)}x_N &&= b_3^{(2)} \\
&\vdots && \\
a_{N-1\,N-1}^{(N-2)}x_{N-1} + a_{N-1\,N}^{(N-2)}x_N &&= b_{N-1}^{(N-2)} \\
a_{NN}^{(N-1)}x_N &&= b_N^{(N-1)}
\end{aligned}
\tag{4}
$$

Gauss Method v

The solution of this system is obvious:

- For the Nth equation of the previous upper-triangular system (4) we get :

$$x_N = \frac{b_N^{(N-1)}}{a_{NN}^{(N-1)}} \quad \text{for} \quad a_{NN}^{(N-1)} \neq 0 \tag{5}$$

- while the rest of the values can be calculated via the relation:

$$x_i = \frac{b_i^{(i-1)} - \sum_{k=i+1}^{N} a_{ik}^{(i-1)} x_k}{a_{ii}^{(i-1)}} \quad \text{for} \quad a_{ii}^{(i-1)} \neq 0 \tag{6}$$

The number of arithmetic operations needed is:

$$\frac{1}{6}\left(4N^3 + 9N^2 - 7N\right) \approx \frac{2}{3}N^3 .$$

Determinant

If a matrix is transformed into an upper-triangular or lower-triangular or diagonal form then

$$\det \mathbf{A} = a_{11} \cdot a_{22}^{(1)} \cdot a_{33}^{(2)} \cdots a_{NN}^{(N-1)} = \prod_{i=1}^{N} a_{ii}^{(i-1)} \tag{7}$$

Pivoting i

DEFINITION: The number a_{ii} in the position (i, i) that is used to eliminate x_i in rows $i+1$, $i+2$, ... ,N is called the ith **pivotal element** and the ith row is called the **pivotal row**.

- **Pivoting to avoid $a_{ii}^{(i-1)} = 0$:** If $a_{ii}^{(i-1)} = 0$, row i cannot be used to eliminate, the elements in column i below the diagonal. It is neccesary to find a row j, where $a_{ji}^{(i-1)} \neq 0$ and $j > i$ and then interchange row i and j so that a nonzero pivot element is obtained.

- **Pivoting to reduce error :** If there is more than one nonzero element in column i that lies on or below the diagonal, there is a choice to determine which rows to interchange.

Because, the computer uses fixed-precision arithmetic, it is possible that a small error is introduced each time an arithmetic operation is performed. One should check the magnitude of all the elements in column i that lie on below the diagonal, and locate a row j in which the element has the largest absolute

Pivoting ii

value, that is $|a_{ji}| = \max\{|a_{ii}|, |a_{i+1\,i}|, \ldots, |a_{N-1\,i}|, |a_{N\,i}|\}$, and then switch row i with row j if $j > i$.

Usually, the larger pivot element will result in a smaller error being propagated.

EXAMPLE

Let's assume that after a number of operations the last two equations of an $N \times N$ system are:

$$\begin{aligned} 0x_{N-1} + x_N &= 1 \\ 2x_{N-1} + x_N &= 3 \end{aligned}$$

The obvious solution is $x_{N-1} = x_N = 1$.

But due to accumulation of numerical errors the system will be:

$$\begin{aligned} \epsilon x_{N-1} + x_N &= 1 \\ 2x_{N-1} + x_N &= 3 \end{aligned} \qquad \begin{aligned} \epsilon x_{N-1} + x_N &= 1 \\ 0 + \left(1 - \frac{2}{\epsilon}\right) x_N &= 3 - \frac{2}{\epsilon} \end{aligned}$$

Pivoting iii

Leading to

$$x_N = \frac{3 - 2/\epsilon}{1 - 2/\epsilon} \approx 1 \quad \text{correct, but} \quad x_{N-1} = \frac{1 - x_N}{\epsilon} \text{ !!!}$$

I.e. x_{N-1} is the ratio of two small numbers with questionable accuracy.

Then this erroneous term will be used for the estimation of $x_{N-2}, \ldots, x_1$ via eqns (6) with disastrous results.

With **pivoting**, the system will be written as:

$$\begin{aligned} 2x_{N-1} + x_N &= 3 \\ \epsilon x_{N-1} + x_N &= 1 \end{aligned} \qquad \begin{aligned} 2x_{N-1} + x_N &= 3 \\ \left(1 - \frac{\epsilon}{2}\right) x_N &= 1 - \frac{3\epsilon}{2} \end{aligned}$$

then $x_N \approx 1.0$ and $x_{N-1} = \frac{3 - x_N}{2} \approx 1.0$.

Gauss - Jordan Method i

It is a variation of Gauss's method. Here instead of transforming the original system into an upper-triangular one, we transform it into a pure diagonal one and then the solution is obvious. [1]

The method follows the steps of Gauss's method with parallel elimination of the elements over the diagonal.

Thus after the 1st step of Gaussian elimination we multiply the 2nd equation with $a_{12}/a_{22}^{(1)}$ and substract it from the 1st.

Thus the system becomes:

$$\begin{aligned}
a_{11}x_1 + \quad 0 + a_{13}^{(2)}x_3 + \ldots + a_{1N}^{(2)}x_N &= b_1^{(2)} \\
0 + a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + \ldots + a_{2N}^{(1)}x_N &= b_2^{(1)} \\
0 \quad + \quad 0 + a_{33}^{(2)}x_3 + \ldots + a_{3N}^{(2)}x_N &= b_3^{(2)} \\
&\vdots \\
0 \quad + \quad 0 + a_{N3}^{(2)}x_3 + \ldots + a_{NN}^{(2)}x_N &= b_N^{(2)}
\end{aligned} \tag{8}$$

Gauss - Jordan Method ii

and in the next step we eliminate the terms with coefficients $a_{13}^{(2)}$ and $a_{23}^{(1)}$.

Thus after $N-1$ steps we get the system

$$\begin{aligned} a_{11}x_1 &= b_1^{(N-1)} \\ a_{22}^{(1)}x_2 &= b_2^{(N-1)} \\ &\vdots \\ a_{NN}^{(N-1)}x_N &= b_N^{(N-1)} \end{aligned} \tag{9}$$

with obvious solution

$$x_i = \frac{b_i^{(N-1)}}{a_{ii}^{(i-1)}} \quad \text{for} \quad i = 1, \ldots, N \tag{10}$$

[1] Some special cases of the method - albeit presented without proof - were known to Chinese mathematicians as early as circa 179 AD. (Wikipedia)

The Jacobi Method (iterative) i

It is a generalization of the method $x = g(x)$ presented in the previous section. Let's assume the following system of N linear equations with N unknowns:

$$\begin{aligned} f_1(x_1, x_2, ..., x_N) &= 0 \\ f_2(x_1, x_2, ..., x_N) &= 0 \\ &..... \\ f_N(x_1, x_2, ..., x_N) &= 0 \end{aligned} \tag{11}$$

this can be easily written in the form:

$$\begin{aligned} x_1 &= g_1(x_2, x_3, ..., x_N) \\ x_2 &= g_2(x_1, x_3, ..., x_N) \\ &... \\ x_N &= g_N(x_1, x_2, ..., x_{N-1}) \end{aligned} \tag{12}$$

The Jacobi Method (iterative) ii

or in a close form:

$$x_i = \frac{b_i}{a_{ii}} - \frac{1}{a_{ii}} \sum_{j=1\,,\,j\neq i}^{N} a_{ij} x_j \quad \text{for} \quad i = 1, \ldots, N \tag{13}$$

Then by giving N initial values $x_1^{(0)}, x_2^{(0)}, \ldots, x_N^{(0)}$, we create the recurrence (with respect to k) relation

$$x_i^{(k+1)} = g_i(x_1^{(k)}, ..., x_N^{(k)}) \quad \text{for} \quad i = 1, \ldots, N \tag{14}$$

which will converge to the solution of the system if:

$$|a_{ii}| > \sum_{j=1\,,\,j\neq i}^{N} |a_{ij}| \tag{15}$$

independent on the choice of the initial values $x_1^{(0)}, x_2^{(0)}, \ldots, x_N^{(0)}$.

The Jacobi Method (iterative) iii

The recurrence relation can be written in a matrix form as:

$$\mathbf{x}^{(k+1)} = \mathbf{D}^{-1}\mathbf{B} - \mathbf{D}^{-1}\mathbf{C}\mathbf{x}^{(k)}$$

where $\mathbf{A} = \mathbf{D} + \mathbf{C}$ i.e. the matrix $\mathbf{D}$ includes only the diagonal elements of $\mathbf{A}$ and the matrix $\mathbf{C}$ all the rest.

EXAMPLE

Let's assume the system:

$$\begin{aligned} 4x - y + z &= 7 \\ -4x + 8y - z &= 21 \\ -2x + y + 5z &= 15 \end{aligned}$$

with solutions $x = 2$, $y = 4$, $z = 3$

The Jacobi Method (iterative) iv

We can create the recurrence relations:

$$\begin{aligned} x^{(k+1)} &= \frac{7 + y^{(k)} - z^{(k)}}{4} \\ y^{(k+1)} &= \frac{21 + 4x^{(k)} + z^{(k)}}{8} \\ z^{(k+1)} &= \frac{15 + 2x^{(k)} - y^{(k)}}{5} \end{aligned}$$

Then assuming $(1, 2, 2)$ we get the following sequence of solutions

$$\begin{aligned} (1,2,2) &\to (1.75, 3.375, 3) &&\to (1.844, 3.875, 3.025) \\ &\to (1.963, 3.925, 2.963) &&\to (1.991, 3.977, 3.0) \\ &\to (1.994, 3.995, 3.001) &&\to \cdots \end{aligned}$$

I.e. with **5** iterations we reached the solution with **3** digits accuracy.

Gauss - Seidel Method i

Gauss - Seidel method is a variation of Jacobi's method.

Let's assume a trial solution $x_1^{(0)}, x_2^{(0)}, \ldots, x_N^{(0)}$.

Then from the 1st equation we estimate $x_1^{(1)}$ and in the 2nd equation we use the following set of trial values $(x_1^{(1)}, x_2^{(0)}, x_3^{(0)}, ..., x_N^{(0)})$ to estimate $x_2^{(1)}$.

Then we use the trial values $(x_1^{(1)}, x_2^{(1)}, x_3^{(0)}, ..., x_N^{(0)})$ to estimate $x_3^{(1)}$ and so on. The recurrence relations will be:

$$
\begin{aligned}
x_1^{(k+1)} &= \frac{1}{a_{11}}\left(b_1 - \sum_{j=2}^{N} a_{1j}x_j^{(k)}\right) \\
x_2^{(k+1)} &= \frac{1}{a_{22}}\left(b_2 - a_{21}x_1^{(k+1)} - \sum_{j=3}^{N} a_{2j}x_j^{(k)}\right) \\
&\ldots \\
x_i^{(k+1)} &= \frac{1}{a_{ii}}\left(b_i - \sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{N} a_{ij}x_j^{(k)}\right) \qquad (16)
\end{aligned}
$$

Gauss - Seidel Method ii

The method will converge if:

$$|a_{ii}| > \sum_{j=1\,,\,j\neq i}^{N} |a_{ij}| \qquad i = 1, 2, \ldots, N \tag{17}$$

This procedure in a "matrix form" will be written as:

$$\mathbf{x}^{(\mathbf{k+1})} = \mathbf{D}^{-1}\left(\mathbf{B} - \mathbf{L}\mathbf{x}^{(\mathbf{k+1})} - \mathbf{U}\mathbf{x}^{(\mathbf{k})}\right) \tag{18}$$

where

$$\mathbf{A} = \underset{lower}{\mathbf{L}} + \underset{diagonal}{\mathbf{D}} + \underset{upper}{\mathbf{U}} \,.$$

The matrix **L** has the elements of below the diagonal **A**, the matrix **D** only the diagonal elements of **A** and finally the matrix **U** the elements of matrix **A** over the diagonal.

Gauss - Seidel Method iii

EXAMPLE

The recurrence relations for the previous example are the same i.e.

$$\begin{aligned} x^{(k+1)} &= \frac{7 + y^{(k)} - z^{(k)}}{4} \\ y^{(k+1)} &= \frac{21 + 4x^{(k+1)} + z^{(k)}}{8} \\ z^{(k+1)} &= \frac{15 + 2x^{(k+1)} - y^{(k+1)}}{5} \end{aligned}$$

But with Gauss - Seidel method we get the following sequence of approximate solutions:

$$\begin{aligned} (1, 2, 2) &\rightarrow (1.75, 3.75, 2, 95) \\ &\rightarrow (1.95, 3.97, 2.99) \\ &\rightarrow (1.996, 3.996, 2.999) \end{aligned}$$

i.e. here we need only **3** iterations to find the solution while with Jacobi's method we need 5 iterations.

Overrelaxation

The typical Gauss-Seidel iteration scheme has the following form:

$$x_i^{(k+1)} = \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{N} a_{ij} x_j^{(k)} \right) \tag{19}$$

we may re-write it as follows:

$$x_i^{(k+1)} = x_i^{(k)} + \frac{1}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i}^{N} a_{ij} x_j^{(k)} \right) \tag{20}$$

because $x_i^{(k)}$ is **added** and **subtracted** from the right side.

The procedure maybe optimized by using the correct overrelaxation factor

$$x_i^{(k+1)} = x_i^{(k)} + \frac{w}{a_{ii}} \left(b_i - \sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i}^{N} a_{ij} x_j^{(k)} \right) \tag{21}$$

Typically, $1 \leq w < 2$.

Inverse of a matrix i

If we are seeking the inverse of the matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

there will be another matrix

$$\mathbf{B} = \begin{pmatrix} x & y \\ z & w \end{pmatrix}$$

with the following property:

$$\mathbf{A} \cdot \mathbf{B} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{I}$$

In order to find the elements (x, y, z, w) of the inverse matrix we must solve two linear systems:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} x \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} y \\ w \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Inverse of a matrix ii

This means that the inverse of a matrix cannot help in the solution of a linear system since one needs to solve the system to find it.

- Notice that in order to solve both of the above systems we need to diagonalize (with the Gauss - Jordan procedure) the same matrix **A**.

- The only difference between the two systems is the vector on the right hand side which corresponds to different column of the unit matrix.

We then construct the following scheme:

$$\begin{array}{cccc|cccc} a_{11} & a_{12} & \dots & a_{1N} & 1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2N} & 0 & 1 & \dots & 0 \\ & & & & & & & \\ a_{N1} & a_{N2} & \dots & a_{NN} & 0 & 0 & \dots & 1 \\ \multicolumn{4}{c|}{\mathbf{A}} & \multicolumn{4}{c}{\mathbf{I}} \end{array} \tag{22}$$

Inverse of a matrix iii

and whatever transformation we do in the left side (matrix **A**) exactly the same we do on the right hand side (matrix **I**).

$$\left(\begin{array}{cccc|cccc} 1 & 0 & \dots & 0 & \tilde{a}_{11} & \tilde{a}_{12} & \dots & \tilde{a}_{1N} \\ 0 & 1 & \dots & 0 & \tilde{a}_{21} & \tilde{a}_{22} & \dots & \tilde{a}_{2N} \\ & & & & & & & \\ 0 & 0 & \dots & 1 & \tilde{a}_{N1} & \tilde{a}_{N2} & \dots & \tilde{a}_{NN} \end{array}\right) \tag{23}$$

Thanks to the Gauss-Jordan method on the left hand side we reduce to the unit matrix matrix **I** and on the right hand side on the inverse $\tilde{a}_{ij} = \mathbf{A}^{-1}$.

Eigenvalues and Eigenvectors i

If $\mathbf{A}$ is a $N \times N$ matrix then the scalars λ for which there exists a non-zero vector $\vec{u}$ such that

$$\mathbf{A} \cdot \vec{u} = \lambda \vec{u} \tag{24}$$

will be called **e-values** of the matrix $\mathbf{A}$ and the vectors $\vec{u}$ **e-vectors**.

In the following example:

$$\underbrace{\begin{pmatrix} 1 & 2 & 3 \\ -1 & 3 & 1 \\ 2 & 0 & 1 \end{pmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}}_{\vec{u}_1} = \underbrace{-1}_{\lambda_1} \cdot \underbrace{\begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}}_{\vec{u}_1}$$

the vector $\vec{u_1} = (2, 1, -2)^\top$ is the e-vector and $\lambda_1 = -1$ the corresponding e-value of $\mathbf{A}$.

Equation (24) is equivalent to:

$$\det(\mathbf{A} - \lambda \mathbf{I}) = 0$$

Eigenvalues and Eigenvectors ii

that is

$$\det \begin{vmatrix} 1-\lambda & 2 & 3 \\ -1 & 3-\lambda & 1 \\ 2 & 0 & 1-\lambda \end{vmatrix} = \lambda^3 - 5\lambda^2 + 3\lambda + 9 = 0 \qquad (25)$$

and the e-values will be the roots of the **characteristic polynomial** (here $\lambda_i = -1$, 3 and 3).

Eigenvalues and Eigenvectors : The power method i

For a matrix **A** there will always be a **dominant e-value** λ_1 (this means that λ_1 is absolutely larger than any other e-value)

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_N| \tag{26}$$

i.e.,

$$\mathbf{A}\vec{u}^{(i)} = \lambda_i \vec{u}^{(i)} \qquad (1 \leq i \leq N) \tag{27}$$

Also any vector $\vec{x}$ can be written as a linear combination of the N e-vectors $\left\{\vec{u}^{(1)}, \vec{u}^{(2)}, \ldots, \vec{u}^{(N)}\right\}$ i.e.,

$$\vec{x} = a_1\vec{u}^{(1)} + a_2\vec{u}^{(2)} + \cdots + a_N\vec{u}^{(N)} \tag{28}$$

If we multiply both sides of (28) with the matrix **A**, we get:

$$\mathbf{A}\vec{x} \equiv \vec{x}^{(1)} = a_1\lambda_1\vec{u}^{(1)} + a_2\lambda_2\vec{u}^{(2)} + \cdots + a_N\lambda_N\vec{u}^{(N)} \tag{29}$$

Eigenvalues and Eigenvectors : The power method ii

If we multiply k times eqn (29) with the matrix $\mathbf{A}$ we get:

$$\begin{aligned}\mathbf{A}^k\mathbf{x} \equiv \mathbf{x}^{(k)} &= a_1\lambda_1^k\mathbf{u}^{(1)} + a_2\lambda_2^k\mathbf{u}^{(2)} + \cdots + a_N\lambda_N^k\mathbf{u}^{(N)} \\ &= \lambda_1^k\left[a_1\mathbf{u}^{(1)} + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\mathbf{u}^{(2)} + \cdots + a_N\left(\frac{\lambda_N}{\lambda_1}\right)^k\mathbf{u}^{(N)}\right] \quad (30)\end{aligned}$$

Since λ_1 is the absolute larger e-value (eqn 26), we will get:

$$(\lambda_j/\lambda_1)^k \to 0 \quad \text{as} k \to \infty\,.$$

Thus

$$\mathbf{x}^{(k)} = \mathbf{A}^k \cdot \mathbf{x} \approx \lambda_1^k \cdot a_1 \cdot \mathbf{u}^{(1)} \quad (31)$$

then the ratio

$$\mathbf{r}^{(k)} \equiv \frac{\mathbf{x}^{(k+1)}}{\mathbf{x}^{(k)}} = \frac{\mathbf{A}^{k+1}\mathbf{x}}{\mathbf{A}^k\vec{x}} \approx \frac{\lambda_1^{k+1}a_1\mathbf{u}^{(1)}}{\lambda_1^k a_1\mathbf{u}^{(1)}} \to \lambda_1 \quad (32)$$

leads to the e-value λ_1 and the vector $\mathbf{x}^{(k)}$ defined in (31) is the corresponding e-vector.

EXAMPLE

The matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ -1 & 2 & 2 \\ 1 & 0 & 3 \end{pmatrix}$$

has e-values:

$\lambda_1 = 3.41421356$, $\lambda_2 = 2$ & $\lambda_3 = 0.585786$

and corresponding e-vectors :

$$\mathbf{u}^{(1)} = \begin{pmatrix} 0.3694 \\ 1 \\ 0.8918 \end{pmatrix} \quad , \quad \mathbf{u}^{(2)} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \& \quad \mathbf{u}^{(2)} = \begin{pmatrix} 0.7735 \\ 1 \\ -0.3204 \end{pmatrix}$$

Then if we assume a trial vector $\mathbf{x} = (1, 2, 1)^\top$, and multiply it with the 5th and 6th power of $\mathbf{A}$

$$\mathbf{A}^5 = \begin{pmatrix} 68 & 0 & 164 \\ 136 & 32 & 428 \\ 164 & 0 & 396 \end{pmatrix} \quad \mathbf{A}^6 = \begin{pmatrix} 232 & 0 & 560 \\ 532 & 64 & 1484 \\ 560 & 0 & 1352 \end{pmatrix}$$

we get : $\mathbf{x}^{(5)} = \mathbf{A}^5\mathbf{x} = \begin{pmatrix} 232 \\ 628 \\ 560 \end{pmatrix}$ & $\mathbf{x}^{(6)} = \mathbf{A}^6\mathbf{x} = \begin{pmatrix} 792 \\ 2144 \\ 1912 \end{pmatrix}$

This leads to

$$\lambda_1 \approx \frac{\mathbf{x}^{(6)}}{\mathbf{x}^{(5)}} = \frac{792}{232} \approx \frac{2144}{628} \approx \frac{1912}{560} \approx 3.414286$$

and the e-vector $\mathbf{u}^{(1)} \approx \mathbf{x}^{(6)} \approx \begin{pmatrix} 0.3694 \\ 1 \\ 0.8918 \end{pmatrix}$

NOTE:
It is numerically faster to construct the vectors $\mathbf{x}^{(k+1)}$ by multiplying the matrix $\mathbf{A}$ with the vector $\mathbf{x}^{(k)}$ i.e. $\mathbf{x}^{(k+1)} = \mathbf{A}\mathbf{x}^{(k)}$ instead of calculating the kth power of $\mathbf{A}$.

Eigenvalues and Eigenvectors : Aitken Acceleration

The elements of the vector $\mathbf{r}^{(k)}$ in eqn (32) are approximations to the exact e-value λ_1 and the error will be $\epsilon_k = |\mathbf{r}_k - \lambda_1|$ where $\mathbf{r}$ is any element of $\mathbf{r}^{(k)}$. Obviously for $\mathbf{r}^{(k+1)}$ we get $\epsilon_{k+1} = |\mathbf{r}_{k+1} - \lambda_1|$. Since the convergence of the method is linear:, i.e.

$$\epsilon_{k+1} = A\,\epsilon_k \tag{33}$$

we can use Aitken's method to accelerate the convergence. That is, we will use the formula derived in the previous section for 3 approximate values: $\mathbf{r}_k$, $\mathbf{r}_{k+1}$ and $\mathbf{r}_{k+2}$ to find a better approximation to the e-value i.e.

$$\lambda_1 \approx \frac{\mathbf{r}_k\mathbf{r}_{k+2} - \mathbf{r}_{k+1}^2}{\mathbf{r}_{k+2} - 2\mathbf{r}_{k+1} + \mathbf{r}_k}\,. \tag{34}$$

Application:

From the example of the previous slide $\mathbf{r}_5 = 3.414$, $\mathbf{r}_4 = 3.413$ and $\mathbf{r}_3 = 3.407$. Then Aitken's method gives $\lambda_1 \approx 3.4142$.

Eigenvalues and Eigenvectors : Inverse Power Method

Theorem: If λ is an e-value of a matrix $\mathbf{A}$, then λ^{-1} is an e-value of the $\mathbf{A}^{-1}$.

Thus if $|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_{N-1}| > |\lambda_N| > 0$, are the e-values of $\mathbf{A}$ then

$$\left|\lambda_N^{-1}\right| > \left|\lambda_{N-1}^{-1}\right| \geq \cdots \geq \left|\lambda_1^{-1}\right| > 0 \tag{35}$$

will be the e-values of $\mathbf{A}^{-1}$. Thus λ_N^{-1} is the largest e-value of $\mathbf{A}^{-1}$.

To find the e-value instead of calculating the inverse of the matrix $\mathbf{A}$ to get the expression $\left(\mathbf{A}^{-1}\right)^{(k+1)}\mathbf{x}$ we can solve by using Gaussian elimination the system $\mathbf{A}\cdot\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$, where $\mathbf{x}^{(k)}$ is defined as $\mathbf{x}^{(k)} = \mathbf{A}^k\mathbf{x}$.

That is for a initial $\mathbf{x}^{(0)}$, we get $\mathbf{x}^{(1)} = \mathbf{A}^{-1}\mathbf{x}^{(0)}$ i.e. $\mathbf{A}\mathbf{x}^{(1)} = \mathbf{x}^{(0)}$. Thus the solution of the linear system defines the value of $\mathbf{x}^{(1)}$.

With this procedure we get the vector $\mathbf{x}^{(k+1)}$ and λ_N^{-1} via

$$\mathbf{x}^{(k+1)} = \mathbf{A}^{-1}\mathbf{x}^{(k)} \quad \Rightarrow \quad \mathbf{A}\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} \tag{36}$$

Eigenvalues and Eigenvectors : Shifting Power Method

Theorem: If the N values λ_i with $(i = 1, \ldots, N)$ are the e-values of a $N \times N$ matrix $\mathbf{A}$, then for any complex number μ the matrix $\mathbf{A} - \mu\mathbf{I}$ (where $\mathbf{I}$ is the unit matrix) will have as e-values the complex numbers $(\lambda_i - \mu)$ for $(i = 1, \ldots N)$.

Thus if in the interval (λ_N, λ_1) we consider a point μ such that for an e-value λ_k to hold $0 < |\lambda_k - \mu| < \epsilon$ while for all the rest we get $|\lambda_i - \mu| > \epsilon$; then $|\lambda_k - \mu|$ will be the minimum e-value of $\mathbf{A} - \mu\mathbf{I}$.

This means that we can estimate it by using the *inverse power method*

Remember that you need to calculate $\mathbf{x}^{(k+1)}$ by solving the system:

$$(\mathbf{A} - \mu\mathbf{I})\,\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)}$$

Thus if we calculate the r_k, the actual e-value λ_i will be:

$$\lambda_i = \frac{1}{r_k} + \mu\,. \tag{37}$$