EIGENVALUE PROBLEMS

A fixed point for an $n \times n$ matrix $\mathbf{A}$ is a vector $\mathbf{x}$ in R^n such that:

$$\mathbf{A}\mathbf{x} = \mathbf{x}$$

Every square matrix $\mathbf{A}$ has at least one fixed point, $\mathbf{x} = \mathbf{0}$, called the trivial fixed point of $\mathbf{A}$.

The general procedure to find fixed points is to rewrite the equation above as an homogeneous linear system:

$$\mathbf{A}\mathbf{x} = \mathbf{x} \implies \mathbf{A}\mathbf{x} = \mathbf{I}\mathbf{x} \implies (\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$$

The following theorem is useful in ascertaining whether a matrix $\mathbf{A}$ has non-trivial fixed points

Theorem

If $\mathbf{A}$ is an $n \times n$ matrix, then the following statements are equivalent:

a) $\mathbf{A}$ has non-trivial fixed points
b) $\mathbf{I} - \mathbf{A}$ is singular
c) $\det(\mathbf{I} - \mathbf{A}) = 0$

Example 1

$$\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix}$$ has only the trivial fixed point because $$\det(\mathbf{I} - \mathbf{A}) = \begin{vmatrix} -2 & -6 \\ -1 & -1 \end{vmatrix} = -4 \neq 0$$

Example 2

$$\mathbf{B} = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix}$$ has no-trivial fixed points because $$\det(\mathbf{I} - \mathbf{B}) = \begin{vmatrix} 1 & -2 \\ 0 & 0 \end{vmatrix} = 0$$

To find them, we must solve the linear system $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$:

$$\begin{bmatrix} 1 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \mathbf{x} = \begin{bmatrix} 2t \\ t \end{bmatrix}$$ for any real value t

A generalization of the fixed point problem for an $n \times n$ matrix $\mathbf{A}$ is to find a vector $\mathbf{x}$ in R^n such that:

$$\mathbf{Ax} = \lambda\mathbf{x}$$

where λ is an arbitrary scalar (real or complex).

A scalar λ is called an eigenvalue of $\mathbf{A}$ if there is a nonzero vector $\mathbf{x}$ such that $\mathbf{Ax}=\lambda\mathbf{x}$. If λ is an eigenvalue of $\mathbf{A}$, then every nonzero vector $\mathbf{x}$ for which $\mathbf{Ax}=\lambda\mathbf{x}$ is called an eigenvector of $\mathbf{A}$ corresponding to λ.

The most direct procedure to find eigenvalues and eigenvectors of a matrix $\mathbf{A}$ is to rewrite the equation above as :

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

and try to determine for which values of λ, if any, this system has non-trivial solutions.

The linear system:

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

has non-trivial solutions when:

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

and this is called the characteristic equation of $\mathbf{A}$.

Also, if λ is an eigenvalue of $\mathbf{A}$, then the linear system above has a nonzero solution space which we call the eigenspace of $\mathbf{A}$ associated with λ.

Since

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_n$$

if we admit complex solutions there are n solutions of the characteristic equation. The set $\{\lambda_1 \ldots \lambda_n\}$ is called the spectrum of $\mathbf{A}$ and the magnitude of the largest eigenvalue is called the spectral radius $\lambda(\mathbf{A})$ of $\mathbf{A}$.

To find the eigenvalues and eigenvectors of

$$\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 1 & 2 \end{bmatrix}$$

we first solve the characteristic equation

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \begin{vmatrix} \lambda - 1 & -3 \\ -4 & \lambda - 2 \end{vmatrix} = \lambda^2 - 3\lambda - 10 = 0 \quad \Rightarrow \quad \begin{matrix} \lambda_1 = -2 \\ \lambda_2 = 5 \end{matrix}$$

and then solve the system

$$\begin{bmatrix} \lambda - 1 & -3 \\ -4 & \lambda - 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

for each of the two eigenvalues λ_1 and λ_2.

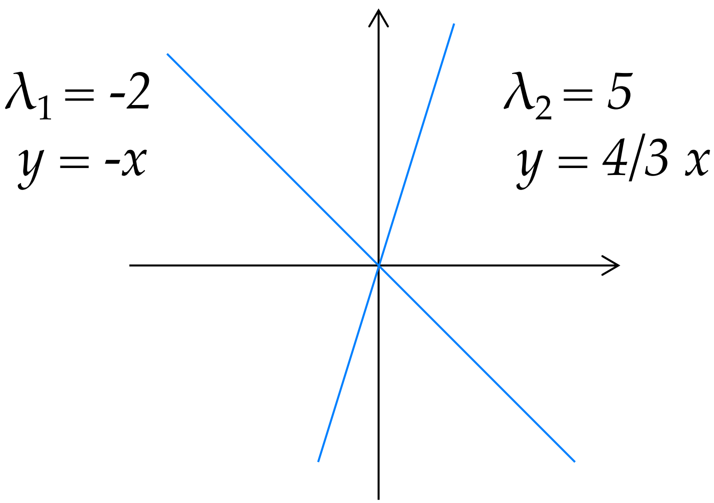
For $\lambda_1 = -2$

$$\begin{bmatrix} -3 & -3 \\ -4 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -t \\ t \end{bmatrix}$$

and for $\lambda_2 = 5$

$$\begin{bmatrix} 4 & -3 \\ -4 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{3}{4}t \\ t \end{bmatrix}$$

Each eigenspace is a line in R^2



The characteristic polynomial of a general 2×2 matrix with real entries

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

is

$$\det(\lambda\mathbf{I}-\mathbf{A}) = \begin{vmatrix} \lambda-a & -b \\ -c & \lambda-d \end{vmatrix} = (\lambda-a)(\lambda-b)-bc = \lambda^2-(a+d)\lambda+(ad-bc)$$

which can be written as

$$\det(\lambda\mathbf{I}-\mathbf{A}) = \lambda^2 - Tr(\mathbf{A})\lambda + \det(\mathbf{A})$$

For the characteristic equation we find

$$\lambda^2 - Tr(\mathbf{A})\lambda + \det(\mathbf{A}) = 0 \quad \Longrightarrow$$

2 different real λ if $Tr(\mathbf{A})^2 + 4\det(\mathbf{A}) > 0$

one repeated real λ if $Tr(\mathbf{A})^2 + 4\det(\mathbf{A}) = 0$

2 conjugate complex λ if $Tr(\mathbf{A})^2 + 4\det(\mathbf{A}) < 0$

The characteristic polynomial of a general symmetric 2×2 matrix with real entries

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

is

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \lambda^2 - (a+d)\lambda + (ad - b^2)$$

for the discriminant of the characteristic equation we find

$$Tr(\mathbf{A})^2 + 4\det(\mathbf{A}) = (a-d)^2 - 4(ad - b^2) = (a-d)^2 + 4b^2 \geq 0$$

and we have that the two eigenvalues must be real. In general, they are different and their corresponding eigenspaces are perpendicular lines through the origin. Only when $a = d$ and $b = 0$ we have two repeated eigenvalues:

$$\mathbf{A} = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix} \Longrightarrow Tr(\mathbf{A})^2 + 4\det(\mathbf{A}) = 0 \Longrightarrow \begin{aligned} \lambda_1 &= a \\ \lambda_2 &= a \end{aligned}$$

If $\mathbf{A}$ is an $n \times n$ matrix with eigenvalues $\lambda_1 \ldots \lambda_n$ (repeated according to multiplicity) then

$$Tr(\mathbf{A}) = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

and

$$\det(\mathbf{A}) = \lambda_1 \lambda_2 \cdots \lambda_n$$

for the discriminant of the characteristic equation we find

Since for a triangular matrix (upper triangular, lower triangular, or diagonal) we have

$$Tr(\mathbf{T}) = t_{11} + t_{22} + \cdots + t_{nn}$$

and

$$\det(\mathbf{T}) = t_{11} t_{22} \cdots t_{nn}$$

$\Rightarrow$ the elements in the diagonal of any triangular matrix coincide with its eigenvalues

Eigenvalues of an $n \times n$ matrix $\mathbf{A}$ are rarely obtained by solving the characteristic equation in real-world applications primarily for two reasons

- In order to construct the characteristic equation we need to expand an $n \times n$ determinant and this task is computationally prohibitive (it involves $n!$ operations) for large matrices.

- There is no algebraic formula or finite algorithm that can be used to obtain exact solutions of polynomial equations with degree $n \geq 5$.

For these reasons numerical methods to solve eigenvalue problems use alternative strategies based on the relations between traces and determinants of matrices and their eigenvalues.

If $\mathbf{A}$ is an $n \times n$ matrix with eigenvalues $\lambda_1 \ldots \lambda_n$ such that

$$|\lambda_1| > |\lambda_2| > \cdots > |\lambda_n|$$

then λ_1 is called the dominant eigenvalue of $\mathbf{A}$ and its associated eigen-vector the dominant eigenvector and both can be found iteratively.
The following theorem applies for matrices with linearly independent eigenvectors (symmetric matrices have this property):

Theorem
Let $\mathbf{A}$ be a symmetric $n \times n$ matrix with a positive dominant eigenvalue λ. If $\mathbf{x}_0$ is a unit vector in R^n that is not orthogonal to the eigenspace correspond-ing to λ then the normalized power sequence

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \frac{\mathbf{A}\mathbf{x}_0}{\left\|\mathbf{A}\mathbf{x}_0\right\|}, \quad \mathbf{x}_2 = \frac{\mathbf{A}^2\mathbf{x}_0}{\left\|\mathbf{A}^2\mathbf{x}_0\right\|}, \ \cdots \ , \quad \mathbf{x}_k = \frac{\mathbf{A}^k\mathbf{x}_0}{\left\|\mathbf{A}^k\mathbf{x}_0\right\|}, \ \cdots$$

converges to a unit dominant eigenvector and the sequence

$$(\mathbf{A}\mathbf{x}_1)\cdot\mathbf{x}_1, (\mathbf{A}\mathbf{x}_2)\cdot\mathbf{x}_2, \cdots, (\mathbf{A}\mathbf{x}_k)\cdot\mathbf{x}_k, \cdots$$

converges to the dominant eigenvalue λ.

To find the dominant eigenvalue of a symmetric $n \times n$ matrix $\mathbf{A}$ follow these steps:

Step 1

Choose an arbitrary nonzero vector and normalize it (if needed) to obtain a unit vector $\mathbf{x}_0$.

Step 2

Compute $\mathbf{x}_1 = \mathbf{A}\mathbf{x}_0$ and $\mathbf{A}\mathbf{x}_1 \cdot \mathbf{x}_1$ to approximate the dominant eigenvector and eigenvalue, respectively.

Step 3 …

Compute $\mathbf{x}_2 = \mathbf{A}\mathbf{x}_1$ and normalize it to obtain the second approximation to the dominant eigenvector. Compute $\mathbf{A}\mathbf{x}_2 \cdot \mathbf{x}_2$ to approximate the dominant eigenvalue.

Stop the process when the difference between the eigenvalue approximations in two successive steps is smaller than a given threshold ε.

1) Power method with Euclidean scaling

Write a program to find the dominant eigenvalue and a unit dominant eigenvector for a symmetric $n \times n$ matrix.

Suggestion:

Try it for the following matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 2 \\ 2 & 3 \end{bmatrix}$$

with

$$\mathbf{x_o} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

A general strategy to obtain eigenvalues and eigenvectors for a given $n \times n$ matrix $\mathbf{A}$ is to transform into another one $\mathbf{B}$ that has the same spectrum and eigenvectors, but for which it is easy to obtain these, for instance, triangular or diagonal matrices.

In this respect, the concept of similarity transformation is essential.

Two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be similar (or related by a similarity transformation) if there exists an invertible matrix $\mathbf{S}$ such that:

$$\mathbf{A} = \mathbf{S}^{-1}\mathbf{B}\mathbf{S}$$

Since

$$\mathbf{B} = \mathbf{S}\mathbf{A}\mathbf{S}^{-1} = (\mathbf{S}^{-1})^{-1}\mathbf{A}(\mathbf{S}^{-1}) = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$$

it is evident that if $\mathbf{A}$ is similar to $\mathbf{B}$, then $\mathbf{B}$ is similar to $\mathbf{A}$.

There are a number of basic properties shared by similar matrices:

- Similar matrices have the same determinant
- Similar matrices have the same trace
- Similar matrices have the same characteristic polynomial

Proof that if $\mathbf{A}$ is similar to $\mathbf{B}$, they have the same characteristic polynomial:

$$\lambda\mathbf{I}-\mathbf{B}=\lambda\mathbf{I}-\mathbf{S}^{-1}\mathbf{A}\mathbf{S}=\lambda\mathbf{S}^{-1}\mathbf{S}-\mathbf{S}^{-1}\mathbf{A}\mathbf{S}=\mathbf{S}^{-1}(\lambda\mathbf{S}-\mathbf{A}\mathbf{S})=\mathbf{S}^{-1}(\lambda\mathbf{I}\mathbf{S}-\mathbf{A}\mathbf{S})$$

$$=\mathbf{S}^{-1}(\lambda\mathbf{I}-\mathbf{A})\mathbf{S}$$

which shows that if $\mathbf{A}$ is similar to $\mathbf{B}$ then $\lambda\mathbf{I}$-$\mathbf{A}$ and $\lambda\mathbf{I}$-$\mathbf{B}$ are similar and

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

Since similar matrices have the same characteristic polynomial:

Theorem
Similar matrices have the same eigenvalues and those eigenvalues have the same algebraic and geometric multiplicities for both matrices

Algebraic multiplicity: how many times is a given eigenvalue λ repeated
Geometric multiplicity: how many independent eigenvectors are associated to a given eigenvalue

The eigenvectors are, however, not necessarily the same:

Theorem
Suppose $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$ and λ is an eigenvalue of $\mathbf{A}$ and $\mathbf{B}$ then:

If $\mathbf{x}$ is an eigenvector of $\mathbf{B}$ corresponding to λ, then $\mathbf{P}\mathbf{x}$ is an eigenvector of $\mathbf{A}$ corresponding to λ, and if $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ corresponding to λ, then $\mathbf{P}^{-1}\mathbf{x}$ is an eigenvector of $\mathbf{B}$ corresponding to λ.

An effective way to find eigenvectors and eigenvalues for a given square matrix is to find a similar diagonal matrix (if it exists)

Diagonalization problem

Given a $n \times n$ matrix $\mathbf{A}$ find an invertible transformation matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}^{-1}\mathbf{D}\mathbf{P}$, where $\mathbf{D}$ is a diagonal matrix.

$$\mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_1 \end{bmatrix} \qquad \mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{2n} \\ \vdots & \vdots & & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix} = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix}$$

eigenvalues in diagonal

eigenvectors in columns

Attention: not all square matrices are diagonalizable!

Diagonalization of real symmetric matrices

A real symmetric matrix $\mathbf{A}$ can be diagonalized by an orthogonal transformation

$$\mathbf{A} = \mathbf{O}^T \mathbf{D} \mathbf{O} \quad \text{where} \quad \mathbf{O}^{-1} = \mathbf{O}^T$$

Diagonalization of hermitian matrices

A hermitian matrix $\mathbf{A}$ can be diagonalized by an unitary transformation

$$\text{if} \quad \mathbf{A} = \mathbf{A}^\dagger \quad \text{then} \quad \mathbf{A} = \mathbf{U}^\dagger \mathbf{D} \mathbf{U} \quad \text{with} \quad \mathbf{U}^{-1} = \mathbf{U}^\dagger$$

Jacobi method

A simple method to diagonalize real symmetric matrices $\mathbf{A}$ is to apply a sequence of orthogonal transformations (plane rotations) to successively anihilate off-diagonal until the matrix becomes diagonal:

$$\mathbf{A} \to \mathbf{O}_1^T \mathbf{A} \mathbf{O}_1 \to \mathbf{O}_2^T (\mathbf{O}_1^T \mathbf{A} \mathbf{O}_1) \mathbf{O}_2 \to \ldots \to \mathbf{D} = \mathbf{P}^T \mathbf{A} \mathbf{P}$$

where

$$\mathbf{P}^T = \mathbf{O}_n^T \cdots \mathbf{O}_2^T \mathbf{O}_1^T$$

$$\mathbf{P} = \mathbf{O}_1 \mathbf{O}_2 \cdots \mathbf{O}_n$$

The Jacobi method allows us to diagonalize a real symmetric 2×2 matrix in a single step:

$$\mathbf{O}_1^T \mathbf{A} \mathbf{O}_1 = \mathbf{D}$$

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ 0 & \lambda_{22} \end{bmatrix}$$

Developing the matrix product and equating the terms out of the diagonal to 0 to get a diagonal matrix one arrives to

$$\tan(2\theta) = \frac{2a_{12}}{a_{11} - a_{22}}$$

When programming the Jacobi method one must take care of the possibility $a_{11} = a_{22}$ and treat it separately

The generalization of the Jacobi method to diagonalize a real symmetric $n \times n$ matrix uses the following type of matrices:

$$
\mathbf{O}_{pq} = \begin{array}{cc}
\begin{bmatrix}
1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\
\vdots & \ddots & \vdots & & \vdots & & \vdots \\
0 & \cdots & \cos\theta & \cdots & -\sin\theta & \cdots & 0 \\
\vdots & & \vdots & \ddots & \vdots & & \vdots \\
0 & \cdots & \sin\theta & \cdots & \cos\theta & \cdots & 0 \\
\vdots & & \vdots & & \vdots & \ddots & \vdots \\
0 & \cdots & 0 & \cdots & 0 & \cdots & 1
\end{bmatrix}
&
\begin{array}{l}
\\ \\ \text{row } p \\ \\ \text{row } q \\ \\ \\
\end{array}
\\
\begin{array}{ccccccc}
 & & \text{column } p & & \text{column } q & &
\end{array}
&
\end{array}
$$

with

$$
\tan(2\theta) = \frac{2a_{pq}}{a_{pp} - a_{qq}}
$$

to successively anihilate pairs of non diagonal elements in rows (columns) p and q.

The operation

$$\mathbf{O}_{pq}^{T}\mathbf{A}\mathbf{O}_{pq} = \mathbf{B}$$

leads to an $n \times n$ matrix with $b_{pq} = b_{qp} = 0$. In order to fully diagonalize the matrix we must apply successive orthogonal transformations until all non-diagonal elements become 0 (actually smaller than a given threshold):

$$\mathbf{A} \rightarrow \mathbf{O}_1^T\mathbf{A}\mathbf{O}_1 \rightarrow \mathbf{O}_2^T(\mathbf{O}_1^T\mathbf{A}\mathbf{O}_1)\mathbf{O}_2 \rightarrow \ldots \rightarrow \mathbf{D} = \mathbf{P}^T\mathbf{A}\mathbf{P}$$

where

$$\mathbf{P}^T = \mathbf{O}_n^T \cdots \mathbf{O}_2^T\mathbf{O}_1^T$$

$$\mathbf{P} = \mathbf{O}_1\mathbf{O}_2 \cdots \mathbf{O}_n$$

The general strategy is to sweep through all non-diagonal elements, pick the largest one (in absolute value) and apply the transformation, repeating this procedure until the largest one is below the threshold, where the process is stopped. Although a given rotation anihilates two non-diagonal elements, the following rotation may change this, but it can be shown that non-diagonal elements become progressively smaller.

1) Jacobi method for 2×2 symmetric matrices

Write a program to diagonalize a symmetric 2×2 matrix. Your program must indicate at the end which are the eigenvalues and the associated eigenvectors of your matrix.

2) Jacobi method for $n \times n$ symmetric matrices

Write a general program to diagonalize a symmetric $n \times n$ matrix. Your program must indicate at the end which are the eigenvalues and the associated eigenvectors of your matrix.

Howard Anton, Robert C. Busby
Contemporary Linear Algebra, Wiley, Hoboken, NJ (2003), Chapters 4 and 8

Gene H. Golub, Charles F. Van Loan
Matrix Computations, 3rd Ed., The Johns Hopkins University Press, Baltimore (1996), Chapters 7 and 8