I. Linear Algebra

I.5. Eigenvalues and Eigenvectors

Lecture based on

https://github.com/gwthomas/math4ml (Garrett Thomas, 2018)

https://mml-book.github.io/ (Deisenroth et al., 2020, Mathematics for Machine Learning)

Prof. Dr. Christoph Lippert

Digital Health & Machine Learning

If $\mathbf{A} \in \mathbb{R}^{M \times N}$, its **transpose** $\mathbf{A}^\top \in \mathbb{R}^{n \times m}$ is given by $(\mathbf{A}^\top)_{ij} = A_{ji}$ for each (i, j).

The transpose has several algebraic properties:

i $(\mathbf{A}^\top)^\top = \mathbf{A}$

ii $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$

iii $(\alpha \mathbf{A})^\top = \alpha \mathbf{A}^\top$

iv $(\mathbf{A}\mathbf{B})^\top = \mathbf{B}^\top \mathbf{A}^\top$

Consider N-dimensional vectors that are formed as a list of N scalars, such as the three-dimensional vectors

$$\mathbf{x} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \quad \text{and} \quad \mathbf{y} = \begin{bmatrix} -20 \\ -60 \\ -80 \end{bmatrix}.$$

These vectors are said to be **scalar multiples** of each other, or **parallel** or **collinear**, if there is a scalar λ such that

$$\mathbf{x} = \lambda \mathbf{y}.$$

Consider the linear transformation of N-dimensional vectors defined by an N by N matrix $\mathbf{A}$,

$$\mathbf{Av} = \mathbf{w}$$

or

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$

where, for each row,

$$w_i = A_{i1}v_1 + A_{i2}v_2 + \cdots + A_{in}v_n = \sum_{j=1}^{n} A_{ij}v_j$$

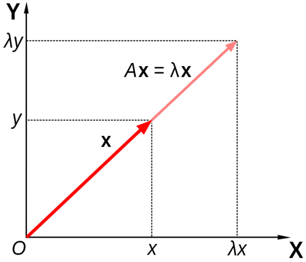
.

If it occurs that $\mathbf{v}$ and $\mathbf{w}$ are scalar multiples, that is if

$$\mathbf{Av} = \mathbf{w} = \lambda\mathbf{v},$$

then $\mathbf{v}$ is an **eigenvector** of the linear transformation $\mathbf{A}$ and the scale factor λ is the **eigenvalue** corresponding to that eigenvector.
This can be stated equivalently as

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



The zero vector is excluded from this definition because $\mathbf{A0} = \mathbf{0} = \lambda\mathbf{0}$ for every λ.

We now give some useful results about how eigenvalues change after various manipulations.

Proposition

Let $\mathbf{x}$ be an eigenvector of $\mathbf{A}$ with corresponding eigenvalue λ. Then

- (i) *For any $\gamma \in \mathbb{R}$, $\mathbf{x}$ is an eigenvector of $\mathbf{A} + \gamma\mathbf{I}$ with eigenvalue $\lambda + \gamma$.*
- (ii) *If $\mathbf{A}$ is invertible, then $\mathbf{x}$ is an eigenvector of $\mathbf{A}^{-1}$ with eigenvalue λ^{-1}.*
- (iii) *$\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$ for any $k \in \mathbb{Z}$ (where $\mathbf{A}^0 = \mathbf{I}$ by definition).*

Proof.

(i) follows readily:

$$(\mathbf{A} + \gamma\mathbf{I})\mathbf{x} = \mathbf{A}\mathbf{x} + \gamma\mathbf{I}\mathbf{x} = \lambda\mathbf{x} + \gamma\mathbf{x} = (\lambda + \gamma)\mathbf{x}$$

(ii) Suppose $\mathbf{A}$ is invertible. Then

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

Dividing by λ, which is valid because the invertibility of $\mathbf{A}$ implies $\lambda \neq 0$, gives $\lambda^{-1}\mathbf{x} = \mathbf{A}^{-1}\mathbf{x}$.
(iii) The case $k \geq 0$ follows immediately by induction on k. Then the general case $k \in \mathbb{Z}$ follows by combining the $k \geq 0$ case with (ii). □

The **determinant** of a square matrix can be defined in several different confusing ways.[1]

But it's good to know the properties:

1. $\det(\mathbf{I}) = 1$
2. $\det\left(\mathbf{A}^\top\right) = \det(\mathbf{A})$
3. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$
4. $\det\left(\mathbf{A}^{-1}\right) = \det(\mathbf{A})^{-1}$
5. $\det(\alpha\mathbf{A}) = \alpha^n \det(\mathbf{A})$

[1] Go look at an introductory linear algebra text (or Wikipedia) if you need a definition.

The determinant of a matrix is equal to the product of its eigenvalues (repeated according to multiplicity):

$$\det(\mathbf{A}) = \prod_i \lambda_i(\mathbf{A})$$

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

has a non-zero solution $\mathbf{v}$ if and only if the **determinant** of $(\mathbf{A} - \lambda\mathbf{I})$ is 0.

Therefore, the eigenvalues of $\mathbf{A}$ are values of λ that satisfy

$$|\mathbf{A} - \lambda\mathbf{I}| = 0$$

Example

$$\mathbf{M} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Taking the determinant of $(\mathbf{M} - \lambda\mathbf{I})$, the characteristic polynomial of $\mathbf{M}$ is

$$|\mathbf{M} - \lambda\mathbf{I}| = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = (2-\lambda)(2-\lambda) - (1)(1) = 3 - 4\lambda + \lambda^2.$$

Setting the characteristic polynomial equal to zero, it has roots at $\lambda = 1$ and $\lambda = 3$, which are the two eigenvalues of $\mathbf{M}$.
The eigenvectors corresponding to each eigenvalue can be found by solving for the components of $\mathbf{v}$ in the equation $\mathbf{M}\mathbf{v} = \lambda\mathbf{v}$.
In this example, the eigenvectors are any non-zero scalar multiples of

$$\mathbf{v}_{\lambda=1} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad \mathbf{v}_{\lambda=3} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Example

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$\begin{vmatrix} 1-\lambda & 1 \\ 0 & 1-\lambda \end{vmatrix} = (1-\lambda)(1-\lambda) - (0)(1) = 0$$

Only a single solution $\lambda = 1$.

Example

$$\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$\begin{vmatrix} 2-\lambda & 0 \\ 0 & 2-\lambda \end{vmatrix} = (2-\lambda)(2-\lambda) = 0$$

Only a single solution $\lambda = 2$.
Every vector is an eigenvector

If the entries of the matrix A are all real numbers, then the coefficients of the characteristic polynomial will also be real numbers, but the eigenvalues may still have non-zero imaginary parts.

Example

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

$$\begin{vmatrix} -\lambda & -1 \\ 1 & -\lambda \end{vmatrix} = (-\lambda)(-\lambda) - (-1)(1) = \lambda^2 + 1 = 0$$

Solutions are $\lambda = i$ or $\lambda = -i$.

The entries of the corresponding eigenvectors therefore may also have non-zero imaginary parts.

Let $\mathbf{A} \in \mathbb{R}^{N \times N}$. Then

$$\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{-1}$$

is called the **eigendecomposition** of $\mathbf{A}$, where $\mathbf{V}$ is the matrix of all eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda}$ is a diagonal matrix holding all the eigenvalues of $\mathbf{A}$.

Equivalently, $\mathbf{A}$ can be diagonalized as

$$\mathbf{V}^{-1}\mathbf{A}\mathbf{V} = \mathbf{\Lambda}$$