Eigenvalues, Eigenvectors (CDT-28)

Preprint · April 2020

DOI: 10.13140/RG.2.2.11761.22880/2

1 author:

Luciano da F. Costa
University of São Paulo
734 PUBLICATIONS **13,156** CITATIONS

SEE PROFILE

Eigenvalues, Eigenvectors (CDT-28)

Luciano da Fontoura Costa
luciano@ifsc.usp.br

São Carlos Institute of Physics – DFCM/USP

14th Apr 2020

Abstract

The concept and properties of eigenvalues and eigenvectors are presented in a concise and introductory manner. The importance of eigenvalues and eigenvectors in several areas is also briefly illustrated with respect to characterization of scalar field extrema, dynamical systems, Markov chains, and multivariate statistics.

"Know thyself."

Delphic proberb.

1 Introduction

Eigenvalues, eigenvectors. These forever companion terms, appear recurrently in many areas of science and technology. Among their many applications, we have the inference of the type of extrema (e.g. maximum, minimum, saddle-point, etc.) of a multivariate function, characterization of the stability and convergence in dynamical systems, study of physical properties such as axes of inertia, characterization of human faces, study of neuronal network dynamics, development of important statistical methods (such as principal component analysis), detection of communities in network research, study of Markov processes... The list, which seems endless, continues to grow, helped by the fact that eigenvalues and eigenvectors can relate not only to matrices, but also to functions, dynamical states, graphs, etc.

Though eigenvalues and eigenvectors (e.g. [1, 2, 3]) are frequently used in several scientific and technological areas, they are not always well-understood conceptually. Yet, they are ultimately intuitive and relatively simple, provided they are approached from the perspective of their properties, such as in preserving the inclination of vectors in linear transformation (see Figure 1).

The present work aims at providing a concise introduction to the concept of eigenvalues and eigenvectors, also covering some of their most important and useful properties, and including a few applications examples. We aimed at addressing applications from the eigenvector and eigenvalue perspective, in the sense that the respective concept and some important properties are presented as a preparation for discussing some of the many representative applications of eigenvalues and eigenvectors.

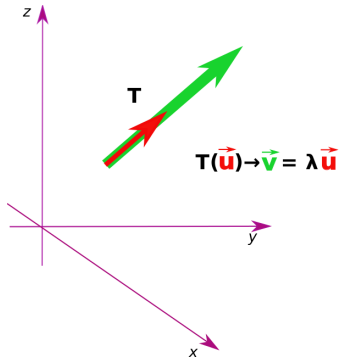

Figure 1: An eigenvector $\vec{u}$ of a linear transformation T will not be modified by it other than by a scaling by λ, its respectively associated eigenvalue.

Before proceeding, it is interesting to consider the origin of these two terms. Since 'values' and 'vectors' are well-known mathematical entities, we focus on the term 'eigen'. Probably it derives from German, where 'eigen' means *peculiar*, *proper*, *own*, *characteristic*,*self*, *inherent*. Actually, the use of this term goes back at least to H. von Helmholtz's 'eigentöne', signifying the tones and resonances produced by vibrating bodies or systems (e.g. [4]).

2 Some Basic Concepts

An $N \times N$ *matrix* A has the general form:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,j} & \dots & a_{1,M} \\ a_{2,1} & a_{2,2} & \dots & a_{2,j} & \dots & a_{2,M} \\ \vdots & \vdots & \dots & \dots & \dots & \vdots \\ a_{i,1} & a_{i,2} & \dots & a_{i,j} & \dots & a_{i,M} \\ \vdots & \vdots & \dots & \dots & \dots & \vdots \\ a_{N,1} & a_{N,2} & \dots & a_{i,M} & \dots & a_{N,M} \end{bmatrix}, \quad (1)$$

A square matrix A is *invertible* if and only if its determinant is different from zero.

The trace of a diagonal matrix A corresponds to the sum of the elements of its main diagonal.

A real, square matrix A is *orthogonal* if and only if $AA^T = A^TA = I$, where I is the *identity matrix*, a diagonal matrix with all elements equal to 1.

The conjugate transpose, or Hermitian transpose of a complex matrix A is $(A^*)^T = (A^T)^* = A^H$. A complex, square Hermitian matrix A is such that $A^HA = AA^H = I$.

The *spectral radius* R_λ of a square matrix A is the maximum magnitude of its eigenvalues.

A square, nonnegative (all entries ≥ 0) matrix A is said to be *primitive* if for some positive integer $k > 0$ we have that all elements of A^k are positive (> 0) values.

A graph can be associated to a real matrix A by making $a_{i,j} = 1$ whenever we have a directed connection from node j to node i, and $a_{i,j} = 0$ otherwise. Such a matrix is called the *adjacency matrix* of the graph.

Let A be a positive square matrix, and make each of its non-zero entries equal to 1, and 0 otherwise. This matrix can be understood as an adjacency matrix defining a respective graph.

A matrix A is *irreducible* if the resulting directed graph is strongly connected, i.e. given any two nodes i and j, there is a directed path from i to j. This can also be interpreted as: given any element $a_{i,j}$ of A there is an integer $k > 0$ such that the corresponding element $p_{i,j}$ of $P = A^k$ is positive (> 0).

So, we have that every primitive matrix (nonnegative) is irreducible, with all its elements becoming nonnegative for the same power of k, but not vice-versa, implying that being primitive is a more restrictive conditions than being irreducible. Also, every positive square matrix is necessarily irreducible.

The linear transformation of a vector $\vec{x}$ into a vector $\vec{y}$ can be expressed in matrix form as:

$$\vec{y} = A\vec{x} \quad (2)$$

where $\vec{x} = (x_1, x_2, \dots, x_N \text{ in } \Re^N)$ is a column vector

Two matrices A and B are said to be *similar* if a third matrix P can be found such that:

$$A = P^{-1}BP \quad (3)$$

A *diagonalizable* real matrix A is similar to a diagonal matrix D, i.e.:

$$A = P^{-1}DP, \quad (4)$$

in case the matrix P exists, $P^{-1}DP$ corresponds to the diagonalization of A.

Let A be a *symmetric* matrix. Matrix A determines a respective *quadratic form*:

$$q_A(\vec{x}) = \vec{x}^T A\vec{x} \quad (5)$$

The sign of the quadratic form of a real symmetric matrix A can be taken into account while classifying it into several types, as described in the following:

Positive definite:	$q_A(\vec{x}) > 0$ for any vector $\vec{x} \neq \vec{0}$.
Negative definite:	$q_A(\vec{x}) < 0$ for any vector $\vec{x} \neq \vec{0}$.
Positive semidefinite:	$q_A(\vec{x}) \geq 0$ for any vector $\vec{x} \neq \vec{0}$.
Negative semidefinite:	$q_A(\vec{x}) \leq 0$ for any vector $\vec{x} \neq \vec{0}$.
Indefinite:	$det(A) \neq 0$ and A is neither positive definite or negative definite.
Null-determinant case:	If $det(A) = 0$, A can be indefinite or positive semidefinite or negative semidefinite.

Observe that a matrix being positive is not the same as being positive definite (and similarly for negative and negative definite).

Let's consider as an example that $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Its quadratic form is:

$$q_A(\vec{x}) = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = x^2 + y^2. \quad (6)$$

Observe that the quadratic form of A corresponds to a second order polynomial on the variables x and y.

This polynomial can only take positive values, from which we conclude that A is *positive definite*, but is not *positive* as it includes the element zero (however, it can be said to be nonnegative).

3 Eigenvalues and Eigenvectors of a Matrix

We now proceed to defining and presenting some of the many important properties of the eigenvalues and respective eigenvectors of a matrix A.

Given a matrix A, which we will considered as being associated to its respectively implemented linear transformation, we can define one of its *eigenvalues* λ and respective *eigenvector* $\vec{v}$ as obeying:

$$\lambda\vec{v} = A\vec{v} \tag{7}$$

Thus, an eigenvector $\vec{v}$ of the A is such that, when transformed by A, yields a new vector $\lambda\vec{v}$ that preserves the inclination of the original vector, though being scaled by the respective eigenvalue λ.

The eigenvalues of a matrix A can be determined by making the respective characteristic polynomial

$$p(\lambda) = det(A - \lambda I) \tag{8}$$

equal to zero, and solving the resulting secular (or characteristic) equation. Then, one possible way to obtain the respective eigenvectors is by substituting the eigenvalues, one-by-one, into Equation 10 and solving the respectively obtained systems of linear equations.

It is also interesting to consider the equation

$$\lambda\vec{r} = \vec{r}^{T}A \tag{9}$$

defining the *right-eigenvectors* of A with respect to λ (the previous ones are called left-eigenvectors).

It can be shown that the eigenvalues λ_i are the same for both Equations 10 and 9, but the respective eigenvectors $\vec{v}$ and $\vec{r}$ are generally not identical, unless A is symmetric. This can be verified by transposing Equation 9 and comparing it with Equation 10.

Let's consider the determination of the eigevalues and respective eigenvectors of $A = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix}$. We have:

$$p(\lambda) = det(A - \lambda I) = det\left(\begin{bmatrix} 1-\lambda & 1-\lambda \\ 0 & 1+\lambda \end{bmatrix}\right) =$$
$$1 - \lambda^2 = 0 \Longrightarrow \lambda_1 = 1 \text{ and } \lambda_2 = -1$$

For $\lambda_1 = 1$, we go back to Equation 10:

$$\lambda\vec{v} = \vec{v} = A\vec{v} \Longrightarrow v_x = 1; v_y = 0. \tag{10}$$

The reader is encouraged to find the eigenvector corresponding to $\lambda_2 = -1$.

Given a matrix A, its *eigenstructure* can be understood as defined by its respective set of eigenvalues and eigenvectors, while the set of eigenvalues is called the *spectrum* of A (typically, the diagonal of A is not considered when calculating its spectrum).

Given an $N \times N$ matrix A, one of its eigenvalues λ_i may corresponds to a multiple root of the characteristic equation, implying that eigenvalue to have respective *algebraic multiplicity* $\mu_A(\lambda_i)$.

An eigenvalue λ_i can also have a respective *geometric multiplicity*, $\gamma_A(\lambda_i)$, which is given as:

$$\gamma_A(\lambda_i) = N - rank(A - \lambda_i I) \tag{11}$$

It can be shown that:

$$1 \leq \gamma_A(\lambda_i) \leq \mu_A(\lambda_i) \leq N \tag{12}$$

If $rank(A - \lambda_i I) > 1$, implying that there are two or more linear independent eigenvectors with the same eigenvalue, the respective λ_i is said to be *degenerate*. In case all eigenvalues have multiplicity 1, the spectrum is said to be *simple*.

The *Perron-Frobenius theorem* (e.g. [5]) provides important information regarding the eigenvalue and eigenvector properties of a square, nonnegative irreducible matrix A. Among other things, it implies that the respective R_λ exists and is a simple root of the respective characteristic polynomial. We also have that there is an eigenvector associated to R_λ that has strictly positive elements.

A list of several additional eigenvalue and eigenvector properties are given in Table 1, while Table 2 includes some eigenvalue and eigenvector properties related to the positive or negative definiteness of a symmetric real matrix A.

Given an $N \times N$ matrix A with respective non-degenerated eigenvalues λ_i, $i = 1, 2, \ldots, N$, with associated linearly-independent eigenvectors $\vec{v_i}$, we can define the following diagonal matrix:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & \lambda_N \end{bmatrix}, \tag{13}$$

as well as the matrix of corresponding eigenvector:

$$V = \begin{bmatrix} \uparrow & \uparrow & \ldots & \uparrow \\ \vec{v_1} & \vec{v_2} & \ldots & \vec{v_N} \\ \downarrow & \downarrow & \ldots & \downarrow \end{bmatrix}, \tag{14}$$

The previous Equation 10 can now be expressed in '*parallel*' manner (e.g. [1]) as:

$$AV = V\Lambda. \tag{15}$$

Therefore, this equation expresses the interrelationships between all eigenvalues and eigenvectors of the original matrix A.

If we left-multiply both sides of Equation 10 by V^{-1}, we get:

$$V^{-1}AV = \Lambda, \tag{16}$$

which is know as the *eigendecomposition* of A. Observer that this decomposition corresponds to the diagonalization of A.

The above result allows us to obtain all the N eigenvalues (in case they exist) given matrix A and its respective eigenvectors.

Observe that a matrix A may not have its inverse. In this case, its single-value decomposition can be used for respective diagonalization (e.g. [1]), being applicable even to non-square matrices.

Now, if we right-multiply both sides of Equation 10 by V^{-1}, and assuming a set of eigenvalues given as Λ, we have:

$$A = V\Lambda V^{-1}. \tag{17}$$

which provides a method for designing a matrix A with pre-specified eigenvalues and eigenvectors.

Given a row vector $\vec{v}$ with at least 2 components, it is easy to obtain a matrix A of which it is an eigenvector. This can be done as:

$$A = \vec{v}\vec{v}^T \tag{18}$$

If $\vec{p}$ is another vector orthogonal to $\vec{v}$ with the same dimension, a matrix A having $\vec{v}$ and $\vec{p}$ as eigenvectors, but *a priori* unspecified eigenvalues, can now be obtained as:

$$A = \vec{v}\vec{v}^T + \vec{p}\vec{p}^T \tag{19}$$

Up to N vectors $N \times 1$, each one orthogonal to all the others, can be combined in this manner so as to obtain an $N \times N$ matrix A having them as eigenvectors.

4 Gershgorin Discs

Proposed by the Russian applied mathematician S. A. Gershgorin [6], the discs, as well as the respective theorem that carry his name, provide an interesting resource for bounding the eigenvalues of complex matrices.

Given an $N \times N$ complex matrix A, let the radius of the discs be calculated as:

$$R_i = \sum_{\substack{j=1,\\ j\neq i}}^{N} |a_{i,j}| \tag{20}$$

Observe that R_i is a real value corresponding to the absolute values of all elements along the i-th row of A. The *Gershgorin disc* for row i, $G(i)$ is a closed circle defined in the complex plan with center at $a_{i,i}$ and radius R_i. Therefore, all such circles are centered at points along the real-axis.

Gershgorin's theorem states that all eigenvalues of A will necessarily be found inside the union of all the N respective Gershgorin discs, which therefore bound the eigenvalues of a matrix.

The following R algorithm can be used to visualize the Gershgorin discs of the $N \times N$ input matrix A. Observe that Nth is the angular resolution for plotting the discs.

Algorithm 1 *Gershgorin*(A, N)

1. $Nth \leftarrow 300$
2. $dth \leftarrow 2 * pi/(Nth - 1)$
3. $th \leftarrow seq(0, 2 * pi, dth)$
4. $a \leftarrow 7$
5. $for(i\ in\ seq(1, N))\{$
   (a) $R \leftarrow sum(abs(Ad[i,]))$
   (b) $Gx \leftarrow R * cos(2 * pi * th) + A[i, i]$
   (c) $Gy \leftarrow R * sin(2 * pi * th)$
   (d) $plot(Gx, Gy, xlim = c(-a, a), \ldots$
   $ylim = c(-a, a), type = "l")$
   (e) $par(new = \text{TRUE})\}$

Figure 2 illustrates the Gershgorin discs (in salmon), as well as the actual three eigenvalues (green triangles, calculated by using an eigenvalue/eigenvector library), of the following matrix:

$$A = \begin{bmatrix} 1 & -1 & 0 \\ -2 & 2 & -1 \\ 1 & 2 & 0 \end{bmatrix}, \tag{21}$$

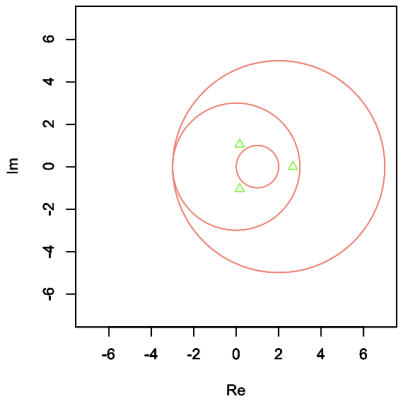

Figure 2: The Gershgorin discs for the matrix A in Eq. 21 shown in salmon, as well as the three respective actual eigenvalues, shown as green triangles.

The above results allow some interesting interpretations and predictions about the eigenvalues of a matrix. For instance, we infer that the the higher the dispersion of the values of the elements along the diagonal of A, while

preserving all the other entries, higher also will be the dispersion of the discs along the real axis.

In addition, the larger the absolute values along respective rows of A, the higher the uncertainty, as inferred by this approach, in bounding the eigenvalues. However, we should not take higher bounding regions to necessarily imply larger eigenvalues.

An interesting situation arises when we have a row i in which only the respective element corresponding to the diagonal, $a_{i,i}$, is non-zero. In this case, one of the eigenvalues will necessarily be equal to $a_{i,i}$.

5 Extrema of Multivariate Functions

In our first application example, we will briefly address the classification of the extrema of a scalar field $\psi(x_1, x_2, \ldots, x_N)$ defined on $\Re^N$.

First, we identify the points yielding null respective gradient, i.e. $||\vec{\nabla}\psi(x_1, x_2, \ldots, x_N)|| = 0$. These so-called *critical* points are candidates for being extrema of ψ. However, additional testing is required considering the *eigenvalues* of the Hessian matrix of A, namely:

$$H_{\psi(\vec{p})} = \begin{bmatrix} \frac{\partial^2\psi}{\partial^2 x} & \frac{\partial^2\psi}{\partial x \partial y} & \frac{\partial^2\psi}{\partial x \partial z} \\ \frac{\partial^2\psi}{\partial y \partial x} & \frac{\partial^2\psi}{\partial^2 y} & \frac{\partial^2\psi}{\partial y \partial z} \\ \frac{\partial^2\psi}{\partial z \partial x} & \frac{\partial^2\psi}{\partial z \partial y} & \frac{\partial^2\psi}{\partial^2 z} \end{bmatrix} \tag{22}$$

The following criteria can then be used while trying to classify the types of extrema of a scalar field:

1. **Positive-definite Hessian:** All eigenvalues of $H(\tilde{x}, \tilde{y}, \tilde{z})$ are *positive* (i.e. A is positive definite) $\Longrightarrow$ $(\tilde{x}, \tilde{y}, \tilde{z})$ is a *local minimum point*;
2. **Negative-definite Hessian:** All eigenvalues of $H(\tilde{x}, \tilde{y}, \tilde{z})$ are *negative* (i.e. A is negative definite) $\Longrightarrow$ $(\tilde{x}, \tilde{y}, \tilde{z})$ is a *local maximum point*;
3. **Indefinite Hessian:** The eigenvalues of $H(\tilde{x}, \tilde{y}, \tilde{z})$ are a mixture of *positive and negative* values $\Longrightarrow$ $(\tilde{x}, \tilde{y}, \tilde{z})$ is a *saddle point*;
4. **Otherwise:** One or more null eigenvalues $\Longrightarrow$ additional analysis is needed.

Observe the importance of the eigenvalues of $H(\psi)$ in identifying the types of extrema of a scalar field.

Let's illustrate the identification of the extrema of the scalar field $\psi = x^2 - y^2$. Its gradient is $\vec{\nabla}(\psi) = 2x\hat{\imath} - 2y\hat{\jmath}$, which has null magnitude only when $x = y = 0$, so that we have $\tilde{\vec{X}} = (0, 0)$ as the sole critical point of ψ. The Hessian of ψ is given as:

$$H(\tilde{\vec{X}}) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}. \tag{23}$$

We have, from property $[P4]$ in Table 1, that this $H(\tilde{\vec{X}})$ has eigenvalues $\lambda_1 = 2$ and $\lambda_2 = -2$, implying the found critical point to be a *saddle point*.

6 Linear Dynamical Systems

A homogeneous (no constant terms in the equations) linear dynamical system involving N variables ($\Re^N$) with constant coefficients can be expressed as:

$$S: \begin{cases} \dot{x}_1(t) = a_{1,1}x_1(t) + a_{1,2}x_2(t) + \ldots + a_{1,N}x_N(t) \\ \dot{x}_2(t) = a_{2,1}x_1(t) + a_{2,2}x_2(t) + \ldots + a_{12,N}x_N(t) \\ \ldots \\ \dot{x}_N(t) = a_{N,1}x_1(t) + a_{N,2}x_2(t) + \ldots + a_{N,N}x_N(t) \end{cases}$$

where a_i, j, $i, j = 1, 2, \ldots, N$ are real values.

The system S can be placed in the equivalent matrix form:

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,N} \\ a_{2,1} & a_{2,2} & \ldots & a_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ a_{N,1} & a_{N,2} & \ldots & a_{N,N} \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix} \tag{24}$$

or, more synthetically:

$$\dot{\vec{x}}(t) = A\vec{x}(t). \tag{25}$$

where A is an $N \times N$ matrix, and both $\dot{\vec{x}}$ and $\vec{x}$ are column vectors.

It can be shown (e.g. [7]) that the general solution of a linear system of ordinary differential equations at constant coefficients can be expressed as:

$$\vec{x}(t) = c_1 e^{-\lambda_1 t}\vec{v}_1 + c_2 e^{-\lambda_2 t}\vec{v}_2 + \ldots + c_N e^{-\lambda_N t}\vec{v}_N \tag{26}$$

where $c_1, c_2, \ldots, c_N$ are constants, $\lambda_1, \lambda_2, \ldots, \lambda_N$ are the eigenvalues of A with respective eigenvectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_N$, provided they can be found.

In case an initial condition $\vec{x}_0$ is provided, the constants can be determined as:

$$\vec{C} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{bmatrix} = \begin{bmatrix} \uparrow & \uparrow & \ldots & \uparrow \\ \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_N \\ \downarrow & \downarrow & \ldots & \downarrow \end{bmatrix}^{-1} \begin{bmatrix} x_{0,1} \\ x_{0,2} \\ \vdots \\ x_{0,N} \end{bmatrix} \tag{27}$$

As an example, let's consider the solution of the linear ODE system with:

$$\begin{bmatrix} 1 & 0 & 3 \\ 2 & 0 & 1 \\ 0 & 1 & 3 \end{bmatrix} \tag{28}$$

We obtain $\lambda_1 = 3.81912...$, $\lambda_1 = 0.09043... + 1.14062...i$, and $\lambda_3 = 0.09043... - 1.14062...i$, with respective eigenvectors:

$$\vec{v}_1 = \begin{bmatrix} 0.63557... \\ 0.48922... \\ 0.59725... \end{bmatrix};$$

$$\vec{v}_2 = \begin{bmatrix} 0.15670... + 0.50738...i \\ 0.80703... \\ -0.24042... - 0.09425i \end{bmatrix};$$

$$\vec{v}_3 = \begin{bmatrix} 0.15670... - 0.50738...i \\ 0.80703... \\ -0.24042... + 0.09425...i \end{bmatrix}. \quad (29)$$

Considering the initial condition:

$$\vec{x}_0 = \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} \quad (30)$$

we get the constants:

$$\vec{C} = \begin{bmatrix} 0.87520... \\ -0.88482... + 0.27487...i \\ -0.88482... - 0.27487...i \end{bmatrix} \quad (31)$$

Figure 3 shows the three obtained solutions $x_1(t), x_2(t), x_3(t)$ for a period of time starting at $t = 0$.

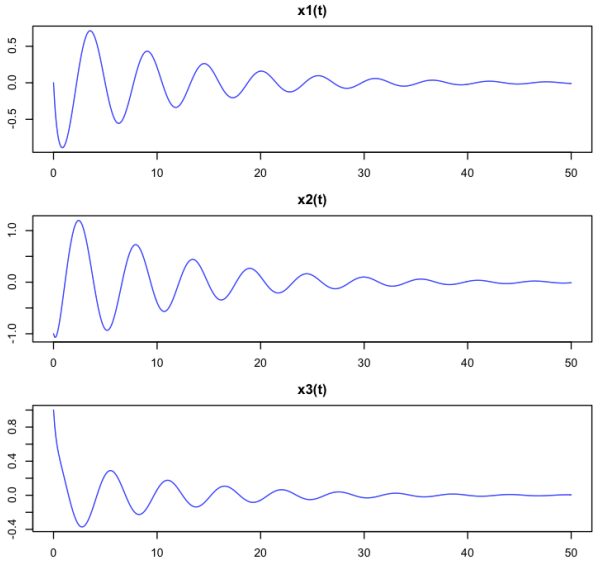

Figure 3: The solutions of the ODE in the considered example unfolding along time.

7 Markov Chains

Markov chains constitute and interesting statistical approach (e.g. [8]) that is largely applicable.

Let the following ordinary differential equation:

$$\dot{\vec{x}}(t) = f(\vec{x}(t)). \quad (32)$$

Its discretization in time yields:

$$\dot{\vec{x}}(t) = lim_{\Delta t \to 0} \frac{\vec{x}(t + \Delta t) - \vec{x}(t)}{\Delta t} \approx$$
$$\approx \frac{\vec{x}(t + \Delta t) - \vec{x}(t)}{\Delta t} \approx f(\vec{x}, t), \quad (33)$$

which implies:

$$\vec{x}(t + \Delta t) \approx f(\vec{x}, t)\Delta t + \vec{x}(t), \quad (34)$$

which corresponds to a first-order *finite difference approximation* of the original equation, which can be used to estimate the solution of that equation by using numerical approaches.

If we make $\Delta t = 1$, and replace the approximation by an equality we get the new system:

$$\vec{x}(t + 1) = f(\vec{x}(t)) + \vec{x}(t), \quad (35)$$

Observe that this is not longer a good approximation of the original equation, corresponding to a distinct dynamic system.

In $f()$ corresponds to a linear transformation, we obtain:

$$\vec{x}(t + 1) = B\vec{x}(t) + \vec{x}(t) = A\vec{x}(t), \quad (36)$$

where $A = B + I$ and in which the next state $\vec{x}^{t+1}$ depends only on a linear combination of the current states corresponding to the entries of $\vec{x}(t)$.

An *initial condition* can be expressed as $\vec{x}(t = 0)$.

Let A be an $N \times N$ real matrix. It is said to be *stochastic* or *Markov* if and only if it only contains nonnegative real entries and the sums of each of its columns is equal to 1. As a simple example, we have:

$$A = \begin{bmatrix} 0.32 & 0.49 \\ \underline{0.68} & \underline{0.51} \\ 1 & 1 \end{bmatrix}$$

Observe that each of the rows of this type of matrix A can be understood as a normalized set of probabilities, in the sense that they add to 1.

It is also interesting to consider a *row*-stochastic matrix, which has all rows adding up to one. In this case, we shall call the previous type of matrix as being *column*-stochastic. A matrix that has these two properties is called a *doubly* stochastic matrix.

Stochastic matrices can be sn to have all eigenvalues with magnitudes smaller or equal to 1. This follows from Gershgorin theorem: we have that each of the rows of a row-stochastic matrix adds up to 1, so we know by that

theorem that they are bound in magnitude to 1. However, as we also know that square matrices have the same left- and right-eigenvalues, we conclude that the right-eigenvalues of a stochastic matrix are also bound by magnitude 1.

In addition, we have that every column-stochastic matrix has at least one left-eigenvector identical to $\vec{1}$ (row vector with all entries identical to 1), because $\vec{1}A = \vec{1}$ effectively implements the sum of each of the columns of A. This vector is associated to the eigenvalue 1, so every row- and column-stochastic matrix have at least one eigenvalue equal to 1.

If A is irreducible, we have from the Perron-Frobenius that the eigenvector associated to $\lambda = 1$ can be placed in a form with strictly positive elements. This eigenvector will be associated to the stationary state of the respective Markov chain, also implying that every respective state will have a non-null probability.

It should be kept in mind that a stochastic matrix A can have: (i) more than one eigenvalue equal to 1; (ii) eigenvalues equal to zero; (iii) negative eigenvalues; (iv) complex eigenvalues.

If A is a stochastic matrix, Equation 36 defines a respective *Markov chain* on the states in $\vec{x}(t)$.

Let's consider that A is irreducible and regular. This effectively means that the state $x_i(t)$ associated to any node i of the graph representing A will, along time, influences any of the other nodes with a non-null contribution.

As already observed, the left-eigenvector associated to the eigenvalue 1 of A corresponds to the *equilibrium* or *stationary* distribution of probabilities of the Markov chain states, i.e.:

$$A\vec{p} = \vec{p} \tag{37}$$

Interestingly, this eigenvector does not depend on the initial state $\vec{x}(t = 0)$, and therefore has no 'memory' of the past dynamics or initial condition.

Let's consider the simple example of Markov chain presented in Figure 4.

The respective transition matrix A can be obtained as:

$$\begin{bmatrix} 0.3 & 0 & 0.1 & 0.6 \\ 0.7 & 0.9 & 0 & 0.1 \\ 0 & 0.1 & 0.5 & 0 \\ 0 & 0 & 0.4 & 0.3 \end{bmatrix}. \tag{38}$$

This matrix can be verified to be irreducible.

The respective eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 0.45461... + 0.30132...i$, $\lambda_3 = 0.4546... - 0.3013...i$, and $\lambda_4 = 0.0907...$,

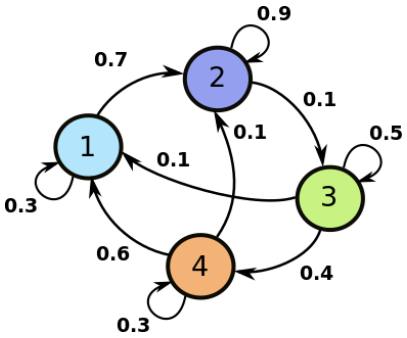

Figure 4: The graph associated to the stochastic matrix in the considered example. It is also interesting to imagine a uniformly random walk performed by a hypothetical agent along this graph, taking the outgoing links according to the respective transition probabilities.

with corresponding eigenvectors:

$$\vec{v}_1 = \begin{bmatrix} -0.122... \\ -0.967... \\ -0.193... \\ -0.110... \end{bmatrix}; \quad \vec{v}_2 = \begin{bmatrix} 0.435... - 0.334...i \\ -0.745... \\ 0.036... + 0.241...i \\ -0.273... + 0.092...i \end{bmatrix};$$

$$\vec{v}_3 = \begin{bmatrix} 0.435... + 0.334...i \\ -0.745... \\ 0.036... - 0.241...i \\ 0.273... - 0.092...i \end{bmatrix}; \quad \vec{v}_4 = \begin{bmatrix} -0.734... \\ 0.600... \\ -0.146... \\ 0.280... \end{bmatrix}$$

Observe the coexistence of real and complex eigenvalues and eigenvectors.

As expected, A has one eigenvalue identical to 1, with an associated real eigenvector corresponding to the stationary state. This can be transformed into probabilities by normalizing $\vec{p}_1 = \vec{v}_1 / sum(\vec{v}_1)$, which yields:

$$\vec{p}_1 = \begin{bmatrix} 0.0878... \\ 0.6940... \\ 0.1388... \\ 0.0793... \end{bmatrix}$$

In case we understand the transition probabilities in A as corresponding to a uniformly random walk on the respective system, the obtained distribution $\vec{p}_1$ indicates that node 2 will be much more frequently visited than the others, followed by the third, first and forth nodes.

Figure 5 illustrates the unfolding of the states values associated to the nodes along the discrete time steps $t = 0, 1, 2, \ldots, 10$.

8 Multivariate Statistics

The *multivariate normal distribution* (e.g. [9]) is particularly important for modeling and trying to make predictions on a whole set of random variables (measurements).

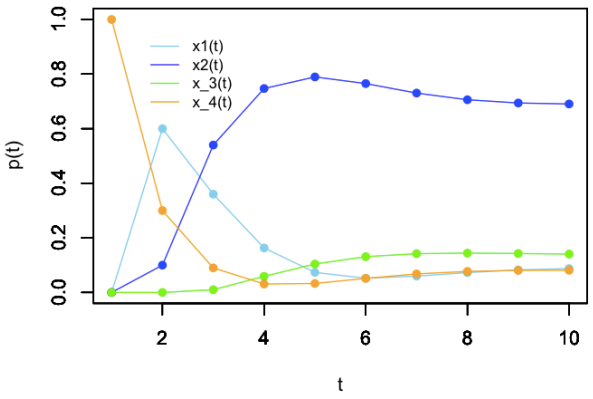

Figure 5: The values of the states probabilities (frequency of visits by a hypothetical agent) along the discrete time t from the initial condition $\vec{x}_0 = [0, 0, 0, 1]$. Observe how, as a consequence of the interconnections and respective probability transitions, the value of the state of node 4, initially equal to 1, decreases quickly, as node 2 progressively concentrates the density of visiting agents.

Considering an $N \times N$ domain, the multivariate normal probability density function with average vector $\vec{\mu}$ and covariance matrix K can be expressed as:

$$g_{\vec{\mu},K}(\vec{X}) = \\ = \frac{1}{(2\pi)^{M/2}}|K|^{-1/2} exp\left\{-\frac{1}{2}\left(\vec{X}-\vec{\mu}\right)^T K^{-1}\left(\vec{X}-\vec{\mu}\right)\right\} \quad (39)$$

Observe the *quadratic form* of K^{-1} in the argument of the exponential above.

Given N random variables X_i, $i = 1, 2, \ldots, N$, represented as the random vector $\vec{X}$, and their respective joint probability density function, the corresponding covariances can be defined as:

$$cov(X_i, X_j) = \int_{-\infty}^{\infty}(X_i - \mu_{X_i})(X_j - \mu_{X_j})p(\vec{X})d\vec{X} \quad (40)$$

The respective unbiased estimator, from M samples of the random variables is given as:

$$cov(X_i, X_j) \approx \frac{1}{M-1}\sum_{k=1}^{M}(X_i - \mu_{X_i})(X_j - \mu_{X_j}) \quad (41)$$

The covariance matrix K can now be defined so that each of its elements $k_{i,j} = cov(X_i, X_j)$. This matrix has important properties, some of which are presented as follows.

First, we have that, as a consequence of its own definition, that it is necessarily real and *symmetric*. Then, as the diagonal elements correspond to the respective *variance* (i.e. $var(X_i) = cov(Xi, Xi)$, an average of squared values), all its elements are necessarily non-zero, and so is its trace, implying that the respective eigenvalues add to a nonnegative value.

Indeed, the covariance can be shown to be positive semidefinite, i.e. its eigenvalues are all larger or equal to 0, therefore its determinant is also nonnegative.

In addition, by being symmetric, its eigenvalues are all real, and the eigenvectors $\vec{v}_i$ corresponding to distinct eigenvalues are orthogonal.

Therefore, if we define the matrix:

$$P = \begin{bmatrix} \leftarrow & \vec{v_1} & \rightarrow \\ \leftarrow & \vec{v_2} & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & \vec{v_N} & \rightarrow \end{bmatrix}, \quad (42)$$

it will orthogonal. We can apply this matrix on the original random vector, yielding a new random variable vector $\vec{Y}$ whose elements are linear combinations of the original random vectors, i.e.:

$$\vec{Y} = P\vec{X} \quad (43)$$

which corresponds to a linear statistical transformation known as discrete *Karhunen-Loève* transform, which provides the basis for the Principal Component Analysis (PCA) methodology [10, 11]. PCA implements a rotation of the original coordinates axes so as to align the first axes with the directions of largest variation, as quantified by the respective variances. The obtained random variables result completely uncorrelated, and their variance is equal to eigenvalues associated to the respective axes.

9 Concluding Remarks

The present work presented, briefly and in an introductory manner, the concept, properties and applications of eigenvalues and eigenvectors from 'eigen-centered' position. By starting with a review of some of their important properties, it was possible to discuss the subsequent applications in a more integrated and systematic way. The interesting Gershgorin approach was also briefly outlined and illustrated.

The addressed eigenvalues and eigenvectors applications included scalar field extrema characterization, solution of linear dynamic systems at constant coefficients, Markov chains, as well as some aspects of multivariate statistics.

The already large potential of theoretical and practical applications of eigenvalues and eigenvectors is being constantly further enhanced thanks to continuing advances in computer science, allowing respective calculations on matrices of ever increasing sizes. This opens up new prospects in theoretical and applied research. It is hoped that the covered presentation may motivate the reader to probe further in this interesting area.

Acknowledgments.

Luciano da F. Costa thanks CNPq (grant no. 307085/2018-0) for sponsorship. This work has benefited from FAPESP grant 15/22308-2.

References

[1] G. H. Golub and C. F. van Loan. *Matrix Computations*. The Johns Hopkins University Press, 1996.

[2] Wikipedia. Eigenvalues and eigenvectors. Wikipedia, the free encyclopedia, 2020. https://en.wikipedia.org/wiki/Eigenvalues_and_eigenvectors. [Online; accessed 10-Apr-2020.].

[3] H. Sagan. *Boundary and eigenvalue problems in mathematical Physics*. Dover, 1989.

[4] H. von Helmholtz. *Die Lehre von den Tonempfindungen als Physiologische Grundlage für die Rhoerie der Musik*. Braunschweig Druk und Verlag von Friedrich Vieweg und Sohn, 1896.

[5] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[6] S. Gershgorin. Über die abgrenzung der eigenwerte einer matrix. *Izv. Akad. Nauk. URSS Otd. Fiz.-Mat. Nauk*, pages 749–754, 1931.

[7] R. K. Nagle, E. B. Saff, and A. D. Snider. *Fundamentals of Differential Equations*. Pearson, 2017.

[8] G. Kemeny and J. L. Snell. *Finite Markov Chains*. van Nostrand Princeton, 1960.

[9] L. da F. Costa. Multivariate statistical modeling. https://www.researchgate.net/publication/340442989_Multivariate_Statistical_Modeling_CDT-26, 2019. [Online; accessed 10-Apr-2020.].

[10] F. Gewers, G. R. Ferreira, H. F. Arruda, F. N. Silva, C. H. Comin, D. R. Amancio, and L. da F. Costa. Principal component analysis: A natural approach to data exploration. Researchgate, 2019. https://www.researchgate.net/publication/324454887_Principal_Component_Analysis_A_Natural_Approach_to_Data_Exploration. accessed 25-Dec-2019.

[11] L. da F. Costa. Features transformation and normalization. Researchgate, 2019. https://www.researchgate.net/publication/340114268_Features_Transformation_and_Normalization_A_Visual_Approach_CDT-24. [Online; accessed 10-Apr-2020.].

Costa's Didactic Texts – CDTs

CDTs intend to be a halfway point between a formal scientific article and a dissemination text in the sense that they: (i) explain and illustrate concepts in a more informal, graphical and accessible way than the typical scientific article; and (ii) provide more in-depth mathematical developments than a more traditional dissemination work.

It is hoped that CDTs can also incorporate new insights and analogies concerning the reported concepts and methods. We hope these characteristics will contribute to making CDTs interesting both to beginners as well as to more senior researchers.

Each CDT focuses on a limited set of interrelated concepts. Though attempting to be relatively self-contained, CDTs also aim at being relatively short. Links to related material are provided in order to complement the covered subjects.

Observe that CDTs, which come with absolutely no warranty, are non distributable and for non-commercial use only.

The complete set of CDTs can be found at: https://www.researchgate.net/project/Costas-Didactic-Texts-CDTs.

Table 1: Some of the some properties of eigenvalues and eigenvectors of a matrix A.

[P1]: Let $\alpha \in \Re$, $\alpha \neq 0$. The matrix αA has the same eigenvectors as A, while the eigenvalues are multiplied by α.
[P2]: Let $\alpha \in \Re$, $\alpha \neq 0$. If $\vec{v}$ is an eigenvector of A, so is $\alpha\vec{v}$. Therefore, the magnitude of an eigenvector cannot be uniquely specified.
[P3]: The characteristic polynomial of an $N \times N$ matrix has degree N, and so has N roots, which may appear with multiplicity.
[P4]: The eigenvalues of a diagonal matrix are the elements of its diagonal.
[P5]: The *trace* of a matrix is equal to the sum of its eigenvalues.
[P6]: The *product* of the eigenvalues of a matrix is equal to its determinant.
[P7]: If λ is an eigenvalue of A, it is also an eigenvalue of A^T.
[P8]: The eigenvalues of A^r, for a positive integer r, are $(\lambda_i)^r$.
[P9]: If λ is an eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}.
[P10]: Eigenvectors associated to *distinct* eigenvalues are *linearly independent*.
[P11]: The eigenvalues of a *real* matrix are real or come in conjugate pairs.
[P12]: All the eigenvalues of a *complex*/real matrix A which is *Hermitian*/symmetric matrix are real.
[P13]: If a *complex*/real matrix A is *Hermitian*/symmetric, all its eigenvectors are linearly independent and mutually orthogonal.
[P14]: If a *complex*/real matrix A is *unitary*/orthogonal, then all its eigenvalues have absolute value equal to 1.
[P15]: A is *invertibe* if and only if all its eigenvalues are non-null.

Table 2: Some positive and negative definiteness-related properties of a *symmetric* real matrix A.

Positive definite:	If and only if all its eigenvalues are positive.
Negative definite:	If and only if all its eigenvalues are negative
Positive semidefinite:	If and only if all its eigenvalues are nonnegative (≥ 0).
Negative semidefinite:	If and only if all its eigenvalues are non-positive (≤ 0).
Indefinite:	If and only it has both positive and negative eigenvalues.