Introduction to Eigenvalues and Eigenvectors

A. Havens

Department of Mathematics
University of Massachusetts, Amherst

April 2 - 6, 2018

Outline

We will study the behavior of *linear endomorphisms of* $\mathbb{R}$*-vector spaces*, i.e., $\mathbb{R}$-linear transformations $T : V \to V$, by studying subspaces $E \subseteq V$ which are preserved via scaling by the endomorphism:

$$T(\mathbf{x}) = \lambda\mathbf{x} \text{ for all } \mathbf{x} \in E\,.$$

Such a subspace is called an *eigenspace* of the endomorphism T, associated to the number λ, which is called an *eigenvalue*. A nonzero vector $\mathbf{x}$ such that $T(\mathbf{x}) = \lambda\mathbf{x}$ for some number λ is called an *eigenvector*.

"Eigen-" is a German adjective which means "characteristic" or "own". Henceforth, we'll bandy the prefix "eigen-" about without apology, whenever we refer to objects which arise from eigenspaces of some linear endomorphism.

Motivating Applications

Understanding eigendata associated to a linear endomorphism T is among the most fruitful ways to analyze linear behavior in applications. There are three primary, non-exclusive themes, which I name *principal directions*, *characteristic dynamical modes*, and *spectral methods*.

- Principal directions arise whenever an eigenvector determines a physically/geometrically relevant axis or direction.
- Characteristic dynamical modes arise in dynamical problems, whenever a general solution is a superposition (i.e., a linear combination) of certain characteristic solutions.
- Spectral methods involve studying the eigenvalues themselves, as an invariant of the object to which they are associated.

A wildly noncomprehensive list of applications follows.

Motivating Applications

- In mechanics, the eigenvectors of the moment of inertia tensor for a rigid body give the principal axes.
- In differential geometry, the eigenvalues of the shape operator of a smooth surface give the principal curvature functions of the surface, and the eigenvectors give tangent vector fields to the lines of curvature.
- In statistics, one may study large data sets via principal component analysis (PCA), which uses *eigendecomposition* to stratify the data into components which are statistically independent (so the covariance vanishes between components). The eigenvectors giving principal component directions are a data-science analog of the principal axes in mechanics.

- First-order linear difference equations $\mathbf{x}_k = \mathrm{A}\mathbf{x}_{k-1}$, which model some discrete dynamical systems and recursive linear equation systems, can be solved using eigentheory.
- A special case is linear Markov chains, which model probabalistic processes, and are used, e.g., in signal and image processing, and also in machine learning.
- Facial recognition software uses the concept of an *eigenface* in facial identification, while voice recognition software employs the concept of an *eigenvoice*. These allow dimension reduction, and are special cases of principal component analysis.

- In the study of continuous dynamical systems, eigenfunctions of a linear differential operator are used to construct general solutions. The eigenvalues may correspond to physically important quantities, like rates or energies, and eigenvectors/eigenfunctions represent solutions of the dynamics.
  In particular:
  - in an oscillatory system, the eigenvalues are called *eigenfrequencies*, while the associated eigenfunctions represent the shapes of corresponding vibrational modes.
  - quantum numbers are eigenvalues, associated to *eigenstates*, which are solutions to the Schrödinger equation.
  - In epidemiology, the basic reproduction number, which measures the average number of infected cases generated by an infected individual in an uninfected population, is the maximum eigenvalue of the "next generation matrix."

- The study of spectral graph theory examines the eigenvalues of adjacency matrices of graphs and their associated discrete Laplacian operators to deduce properties of graphs. Such eigenanalysis made the Google era possible (as the original Google PageRank algorithm is based on spectral graph analysis.)
- The spectra of smooth Laplacians are of interest in the study of elastic, vibrating membranes, such as a drum head. A famous problem in continuum mechanics, as phrased by Lipman Bers, is "can you hear the shape of a drum?"
- The most tenable application for us, in this class, is the complete classification of the geometry of linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$.

We will develop the theory of real eigenvectors and eigenvalues of real square matrices and examine a few simple applications.

Many of the examples listed above require more sophisticated mathematics, as well as additional application-specific background beyond the scope of this course.

Nevertheless, you shall discover the power of eigenstuffs in a few examples.

Formal Definitions

Definition

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation of $\mathbb{R}^n$. Then a nonzero vector $\mathbf{x} \in \mathbb{R}^n - \{\mathbf{0}\}$ is called an *eigenvector* of T if there exists some number $\lambda \in \mathbb{R}$ such that

$$T(\mathbf{x}) = \lambda \mathbf{x}.$$

The real number λ is called a *real eigenvalue* of the real linear transformation T.

Let A be an $n \times n$ matrix representing the linear transformation T. Then, $\mathbf{x}$ is an eigenvector of the matrix A if and only if it is an eigenvector of T, if and only if

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

for an eigenvalue λ.

Remark

We will prioritize the study of real eigenstuffs, primarily using 2×2 and 3×3 matrices, which give a good general sense of the theory.

However, it will later be fruitful, even for real matrices, to allow $\lambda \in \mathbb{C}$, and $\mathbf{x} \in \mathbb{C}^n$ (the case of complex eigenvalues is related to the geometry of rotations, and occurs in dynamical systems featuring oscillatory behavior).

Thus, we will have a modified definition for eigenvalues and eigenvectors in the future, when we are ready to study complex eigentheory for real matrices.

Examples in 2-Dimensions

Example

Let $\mathbf{v} \in \mathbb{R}^2$ be a nonzero vector, and $\ell = \mathrm{Span}\,\{\mathbf{v}\}$. Let $\mathrm{Ref}_\ell : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation of the plane given by reflection through the line ℓ.

Then, since $\mathrm{Ref}_\ell(\mathbf{v}) = 1\mathbf{v}$, $\mathbf{v}$ is an eigenvector of Ref_ℓ with eigenvalue 1, and $\ell = \mathrm{Span}\,\{\mathbf{v}\}$ is an *eigenline* or *eigenspace* of the reflection. Note, any nonzero multiple of $\mathbf{v}$ is also an eigenvector with eigenvalue 1, by linearity.

Can you describe another eigenvector of Ref_ℓ, with a different associated eigenvalue? What is the associated eigenspace?

Examples in 2-Dimensions

Example

Let $\mathbf{v} \in \mathbb{R}^2$ be a nonzero vector, and $\ell = \mathrm{Span}\,\{\mathbf{v}\}$. Let $\mathrm{Ref}_\ell : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear transformation of the plane given by reflection through the line ℓ.

Then, since $\mathrm{Ref}_\ell(\mathbf{v}) = 1\mathbf{v}$, $\mathbf{v}$ is an eigenvector of Ref_ℓ with eigenvalue 1, and $\ell = \mathrm{Span}\,\{\mathbf{v}\}$ is an *eigenline* or *eigenspace* of the reflection. Note, any nonzero multiple of $\mathbf{v}$ is also an eigenvector with eigenvalue 1, by linearity.

Can you describe another eigenvector of Ref_ℓ, with a different associated eigenvalue? What is the associated eigenspace?

If $\mathbf{u} \in \mathbb{R}^2$ is any nonzero vector perpendicular to $\mathbf{v}$, then $\mathbf{u}$ is an eigenvector of Ref_ℓ with eigenvalue -1. The line spanned by $\mathbf{u}$ is also an eigenspace.

Examples in 2-Dimensions

Example

For $\mathbf{v}$ and ℓ as above, the orthogonal projection $\mathrm{proj}_\ell(\mathbf{x}) = \frac{\mathbf{v}\cdot\mathbf{x}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{x}$ has the same eigenspaces as Ref_ℓ, but a different eigenvalue for the line $\ell^\perp = \mathrm{Span}\,\{\mathbf{u}\}$ for $\mathbf{u} \in \mathbb{R}^2 - \{\mathbf{0}\}$ with $\mathbf{u}\cdot\mathbf{v} = 0$.

Indeed, $\mathrm{proj}_\ell\mathbf{u} = \mathbf{0}$, whence, $\mathbf{u}$ is an eigenvector whose associated eigenvalue is 0.

It is crucial to remember: eigenvectors must be nonzero, but eigenvalues may be zero, or any other real number.

Examples in 2-Dimensions

Example

Let $\mathrm{A} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$, for a nonzero real number k.

The map $\mathbf{x} \mapsto \mathrm{A}\mathbf{x}$ is a shearing transformation of $\mathbb{R}^2$.

Given that 1 is the only eigenvalue of A, describe a basis of the associated eigenspace.

Examples in 2-Dimensions

Example

An eigenvector $\mathbf{x}$ of the shearing matrix A with eigenvalue 1 must satisfy $\mathrm{A}\mathbf{x} = \mathbf{x}$, whence $\mathbf{x}$ is a solution of the homogeneous equation $\mathrm{A}\mathbf{x} - \mathrm{I}_2\mathbf{x} = (\mathrm{A} - \mathrm{I}_2)\mathbf{x} = \mathbf{0}$.

Examples in 2-Dimensions

Example

An eigenvector **x** of the shearing matrix A with eigenvalue 1 must satisfy $\mathrm{A}\mathbf{x} = \mathbf{x}$, whence **x** is a solution of the homogeneous equation $\mathrm{A}\mathbf{x} - \mathrm{I}_2\mathbf{x} = (\mathrm{A} - \mathrm{I}_2)\mathbf{x} = \mathbf{0}$.

Therefore the components x_1 and x_2 of **x** must satisfy

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1-1 & k \\ 0 & 1-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0x_1 + kx_2 \\ 0x_1 + 0x_2 \end{bmatrix}$$

Examples in 2-Dimensions

Example

An eigenvector $\mathbf{x}$ of the shearing matrix A with eigenvalue 1 must satisfy $\mathrm{A}\mathbf{x} = \mathbf{x}$, whence $\mathbf{x}$ is a solution of the homogeneous equation $\mathrm{A}\mathbf{x} - \mathrm{I}_2\mathbf{x} = (\mathrm{A} - \mathrm{I}_2)\mathbf{x} = \mathbf{0}$.

Therefore the components x_1 and x_2 of $\mathbf{x}$ must satisfy

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1-1 & k \\ 0 & 1-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0x_1 + kx_2 \\ 0x_1 + 0x_2 \end{bmatrix}$$
$$\implies kx_2 = 0 \implies x_2 = 0\,.$$

Examples in 2-Dimensions

Example

Thus, $\mathbf{x} = \begin{bmatrix} t \\ 0 \end{bmatrix}$, $t \in \mathbb{R} - \{0\}$ is an eigenvector of the shearing matrix A, with eigenvalue 1, and the x_1 axis is the corresponding eigenspace.

One can check directly that there are no other eigenvalues or eigenspaces (a good exercise!).

Examples in 2-Dimensions

Example

The matrix $\mathrm{J} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ has no real eigenvectors.

Indeed, the only proper subspace of $\mathbb{R}^2$ preserved by the map $\mathbf{x} \mapsto \mathrm{J}\mathbf{x}$ is the trivial subspace.

All lines through $\mathbf{0}$ are rotated by $\pi/2$. We will later see that this matrix has *purely imaginary* eigenvalues, as will be the case with other rotation matrices.

Example: a 3×3 Upper triangular Matrix

Example

Consider the upper triangular matrix

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}.$$

Show that the eigenvalues are the entries a_{11}, a_{22} and a_{33} along the main diagonal.

Example: a 3×3 Upper triangular Matrix

Example

Consider the upper triangular matrix

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}.$$

Show that the eigenvalues are the entries a_{11}, a_{22} and a_{33} along the main diagonal.

If λ is an eigenvalue of A, then there exists a nonzero vector $\mathbf{x} \in \mathbb{R}^3$ such that $\mathrm{A}\mathbf{x} = \lambda\mathbf{x}$. But then $\mathrm{A}\mathbf{x} - \lambda\mathbf{x} = (\mathrm{A} - \lambda\mathrm{I}_3)\mathbf{x} = \mathbf{0}$ must have a nontrivial solution.

Example: a 3×3 Upper triangular Matrix

Example

Consider the upper triangular matrix

$$\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix}.$$

Show that the eigenvalues are the entries a_{11}, a_{22} and a_{33} along the main diagonal.

If λ is an eigenvalue of A, then there exists a nonzero vector $\mathbf{x} \in \mathbb{R}^3$ such that $\mathrm{A}\mathbf{x} = \lambda\mathbf{x}$. But then $\mathrm{A}\mathbf{x} - \lambda\mathbf{x} = (\mathrm{A} - \lambda\mathrm{I}_3)\mathbf{x} = \mathbf{0}$ must have a nontrivial solution.

But the homogeneous equation has a nontrivial solution if and only if the square matrix $\mathrm{A} - \lambda\mathrm{I}_3$ has determinant equal to 0.

Example: a 3×3 Upper triangular Matrix

Example

Since

$$A - \lambda I_3 = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix},$$

Example: a 3×3 Upper triangular Matrix

Example

Since

$$\mathrm{A} - \lambda \mathrm{I}_3 = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix},$$

$$\det(\mathrm{A} - \lambda \mathrm{I}_3) = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda)\,.$$

Example: a 3×3 Upper triangular Matrix

Example

Since

$$\mathrm{A} - \lambda \mathrm{I}_3 = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix},$$

$$\det(\mathrm{A} - \lambda \mathrm{I}_3) = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda)\,.$$

Thus $\det(\mathrm{A} - \lambda \mathrm{I}_3) = 0 \iff (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) = 0$, which holds if and only if $\lambda \in \{a_{11}, a_{22}, a_{33}\}$.

Example: a 3×3 Upper triangular Matrix

Example

Since

$$\mathrm{A} - \lambda \mathrm{I}_3 = \begin{bmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ 0 & a_{22} - \lambda & a_{23} \\ 0 & 0 & a_{33} - \lambda \end{bmatrix},$$

$$\det(\mathrm{A} - \lambda \mathrm{I}_3) = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda)\,.$$

Thus $\det(\mathrm{A} - \lambda \mathrm{I}_3) = 0 \iff (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) = 0$, which holds if and only if $\lambda \in \{a_{11}, a_{22}, a_{33}\}$.

The equation $\det \mathrm{A} - \lambda \mathrm{I}_3 = (a_{11} - \lambda)(a_{22} - \lambda)(a_{33} - \lambda) = 0$ is an example of a *characteristic equation*.

A Theorem: Eigenvalues of an Upper triangular Matrix

We can extend the idea of the above example to prove the following theorem.

Theorem

If A is an $n \times n$ triangular matrix, then the eigenvalues of A are precisely the elements on the main diagonal.

In particular, the eigenvalues of a diagonal matrix are the entries $\{a_{11}, \dots, a_{nn}\}$ of the main diagonal.

A Theorem: Independence of Eigenvectors with Distinct Eigenvalues

Theorem

If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are eigenvectors that correspond respectively to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of an $n \times n$ matrix A, then the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is linearly independent.

Proof.

We proceed by contradiction. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a linearly dependent set.

A Theorem: Independence of Eigenvectors with Distinct Eigenvalues

Theorem

If $\mathbf{v}_1, \ldots, \mathbf{v}_r$ *are eigenvectors that correspond respectively to distinct eigenvalues* $\lambda_1, \ldots, \lambda_r$ *of an* $n \times n$ *matrix* A*, then the set* $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ *is linearly independent.*

Proof.

We proceed by contradiction. Suppose $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ is a linearly dependent set.

Observe that, being a set of eigenvectors, $\mathbf{v}_i \neq \mathbf{0}$ for any $i = 1, \ldots r$, and by linear dependence we can find an index p, $1 < p < r$ such that $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is linearly independent, and $\mathbf{v}_{p+1} \in \mathrm{Span}\,\{\mathbf{v}_1, \ldots \mathbf{v}_p\}$.

A Proof

Proof (continued.)

So there exist constants $c_1, \ldots, c_{p-1}$ not all zero such that

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + \ldots c_p\mathbf{v}_p\,.$$

A Proof

Proof (continued.)

So there exist constants $c_1, \ldots, c_{p-1}$ not all zero such that

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + \ldots c_p\mathbf{v}_p \,.$$

Left-multiplying both sides of this relation by A, we obtain

$$\mathrm{A}\mathbf{v}_{p+1} = \mathrm{A}(c_1\mathbf{v}_1 + \ldots c_p\mathbf{v}_p)$$

A Proof

Proof (continued.)

So there exist constants $c_1, \ldots, c_{p-1}$ not all zero such that

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + \ldots c_p\mathbf{v}_p \,.$$

Left-multiplying both sides of this relation by A, we obtain

$$\begin{aligned} \mathrm{A}\mathbf{v}_{p+1} &= \mathrm{A}(c_1\mathbf{v}_1 + \ldots c_p\mathbf{v}_p) \implies \\ \lambda_{p+1}\mathbf{v}_{p+1} &= c_1\lambda_1\mathbf{v}_1 + \ldots c_p\lambda_p\mathbf{v}_p \end{aligned}$$

A Proof

Proof (continued.)

So there exist constants $c_1, \ldots, c_{p-1}$ not all zero such that

$$\mathbf{v}_{p+1} = c_1\mathbf{v}_1 + \ldots c_p\mathbf{v}_p\,.$$

Left-multiplying both sides of this relation by A, we obtain

$$\mathrm{A}\mathbf{v}_{p+1} = \mathrm{A}(c_1\mathbf{v}_1 + \ldots c_p\mathbf{v}_p) \implies$$
$$\lambda_{p+1}\mathbf{v}_{p+1} = c_1\lambda_1\mathbf{v}_1 + \ldots c_p\lambda_p\mathbf{v}_p$$

Scaling the original relation by λ_{p+1}, and subtracting the relations, we obtain

$$\mathbf{0} = c_1(\lambda_1 - \lambda_{p+1})\mathbf{v}_1 + \ldots c_p(\lambda_p - \lambda_{p+1})\mathbf{v}_p\,.$$

Proof (continued.)

This final relation is impossible:

Proof (continued.)

This final relation is impossible:

since the set $\{\mathbf{v}_1, \dots \mathbf{v}_p\}$ is linearly independent, this equation requires that the scalar weights all vanish, but we know at least one c_i must be nonzero since $\mathbf{v}_{p+1}$ is an eigenvector (and hence nonzero), while since the eigenvalues are all distinct, $\lambda_i - \lambda_{p+1} \neq 0$ for any $i = 1, \dots p$.

Proof (continued.)

This final relation is impossible:

since the set $\{\mathbf{v}_1, \ldots \mathbf{v}_p\}$ is linearly independent, this equation requires that the scalar weights all vanish, but we know at least one c_i must be nonzero since $\mathbf{v}_{p+1}$ is an eigenvector (and hence nonzero), while since the eigenvalues are all distinct, $\lambda_i - \lambda_{p+1} \neq 0$ for any $i = 1, \ldots p$.

Thus our assumption that the set $\{\mathbf{v}_1, \ldots, \mathbf{v}_r\}$ was linearly dependent is untenable. $\square$

An Eigenvalue of 0

Let A be a matrix which has an eigenvector $\mathbf{x}$ such that the associated eigenvalue is $\lambda = 0$. Then the eigenspace associated to the zero eigenvalue is the null space of A.

This is easy to see. Let E_0 be the 0-eigenspace. Then for any $\mathbf{x} \in E_0$, $\mathrm{A}\mathbf{x} = 0\mathbf{x} = \mathbf{0}$, whence $\mathbf{x} \in \mathrm{Nul}\,\mathrm{A}$ for all $\mathbf{x} \in E_0$, which implies $E_0 \subseteq \mathrm{Nul}\,\mathrm{A}$. Conversely, for any $\mathbf{x} \in \mathrm{Nul}\,\mathrm{A}$, $\mathrm{A}\mathbf{x} = \mathbf{0} = 0\mathbf{x}$, so $\mathbf{x} \in E_0$, and $\mathrm{Nul}\,\mathrm{A} \subseteq E_0$. Thus $E_0 = \mathrm{Nul}\,\mathrm{A}$.

It follows that A is invertible if and only if 0 is not a eigenvalue of A.

Determinant Via Row-Ops

We recall some facts about determinants.

Suppose a square matrix A can be row reduced to an echelon form $\mathrm{B} = (b_{ij})$ using only r row interchanges, and elementary row replacements $R_i - sR_j \mapsto R_i$, without row scalings $sR_i \mapsto R_i$. Then

$$\det \mathrm{A} = \begin{cases} (-1)^r \prod_{i=1}^n b_{ii} & \text{when } \mathrm{A} \text{ is invertible} \\ 0 & \text{when } \mathrm{A} \text{ is not invertible} \end{cases} .$$

In particular, A is invertible if and only if $\det \mathrm{A} \neq 0$.

Review of Determinant Properties

Theorem

Let $\mathrm{A}, \mathrm{B} \in \mathbb{R}^{n\times n}$*. Then*

a. *If* $\mathrm{A} = (a_{ij})$ *is triangular, then* $\det \mathrm{A} = \prod_{i=1}^{n} a_{ii}$*, the product of the diagonal entries.*
b. $\det(\mathrm{AB}) = (\det \mathrm{A})(\det \mathrm{B})$.
c. $\det \mathrm{A}^{\mathrm{t}} = \det \mathrm{A}$.
d. A *is invertible if and only if* $\det \mathrm{A} \neq 0$.
e. *A row replacement operation on* A *does not alter* $\det \mathrm{A}$*. A row swap operation on* A *reverses the sign of* $\det \mathrm{A}$*. A row scaling by* s *of a row of* A *scales the* $\det \mathrm{A}$ *by* s.

Finding Eigenvalues

Given eigenvalues of A, it is straightforward to solve for associated eigenvectors using our knowledge of linear systems. But how do we find the eigenvalues of A?

The observations about the determinant and invertibility are the key.

We'll construct a determinant equation, yielding a polynomial, such that its solutions are the eigenvalues.

Determinants and Characteristic Equations

Let $\mathrm{A} \in \mathbb{R}^{n\times n}$. Suppose λ is an eigenvalue of A with eigenvector $\mathbf{x}$. Then

$$\mathrm{A}\mathbf{x} = \lambda\mathbf{x} \implies (\mathrm{A} - \lambda\mathrm{I}_n)\mathbf{x} = \mathbf{0} \implies \mathbf{x} \in \mathrm{Nul}\,(\mathrm{A} - \lambda\mathrm{I}_n)\,.$$

Since $\mathbf{x} \neq \mathbf{0}$ (being an eigenvector), we deduce that $\mathrm{Nul}\,(\mathrm{A} - \lambda\mathrm{I}_n)$ is nontrivial, whence it is noninvertible and $\det(\mathrm{A} - \lambda\mathrm{I}_n) = 0$.

Definition

Given a matrix $\mathrm{A} \in \mathbb{R}^{n\times n}$, the *characteristic equation for* A is

$$\det(\mathrm{A} - \lambda\mathrm{I}_n) = 0\,.$$

The left hand expression $\det(\mathrm{A} - \lambda\mathrm{I}_n)$ determines a polynomial in λ, called the *characteristic polynomial*, whose real roots are precisely the real eigenvalues of A.

A 2×2 Example

Example

Let $A = \begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A.

A 2×2 Example

Example

Let $A = \begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A.

Solution: The characteristic equation is

$$0 = \det\left(\begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right)$$

.

A 2×2 Example

Example

Let $A = \begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A.

Solution: The characteristic equation is

$$\begin{aligned} 0 &= \det\left(\begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\ &= (6-\lambda)(-4-\lambda) - (-2)(8) \end{aligned}$$

.

A 2×2 Example

Example

Let $A = \begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A.

Solution: The characteristic equation is

$$\begin{aligned} 0 &= \det\left(\begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\ &= (6-\lambda)(-4-\lambda) - (-2)(8) \\ &= -24 + 4\lambda - 6\lambda + \lambda^2 + 16 = \lambda^2 - 2\lambda - 8 \end{aligned}$$

.

A 2×2 Example

Example

Let $\mathrm{A} = \begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A.

Solution: The characteristic equation is

$$\begin{aligned} 0 &= \det\left(\begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\ &= (6-\lambda)(-4-\lambda) - (-2)(8) \\ &= -24 + 4\lambda - 6\lambda + \lambda^2 + 16 = \lambda^2 - 2\lambda - 8 \\ &= (\lambda+2)(\lambda-4)\,. \end{aligned}$$

A 2×2 Example

Example

Let $A = \begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix}$. Find the eigenvalues and eigenvectors of A.

Solution: The characteristic equation is

$$\begin{aligned} 0 &= \det\left(\begin{bmatrix} 6 & 8 \\ -2 & -4 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}\right) \\ &= (6-\lambda)(-4-\lambda) - (-2)(8) \\ &= -24 + 4\lambda - 6\lambda + \lambda^2 + 16 = \lambda^2 - 2\lambda - 8 \\ &= (\lambda+2)(\lambda-4)\,. \end{aligned}$$

Thus $\lambda_1 = -2$ and $\lambda_2 = 4$ are the eigenvalues of A.

A 2×2 Example

Example

To obtain the eigenvectors, we must solve systems associated to each eigenvalue:

$$\big(\mathrm{A} - (-2)\mathrm{I}_2\big)\mathbf{x} = \mathbf{0} \text{ and } \big(\mathrm{A} - (4)\mathrm{I}_2\big)\mathbf{x} = \mathbf{0}.$$

A 2×2 Example

Example

To obtain the eigenvectors, we must solve systems associated to each eigenvalue:

$$\big(\mathrm{A} - (-2)\mathrm{I}_2\big)\mathbf{x} = \mathbf{0} \text{ and } \big(\mathrm{A} - (4)\mathrm{I}_2\big)\mathbf{x} = \mathbf{0}.$$

For $\lambda_1 = -2$, this yields a homogeneous system with augmented matrix

$$\left[\begin{array}{cc|c} 8 & 8 & 0 \\ -2 & -2 & 0 \end{array}\right],$$

which is solved so long as the components x_1 and x_2 of $\mathbf{x}$ satisfy $x_2 = -x_1$,

A 2×2 Example

Example

To obtain the eigenvectors, we must solve systems associated to each eigenvalue:

$$\big(\mathrm{A} - (-2)\mathrm{I}_2\big)\mathbf{x} = \mathbf{0} \text{ and } \big(\mathrm{A} - (4)\mathrm{I}_2\big)\mathbf{x} = \mathbf{0}.$$

For $\lambda_1 = -2$, this yields a homogeneous system with augmented matrix

$$\left[\begin{array}{cc|c} 8 & 8 & 0 \\ -2 & -2 & 0 \end{array}\right],$$

which is solved so long as the components x_1 and x_2 of $\mathbf{x}$ satisfy $x_2 = -x_1$,

Thus, e.g., $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ spans the -2-eigenspace.

A 2×2 Example

Example

For $\lambda_1 = 4$ the corresponding homogeneous system has augmented matrix

$$\left[\begin{array}{cc|c} 2 & 8 & 0 \\ -2 & -8 & 0 \end{array}\right],$$

which is solved whenever the components x_1 and x_2 satisfy $x_1 = -4x_2$.

A 2×2 Example

Example

For $\lambda_1 = 4$ the corresponding homogeneous system has augmented matrix

$$\left[\begin{array}{cc|c} 2 & 8 & 0 \\ -2 & -8 & 0 \end{array}\right],$$

which is solved whenever the components x_1 and x_2 satisfy $x_1 = -4x_2$.

Thus, e.g., $\begin{bmatrix} 4 \\ -1 \end{bmatrix}$ spans the 4-eigenspace.

A 3×3 example

Example

Let $A = \begin{bmatrix} 2 & -2 & -1 \\ -1 & 1 & -1 \\ -1 & -2 & 2 \end{bmatrix}$ and let $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

A 3×3 example

Example

Let $\mathrm{A} = \begin{bmatrix} 2 & -2 & -1 \\ -1 & 1 & -1 \\ -1 & -2 & 2 \end{bmatrix}$ and let $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

(a) Show that $\mathbf{v}$ is an eigenvector of A. What is its associated eigenvalue?

(b) Find the characteristic equation of A.

(c) Find the remaining eigenvalue(s) of A, and describe the associated eigenspace(s).

A 3×3 example

Example

Solution:

(a) It is easy to check that $A\mathbf{v} = -\mathbf{v}$, whence $\mathbf{v}$ is an eigenvector with associated eigenvalue -1.

A 3×3 example

Example

Solution:

(a) It is easy to check that $\mathrm{A}\mathbf{v} = -\mathbf{v}$, whence $\mathbf{v}$ is an eigenvector with associated eigenvalue -1.

Observe that since $\mathbf{v} = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$, this can only be the case since the sum of entries in each row of A is -1.

A 3×3 example

Example

Solution:

(a) It is easy to check that $A\mathbf{v} = -\mathbf{v}$, whence $\mathbf{v}$ is an eigenvector with associated eigenvalue -1.

Observe that since $\mathbf{v} = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$, this can only be the case since the sum of entries in each row of A is -1. More generally, $\mathbf{e}_1 + \mathbf{e}_2 + \ldots + \mathbf{e}_n$ is an eigenvalue of an $n \times n$ matrix if and only if the sum of all entries in the rows of the matrix equal a constant λ, which is then the eigenvalue for $\mathbf{v}$.

A 3×3 example

Example

Solution:

(a) It is easy to check that $\mathrm{A}\mathbf{v} = -\mathbf{v}$, whence $\mathbf{v}$ is an eigenvector with associated eigenvalue -1.

Observe that since $\mathbf{v} = \mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3$, this can only be the case since the sum of entries in each row of A is -1. More generally, $\mathbf{e}_1 + \mathbf{e}_2 + \ldots + \mathbf{e}_n$ is an eigenvalue of an $n \times n$ matrix if and only if the sum of all entries in the rows of the matrix equal a constant λ, which is then the eigenvalue for $\mathbf{v}$.

(b) Let $\chi_{\mathrm{A}}(\lambda) = \det\left(\mathrm{A} - \lambda \mathrm{I}_3\right)$ be the characteristic polynomial. To find $\chi_{\mathrm{A}}(\lambda)$, we thus need to calculate the determinant of the 3×3 matrix $\mathrm{A} - \lambda \mathrm{I}_3$.

A 3×3 example

Example

(b) (continued.)

$$\chi_A(\lambda) = \begin{vmatrix} 2-\lambda & -2 & -1 \\ -1 & 1-\lambda & -1 \\ -1 & -2 & 2-\lambda \end{vmatrix}$$

A 3 × 3 example

Example

(b) (continued.)

$$\begin{aligned}\chi_A(\lambda) &= \begin{vmatrix} 2-\lambda & -2 & -1 \\ -1 & 1-\lambda & -1 \\ -1 & -2 & 2-\lambda \end{vmatrix} \\ &= (2-\lambda)(1-\lambda)(2-\lambda) - 2 - 2 \\ &\qquad - (1-\lambda) - 2(2-\lambda) - 2(2-\lambda)\end{aligned}$$

A 3×3 example

Example

(b) (continued.)

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

A 3×3 example

Example

(c) Any eigenvalue λ of A satisfies the characteristic equation, and thus is a root of the characteristic polynomial $\chi_{\mathrm{A}}(\lambda)$.

A 3×3 example

Example

(c) Any eigenvalue λ of A satisfies the characteristic equation, and thus is a root of the characteristic polynomial $\chi_{\mathrm{A}}(\lambda)$. Thus, we seek solutions of the polynomial equation

$$0 = \chi_{\mathrm{A}}(\lambda) \implies 0 = -\lambda^3 + 5\lambda^2 - 3\lambda - 9, \text{ or equivalently,}$$

$$0 = \lambda^3 - 5\lambda^2 + 3\lambda + 9$$

A 3×3 example

Example

(c) Any eigenvalue λ of A satisfies the characteristic equation, and thus is a root of the characteristic polynomial $\chi_{\mathrm{A}}(\lambda)$. Thus, we seek solutions of the polynomial equation

$$0 = \chi_{\mathrm{A}}(\lambda) \implies 0 = -\lambda^3 + 5\lambda^2 - 3\lambda - 9, \text{ or equivalently,}$$
$$0 = \lambda^3 - 5\lambda^2 + 3\lambda + 9$$

We already know -1 is an eigenvalue, so we can divide by $\lambda + 1$ to obtain a quadratic:

$$0 = \lambda^2 - 6\lambda + 9 = (\lambda - 3)^2 .$$

A 3×3 example

Example

(c) Any eigenvalue λ of A satisfies the characteristic equation, and thus is a root of the characteristic polynomial $\chi_{\mathrm{A}}(\lambda)$. Thus, we seek solutions of the polynomial equation

$$0 = \chi_{\mathrm{A}}(\lambda) \implies 0 = -\lambda^3 + 5\lambda^2 - 3\lambda - 9, \text{ or equivalently,}$$
$$0 = \lambda^3 - 5\lambda^2 + 3\lambda + 9$$

We already know -1 is an eigenvalue, so we can divide by $\lambda + 1$ to obtain a quadratic:

$$0 = \lambda^2 - 6\lambda + 9 = (\lambda - 3)^2 .$$

Thus there is precisely one other eigenvalue, $\lambda = 3$.

A 3×3 example

Example

(c) (continued.) To find the associated eigenvector(s), we need to solve the homogeneous system $(\mathrm{A} - 3\mathrm{I}_3)\mathbf{x} = \mathbf{0}$.

$$\mathrm{RREF}\left[\begin{array}{ccc|c} -1 & -2 & -1 & 0 \\ -1 & -2 & -1 & 0 \\ -1 & -2 & -1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 2 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

The solutions thus have the form

$$\mathbf{x} = \begin{bmatrix} -2s - t \\ s \\ t \end{bmatrix} = s\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} + t\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

A 3×3 example

Example

(c) (continued.) Observe that this solution space is merely the plane with equation $x_1 + 2x_2 + x_3 = 0$, and it is spanned by the vectors

$$\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

Thus the 3-eigenspace is two dimensional.

Repeated Eigenvalues: Multiplicities

Remark

In the preceding example, the eigenvalue 3 appeared as a double root of the characteristic polynomial. We say that 3 has *algebraic multiplicity* 2.

The associated eigenspace was spanned by two independent eigenvectors, so the eigenvalue 3 is said in this case to also have *geometric multiplicity* 2.

Multiplicities Defined

Definition

Let $\mathrm{A} \in \mathbb{R}^{n\times n}$ be a real square matrix with characteristic polynomial $\chi_{\mathrm{A}}(\lambda)$. Suppose $\nu \in \mathbb{R}$ is an eigenvalue of A, so $\chi_{\mathrm{A}}(\nu) = 0$. Let $E_\nu := \{\mathbf{x} \in \mathbb{R}^n \mid \mathrm{A}\mathbf{x} = \nu\mathbf{x}\} \subseteq \mathbb{R}^n$ be the ν-eigenspace.

- The *algebraic multiplicity* $m := m(\nu)$ of the eigenvalue ν is the largest integer such that $\lambda - \nu$ divides $\chi_{\mathrm{A}}(\lambda)$:
$$\chi_{\mathrm{A}}(\lambda) = (\lambda - \nu)^m q(\lambda),$$
where $q(\lambda)$ is a polynomial of degree $n - m$ with $q(\nu) \neq 0$.
- The *geometric multiplicity* $\mu := \mu(\nu)$ of the eigenvalue ν is the dimension of E_ν: $\mu_\nu = \dim E_\nu$.

Algebraic versus Geometric Multiplicity

An important question, whose answer is relevant for our forthcoming discussion of *similar matrices* and *diagonalization* is the following:

For a given real eigenvalue ν of a real $n \times n$ matrix A, are $m(\nu)$ and $\mu(\nu)$ equal?

A little thought about previous examples shows they are not. Indeed, consider the shearing transform of $\mathbb{R}^2$ discussed above.

Let $k \in \mathbb{R}$ and recall that the matrix $\mathrm{A} = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix}$ has eigenvalue $\lambda = 1$ with algebraic multiplicity 2.

If $k \neq 0$, then the 1-eigenspace is the line $\mathrm{Span}\{\mathbf{e}_1\}$ whence the geometric multiplicity of $\lambda = 1$ is 1.

On the other hand, if $k = 0$, then the algebraic and geometric multiplicity are equal, since the entire plane $\mathbb{R}^2$ becomes the 1-eigenspace.

Thus a discrepancy can occur between algebraic and geometric multiplicities of eigenvalues.

An Inequality

Proposition

An eigenvalue ν of an $n \times n$ matrix A has algebraic multiplicity at least as large as its geometric multiplicity:

$$1 \leq \mu(\nu) \leq m(\nu) \leq n\,.$$

The inequalities $1 \leq \mu(\nu)$ and $m(\nu) \leq n$ should be clear. The interesting thing to prove is that $\mu(\nu) \leq m(\nu)$.

Before we prove this, we introduce a useful equivalence relation for square matrices, called *similarity*, which implies strong relationships between the eigendata of matrices among a given similarity equivalence class.

An Equivalence Relation: Similarity of Matrices

Definition

Given two $n \times n$ matrices A and B, the matrix A is said to be *similar to* B if there exists an invertible matrix P such that $\mathrm{A} = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}$.

Observe that if A is similar to B via some invertible P, then taking $\mathrm{Q} = \mathrm{P}^{-1}$, one has $\mathrm{B} = \mathrm{Q}\mathrm{A}\mathrm{Q}^{-1}$, whence B is similar to A. Thus we can say unambiguously that A and B are *similar matrices*.

It is easy to check the remaining conditions to show that similarity is an equivalence relation of square matrices: convince yourself that A is always similar to itself, and that if A is similar to B, and B is similar to C, then A and C are also similar.

Similarity and Characteristic Polynomials

Similar matrices are not necessarily row equivalent, but there is a relationship between their characteristic polynomials, and correspondingly, their eigenvalues:

Theorem

Let A and B be similar matrices. Then:

- $\chi_{\mathrm{A}} = \chi_{\mathrm{B}}$, *and thus A and B share eigenvalues and respective algebraic multiplicities,*
- *for any eigenvalue λ of A and B, the geometric multiplicity of λ for A is the same as for B.*

A Proof

Proof.

Assume $\mathrm{A} = \mathrm{PBP}^{-1}$ for some invertible matrix $\mathrm{P} \in \mathbb{R}^{n\times n}$. Observe that

$$\mathrm{A}-\lambda\mathrm{I}_n = \mathrm{PBP}^{-1}-\lambda\mathrm{PP}^{-1}$$

A Proof

Proof.

Assume $\mathrm{A} = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}$ for some invertible matrix $\mathrm{P} \in \mathbb{R}^{n\times n}$. Observe that

$$\mathrm{A} - \lambda \mathrm{I}_n = \mathrm{P}\mathrm{B}\mathrm{P}^{-1} - \lambda \mathrm{P}\mathrm{P}^{-1} = \mathrm{P}\mathrm{B}\mathrm{P}^{-1} - \mathrm{P}\lambda \mathrm{I}_n \mathrm{P}^{-1}$$

A Proof

Proof.

Assume $\mathrm{A} = \mathrm{PBP}^{-1}$ for some invertible matrix $\mathrm{P} \in \mathbb{R}^{n\times n}$. Observe that

$\mathrm{A}-\lambda\mathrm{I}_n = \mathrm{PBP}^{-1}-\lambda\mathrm{PP}^{-1} = \mathrm{PBP}^{-1}-\mathrm{P}\lambda\mathrm{I}_n\mathrm{P}^{-1} = \mathrm{P}(\mathrm{B}-\lambda\mathrm{I}_n)\mathrm{P}^{-1}$,

whence

$$\chi_{\mathrm{A}}(\lambda) = \det(A - \lambda\mathrm{I}_n)$$

A Proof

Proof.

Assume $\mathrm{A} = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}$ for some invertible matrix $\mathrm{P} \in \mathbb{R}^{n\times n}$. Observe that

$\mathrm{A}-\lambda\mathrm{I}_n = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}-\lambda\mathrm{P}\mathrm{P}^{-1} = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}-\mathrm{P}\lambda\mathrm{I}_n\mathrm{P}^{-1} = \mathrm{P}(\mathrm{B}-\lambda\mathrm{I}_n)\mathrm{P}^{-1}$,

whence

$$\begin{aligned}\chi_{\mathrm{A}}(\lambda) &= \det(A - \lambda\mathrm{I}_n)\\ &= \det(\mathrm{P}(\mathrm{B} - \lambda\mathrm{I}_n)\mathrm{P}^{-1})\end{aligned}$$

A Proof

Proof.

Assume $\mathrm{A} = \mathrm{PBP}^{-1}$ for some invertible matrix $\mathrm{P} \in \mathbb{R}^{n\times n}$. Observe that

$\mathrm{A}-\lambda\mathrm{I}_n = \mathrm{PBP}^{-1}-\lambda\mathrm{PP}^{-1} = \mathrm{PBP}^{-1}-\mathrm{P}\lambda\mathrm{I}_n\mathrm{P}^{-1} = \mathrm{P}(\mathrm{B}-\lambda\mathrm{I}_n)\mathrm{P}^{-1}$,

whence

$$\begin{aligned}\chi_{\mathrm{A}}(\lambda) &= \det(A - \lambda\mathrm{I}_n)\\ &= \det(\mathrm{P}(\mathrm{B} - \lambda\mathrm{I}_n)\mathrm{P}^{-1})\\ &= \det(\mathrm{P})\det(\mathrm{B} - \lambda\mathrm{I}_n)\det(\mathrm{P}^{-1})\end{aligned}$$

A Proof

Proof.

Assume $\mathrm{A} = \mathrm{PBP}^{-1}$ for some invertible matrix $\mathrm{P} \in \mathbb{R}^{n\times n}$.
Observe that

$$\mathrm{A}-\lambda\mathrm{I}_n = \mathrm{PBP}^{-1}-\lambda\mathrm{PP}^{-1} = \mathrm{PBP}^{-1}-\mathrm{P}\lambda\mathrm{I}_n\mathrm{P}^{-1} = \mathrm{P}(\mathrm{B}-\lambda\mathrm{I}_n)\mathrm{P}^{-1},$$

whence

$$\begin{aligned}\chi_{\mathrm{A}}(\lambda) &= \det(A-\lambda\mathrm{I}_n)\\ &= \det(\mathrm{P}(\mathrm{B}-\lambda\mathrm{I}_n)\mathrm{P}^{-1})\\ &= \det(\mathrm{P})\det(\mathrm{B}-\lambda\mathrm{I}_n)\det(\mathrm{P}^{-1})\\ &= \det(\mathrm{B}-\lambda\mathrm{I}_n)\end{aligned}$$

A Proof

Proof.

Assume $\mathrm{A} = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}$ for some invertible matrix $\mathrm{P} \in \mathbb{R}^{n\times n}$. Observe that

$$\mathrm{A}-\lambda\mathrm{I}_n = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}-\lambda\mathrm{P}\mathrm{P}^{-1} = \mathrm{P}\mathrm{B}\mathrm{P}^{-1}-\mathrm{P}\lambda\mathrm{I}_n\mathrm{P}^{-1} = \mathrm{P}(\mathrm{B}-\lambda\mathrm{I}_n)\mathrm{P}^{-1},$$

whence

$$\begin{aligned}
\chi_{\mathrm{A}}(\lambda) &= \det(A - \lambda\mathrm{I}_n)\\
&= \det(\mathrm{P}(\mathrm{B} - \lambda\mathrm{I}_n)\mathrm{P}^{-1})\\
&= \det(\mathrm{P})\det(\mathrm{B} - \lambda\mathrm{I}_n)\det(\mathrm{P}^{-1})\\
&= \det(\mathrm{B} - \lambda\mathrm{I}_n)\\
&= \chi_{\mathrm{B}}(\lambda)\,.
\end{aligned}$$

Proof (continued.)

Now, suppose λ is an eigenvalue of both A and B, and suppose the geometric multiplicity of λ for A is μ. Then there exist linearly independent vectors $\mathbf{v}_1, \ldots \mathbf{v}_\mu$ spanning the λ-eigenspace of A, and for any $\mathbf{v}_i$, $i = 1, \ldots, \mu$, $\mathrm{A}\mathbf{v}_i = \lambda\mathbf{v}_i$.

Then

$$\lambda \mathrm{P}^{-1}\mathbf{v}_i = \mathrm{P}^{-1}\mathrm{A}\mathbf{v}_i = \mathrm{B}(\mathrm{P}^{-1}\mathbf{v}_i)\,,$$

whence $\mathrm{P}^{-1}\mathbf{v}_i$ is an eigenvector for B with eigenvalue λ. Since P is invertible, the map $\mathbf{x} \mapsto \mathrm{P}^{-1}\mathbf{x}$ is an isomorphism, whence this induces a one-to-one correspondence of eigenvectors of A and B with eigenvalue λ. Thus, the geometric multiplicity of λ for B is also μ. $\square$

Proving the inequality

We can now prove the inequality $\mu(\nu) \leq m(\nu)$ for an eigenvalue ν of an $n \times n$ matrix A.

Proof.

Let ν be an eigenvalue of A with geometric multiplicity $\mu := \mu(\nu)$. Thus, there exists an eigenbasis $\mathbf{v}_1, \ldots, \mathbf{v}_\mu$ spanning the ν-eigenspace E_ν,

Proving the inequality

We can now prove the inequality $\mu(\nu) \leq m(\nu)$ for an eigenvalue ν of an $n \times n$ matrix A.

Proof.

Let ν be an eigenvalue of A with geometric multiplicity $\mu := \mu(\nu)$. Thus, there exists an eigenbasis $\mathbf{v}_1, \ldots, \mathbf{v}_\mu$ spanning the ν-eigenspace E_ν, and this eigenbasis can be extended to a basis $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_\mu, \mathbf{u}_1, \ldots \mathbf{u}_{n-\mu}\}$ of $\mathbb{R}^n$.

Proving the inequality

We can now prove the inequality $\mu(\nu) \leq m(\nu)$ for an eigenvalue ν of an $n \times n$ matrix A.

Proof.

Let ν be an eigenvalue of A with geometric multiplicity $\mu := \mu(\nu)$. Thus, there exists an eigenbasis $\mathbf{v}_1, \ldots, \mathbf{v}_\mu$ spanning the ν-eigenspace E_ν, and this eigenbasis can be extended to a basis $\mathscr{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_\mu, \mathbf{u}_1, \ldots \mathbf{u}_{n-\mu}\}$ of $\mathbb{R}^n$.

Let

$$\mathrm{P} = \left[\begin{array}{cccccc} \mathbf{v}_1 & \ldots & \mathbf{v}_\mu & \mathbf{u}_1 & \ldots & \mathbf{u}_{n-\mu} \end{array}\right].$$

Proving the inequality

We can now prove the inequality $\mu(\nu) \leq m(\nu)$ for an eigenvalue ν of an $n \times n$ matrix A.

Proof.

Let ν be an eigenvalue of A with geometric multiplicity $\mu := \mu(\nu)$. Thus, there exists an eigenbasis $\mathbf{v}_1, \dots, \mathbf{v}_\mu$ spanning the ν-eigenspace E_ν, and this eigenbasis can be extended to a basis $\mathscr{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_\mu, \mathbf{u}_1, \dots \mathbf{u}_{n-\mu}\}$ of $\mathbb{R}^n$.

Let

$$\mathrm{P} = \left[\begin{array}{cccccc} \mathbf{v}_1 & \dots & \mathbf{v}_\mu & \mathbf{u}_1 & \dots & \mathbf{u}_{n-\mu} \end{array}\right].$$

Consider the product AP.

Proof (continued.)

$$\mathrm{AP} = \begin{bmatrix} \mathrm{A}\mathbf{v}_1 & \dots & \mathrm{A}\mathbf{v}_\mu & \mathrm{A}\mathbf{u}_1 & \dots & \mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix}$$

Proof (continued.)

$$\begin{aligned} AP &= \begin{bmatrix} A\mathbf{v}_1 & \dots & A\mathbf{v}_\mu & A\mathbf{u}_1 & \dots & A\mathbf{u}_{n-\mu} \end{bmatrix} \\ &= \begin{bmatrix} \nu\mathbf{v}_1 & \dots & \nu\mathbf{v}_\mu & A\mathbf{u}_1 & \dots & A\mathbf{u}_{n-\mu} \end{bmatrix} \end{aligned}$$

Proof (continued.)

$$\begin{aligned} \mathrm{AP} &= \begin{bmatrix} \mathrm{A}\mathbf{v}_1 & \dots & \mathrm{A}\mathbf{v}_\mu & \mathrm{A}\mathbf{u}_1 & \dots & \mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix} \\ &= \begin{bmatrix} \nu\mathbf{v}_1 & \dots & \nu\mathbf{v}_\mu & \mathrm{A}\mathbf{u}_1 & \dots & \mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix} \end{aligned}$$

Observe that since $\mathscr{B}$ is a basis, P is invertible, whence we can compute $\mathrm{P}^{-1}\mathrm{AP}$:

Proof (continued.)

$$\mathrm{AP} = \begin{bmatrix} \mathrm{A}\mathbf{v}_1 & \dots & \mathrm{A}\mathbf{v}_\mu & \mathrm{A}\mathbf{u}_1 & \dots & \mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix}$$
$$= \begin{bmatrix} \nu\mathbf{v}_1 & \dots & \nu\mathbf{v}_\mu & \mathrm{A}\mathbf{u}_1 & \dots & \mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix}$$

Observe that since $\mathscr{B}$ is a basis, P is invertible, whence we can compute $\mathrm{P}^{-1}\mathrm{AP}$:

$$\mathrm{P}^{-1}\mathrm{AP} = \begin{bmatrix} \nu\mathbf{e}_1 & \dots & \nu\mathbf{e}_\mu & \mathrm{P}^{-1}\mathrm{A}\mathbf{u}_1 & \dots & \mathrm{P}^{-1}\mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix}$$

Proof (continued.)

$$\begin{aligned} \mathrm{AP} &= \begin{bmatrix} \mathrm{A}\mathbf{v}_1 & \dots & \mathrm{A}\mathbf{v}_\mu & \mathrm{A}\mathbf{u}_1 & \dots & \mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix} \\ &= \begin{bmatrix} \nu\mathbf{v}_1 & \dots & \nu\mathbf{v}_\mu & \mathrm{A}\mathbf{u}_1 & \dots & \mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix} \end{aligned}$$

Observe that since $\mathscr{B}$ is a basis, P is invertible, whence we can compute $\mathrm{P}^{-1}\mathrm{AP}$:

$$\begin{aligned} \mathrm{P}^{-1}\mathrm{AP} &= \begin{bmatrix} \nu\mathbf{e}_1 & \dots & \nu\mathbf{e}_\mu & \mathrm{P}^{-1}\mathrm{A}\mathbf{u}_1 & \dots & \mathrm{P}^{-1}\mathrm{A}\mathbf{u}_{n-\mu} \end{bmatrix} \\ &= \left[\begin{array}{c|c} \nu\mathrm{I}_\mu & * \\ \hline \mathbf{0}_{(n-\mu)\times(n-\mu)} & * \end{array}\right]. \end{aligned}$$

Proof (continued.)

Since there is a diagonal block of $\nu \mathrm{I}_\mu$ in $\mathrm{P}^{-1}\mathrm{AP}$, we see that $\mathrm{P}^{-1}\mathrm{AP}$ has a factor of $(\nu - \lambda)^\mu$ in its characteristic polynomial.

Proof (continued.)

Since there is a diagonal block of $\nu \mathrm{I}_\mu$ in $\mathrm{P}^{-1}\mathrm{AP}$, we see that $\mathrm{P}^{-1}\mathrm{AP}$ has a factor of $(\nu - \lambda)^\mu$ in its characteristic polynomial.

But since A and $\mathrm{P}^{-1}\mathrm{AP}$ are similar, they share the same characteristic polynomial.

Proof (continued.)

Since there is a diagonal block of $\nu \mathrm{I}_\mu$ in $\mathrm{P}^{-1}\mathrm{AP}$, we see that $\mathrm{P}^{-1}\mathrm{AP}$ has a factor of $(\nu - \lambda)^\mu$ in its characteristic polynomial.

But since A and $\mathrm{P}^{-1}\mathrm{AP}$ are similar, they share the same characteristic polynomial.

Thus, the algebraic multiplicity $m(\nu)$ for the eigenvalue ν of A is at least μ. □

Proof (continued.)

Since there is a diagonal block of $\nu \mathrm{I}_\mu$ in $\mathrm{P}^{-1}\mathrm{AP}$, we see that $\mathrm{P}^{-1}\mathrm{AP}$ has a factor of $(\nu-\lambda)^\mu$ in its characteristic polynomial.

But since A and $\mathrm{P}^{-1}\mathrm{AP}$ are similar, they share the same characteristic polynomial.

Thus, the algebraic multiplicity $m(\nu)$ for the eigenvalue ν of A is at least μ. $\square$

Observation

If $\mu(\nu) = m(\nu)$, then we get a maximal diagonal block $\nu \mathrm{I}_m$ in $\mathrm{P}^{-1}\mathrm{AP}$; if χ_{A} factors completely into a product of terms $(\nu_i - \lambda)^{\mu(\nu_i)}$ with $\sum_i \mu(\nu_i) = n$ for real numbers ν_i, then $\mathrm{P}^{-1}\mathrm{AP}$ will be a completely diagonal matrix. We'll study the process of *diagonalization* shortly.

Linear transformations $T : \mathbb{R}^2 \to \mathbb{R}^2$

We'll briefly discuss the role of eigenanalysis in studying the geometry of linear transformations of the plane $\mathbb{R}^2$.

First, we remark that there is a dichotomy: linear maps $T : \mathbb{R}^2 \to \mathbb{R}^2$ are either invertible or non-invertible. We know that the map $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ is non-invertible if and only if A is singular, if and only if 0 is an eigenvalue of A.

Thus, let us first understand the geometry of maps $\mathbf{x} \to \mathrm{A}\mathbf{x}$ where A has a 0 eigenvalue.

Projections

Since χ_{A} is a degree two polynomial for any $\mathrm{A} \in \mathbb{R}^{2\times 2}$, there are two possibilities for zero eigenvalues: a single zero eigenvalue and one nonzero eigenvalue λ, or a zero eigenvalue with algebraic multiplicity $m = 2$.

If the eigenvalues of A are 0 and $\lambda \neq 0$, then A is similar to $\begin{bmatrix} \lambda & 0 \\ 0 & 0 \end{bmatrix}$, which represents a *stretched projection map* $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ projecting onto its nonzero eigenspace E_λ, with stretching factor λ:

- if $\lambda = 1$ then T is an unstretched orthogonal or oblique projection onto the eigenline E_λ,
- if $|\lambda| < 1$ then T is a contracted projection onto E_λ,
- if $|\lambda| > 1$ then T is a dilated projection onto E_λ,
- if $\lambda < 0$ then T is a additionally acts by reflection, reversing the eigenline E_λ.

Nilpotent maps

In the case of a zero eigenvalue of algebraic multiplicity 2, there are two possibilities: the zero matrix, or a *nilpotent matrix*. Nilpotent matrices are (nonzero) square matrices $\mathrm{N} \in \mathbb{R}^{n\times n}$ for which there exists a positive integer power r such that $\mathrm{N}^r = \mathbf{0}_{n\times n}$.

Every 2×2 nilpotent matrix is similar to $\mathrm{N} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$.

Can you give a geometric interpretation of nilpotence, given the similarity to N? (Once we study diagonalization, you will hopefully see how to show this claim, and interpret it...)

Linear Automorphisms of the Plane

Now we can examine invertible matrices A, which always determine linear automorphisms of $\mathbb{R}^2$.

This classification is more subtle than the classification of singular A, and will require some additional results from the theory of diagonalization and the theory of complex eigenvalues, which we will visit later.

To get a better picture, we first examine the general form of the characteristic polynomial of a 2×2 matrix.

The Characteristic Polynomial of $\mathrm{A} \in \mathbb{R}^{2\times 2}$

If $\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then the characteristic polynomial satisfies

$$\chi_{\mathrm{A}}(\lambda) = \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21}$$

The Characteristic Polynomial of $A \in \mathbb{R}^{2\times 2}$

If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then the characteristic polynomial satisfies

$$\begin{aligned} \chi_A(\lambda) &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} \\ &= \lambda^2 - \operatorname{tr}(A)\lambda + \det(A) \end{aligned}$$

where $\operatorname{tr}(A) = a_{11} + a_{22}$ is called the trace of the matrix A.

The Characteristic Polynomial of $\mathrm{A} \in \mathbb{R}^{2\times 2}$

If $\mathrm{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, then the characteristic polynomial satisfies

$$\begin{aligned} \chi_{\mathrm{A}}(\lambda) &= \lambda^2 - (a_{11} + a_{22})\lambda + a_{11}a_{22} - a_{12}a_{21} \\ &= \lambda^2 - \mathrm{tr}\,(\mathrm{A})\lambda + \det(\mathrm{A}) \end{aligned}$$

where $\mathrm{tr}\,(\mathrm{A}) = a_{11} + a_{22}$ is called the trace of the matrix A.

By the fundamental theorem of algebra, the characteristic polynomial factors into linear factors as $\chi_{\mathrm{A}}(\lambda) = (\lambda - \nu_1)(\lambda - \nu_2)$ where ν_1, ν_2 may be complex and are not necessarily distinct.

Eigenvalues, Trace, and Determinant

Multiplying this factorization out, observe that

$$\lambda^2 - (\nu_1 + \nu_2)\lambda + \nu_1\nu_2 = \lambda^2 - \operatorname{trace}(\mathrm{A})\lambda + \det \mathrm{A}\,,$$

whence $\nu_1\nu_2 = \det \mathrm{A}$ and $\nu_1 + \nu_2 = \operatorname{tr} \mathrm{A}$. That is, the product of the eigenvalues is the determinant, and the sum of the eigenvalues is the trace. This rule holds generally for any size of square matrix.

By the quadratic formula, we can also express the eigenvalues of a 2×2 matrix directly in terms of its trace and determinant.

The following proposition gives the explicit formulae, and describes easily proved results characterizing eigendata and geometry of a linear map $\mathbf{x} \mapsto \mathrm{A}\mathbf{x}$.

Proposition

Let A *be a matrix determining a linear map* $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ *of* $\mathbb{R}^2$ *and let* $\Delta = \det(\mathrm{A})$, $\tau = \mathrm{tr}\,(\mathrm{A})$. *Then the eigenvalues of* A *are*

$$\lambda_+ := \frac{1}{2}\left(\tau + \sqrt{\tau^2 - 4\Delta}\right), \lambda_- := \frac{1}{2}\left(\tau - \sqrt{\tau^2 - 4\Delta}\right).$$

- A *has a repeated eigenvalue if and only if* $\tau = \pm 2\sqrt{\Delta}$, *and otherwise has two distinct eigenvalues.*
- A *has a zero eigenvalue if and only if* $\Delta = 0$, *and if in addition* $\tau = 0$, *then the matrix is either nilpotent or the zero matrix.*
- *If* $\tau^2 \geq 4\Delta$ *then the eigenvalues* $\lambda_\pm$ *are real. Otherwise, if* $\tau^2 < 4\Delta$ *then the matrix has distinct complex eigenvalues with strictly nonzero imaginary parts, occuring as a conjugate pair* $\lambda = a + bi$, $\bar{\lambda} = a - bi$. *Moreover, the determinant in this case is* $|\lambda|^2 = \lambda\bar{\lambda} = a^2 + b^2 > 0$.

Proposition (Proposition (continued.))

- *T is area preserving if and only if $|\Delta| = 1$, contracts areas if and only if $|\Delta| < 1$, and expands areas if and only if $|\Delta| > 1$.*
- *Assuming no zero eigenvalues, T is orientation preserving if and only if $\Delta > 0$, and orientation reversing if and only if $\Delta < 0$. If there is one zero eigenvalue and one nonzero eigenvalue λ, it reverses the eigenline if and only if $\lambda < 0$.*

Henceforth, assume that the eigenvalues of A are both nonzero, so $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ is an automorphism of $\mathbb{R}^2$.

We'll characterize the possible geometric actions of this map from the eigendata.

We first consider repeated eigenvalues, where the possibilities are quite limited. We then investigate distinct eigenvalues.

Generalized Shearing

In the repeated eigenvalue case with eigenvalue λ of algebraic multiplicity 2, the matrix is either $\pm\mathrm{I}_2$, a contraction or dilation matrix obtained from scaling $\pm\mathrm{I}_2$ by λ, or the matrix of a *generalized shearing map*.

A generalized shearing map with eigenvalue λ is a map $\mathbf{x} \to \mathrm{A}\mathbf{x}$ such that $m(\lambda) = 2$ but $\mu(\lambda) = 1$ and such that A similar to a matrix of the form

$$\mathrm{J}_\lambda = \lambda \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

The invertible matrix $\mathrm{P} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix}$ giving the similarity $\mathrm{A} = \mathrm{P}\mathrm{J}_\lambda\mathrm{P}^{-1}$ consists of a vector $\mathbf{b}_1$ spanning E_λ and a vector $\mathbf{b}_2$ which is the pre-image of $\lambda\mathbf{b}_1$ under the transformation $\mathbf{x} \mapsto (A - \lambda\mathrm{I}_2)\mathbf{x}$.

The Geometry of Plane Linear Automorphisms

Next we consider the case where A has two distinct eigenvalues. The map $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$ may then be classified by the following geometric considerations:

1. the effect of the map on areas,
2. the effect of the map on orientations,
3. the effect of the map on distances from the origin,
4. the existence or nonexistence of an *eigenframe*, or equivalently, the nonexistence or existence respectively of a minimum rotation angle between $\mathbf{x}$ and the line spanned by its image $T(\mathbf{x})$.

We'll unpack each of these effects via conditions on the eigenvalues.

The Geometry of Plane Linear Automorphisms

It's clear that condition (4) is related to whether the eigenvalues are real or complex.

If the eigenvalues are real, since they are assumed distinct, we know there are two linearly independent eigenvectors spanning distinct eigenlines.

Any such pair gives an *eigenframe*, which is a frame of vectors giving a basis of $\mathbb{R}^2$ such that it remains invariant under the action of the linear map $T(\mathbf{x}) = \mathrm{A}\mathbf{x}$.

We thus will need to examine the other geometric effects to understand the map.

The Geometry of Plane Linear Automorphisms

On the other hand, if the eigenvalues are complex, we'll be able to prove that all vectors are rotated by T (not necessarily equally) and there will be some minimum rotation angle between a vector and the line spanned by its image.

The map will necessarily be orientation preserving (as the determinant is positive), but the other geometric considerations still apply.

In every case, knowing the eigenvalues, we can construct a matrix similar to A which captures the essential geometry in a suitable coordinate system.

The Geometry of Plane Linear Automorphisms

We summarize the results in a table:

▸ Click here for the table!

Definition

An *n*-th order recurrence relation is a discrete relation of the form

$$x_k = f\left(x_{k-n}, x_{k-n+1}, \ldots x_{k-1}\right),$$

for integers $k \geq n$ where f is some function.

Such a relation, if solvable, defines a sequence $\{x_k : k \in \mathbb{Z}_{\geq 0}\}$ determined by the first n terms $\{x_0, \ldots, x_{n-1}\}$.

An initial value recurrence problem for such a recurrence relation is given if one knows the function f, and the values of the first n terms $x_0, x_1, \ldots, x_{n-1}$, and wishes to solve the recurrence to express the general term x_k, $k \geq n$ as a function of k.

Definition

An n-th order recurrence is *linear homogeneous* if f is a homogeneous linear function, i.e., if the recurrence relation is of the form

$$x_k = a_0 x_{k-n} + a_1 x_{k-n+1} + \ldots + a_{n-1} x_{k-1} = \textstyle\sum_{i=0}^{n-1} a_i x_{k-n+i}\,,$$

for numbers $a_0, \ldots, a_{n-1}$.

Observe that for an n-th order linear recurrence, x_n satisfies

$$x_n = a_0 x_0 + a_1 x_1 + \ldots + a_{n-1} x_{n-1}\,,$$

whence the sequence $\{x_k : k \in \mathbb{Z}_{\geq 0}\}$ is determined uniquely by the initial values $x_0, x_1, \ldots x_{n-1}$.

Observe that a linear recurrence can be written in the form

$$x_k = \mathbf{a}\cdot\mathbf{x}_{k-1} = \mathbf{a}^{\mathrm{t}}\mathbf{x}_{k-1}\,, \quad \mathbf{a} := \begin{bmatrix} a_0 \\ \vdots \\ a_{n-1} \end{bmatrix}, \quad \mathbf{x}_{k-1} := \begin{bmatrix} x_{k-n} \\ \vdots \\ x_{k-1} \end{bmatrix}.$$

We can define a vector $\mathbf{x}_k$ consisting of n consecutive terms ending with x_k, and a matrix C, called a *companion matrix*, such that $\mathbf{x}_k = \mathrm{C}\mathbf{x}_{k-1}$:

$$\mathbf{x}_k := \begin{bmatrix} x_{k-n+1} \\ x_{k-n+2} \\ \vdots \\ x_{k-1} \\ x_k \end{bmatrix}, \ \mathrm{C} := \left[\begin{array}{c|c} \mathbf{0} & \mathrm{I}_n \\ \hline \multicolumn{2}{c}{\mathbf{a}^{\mathrm{t}}} \end{array}\right] = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 \\ a_0 & a_1 & \cdots & \cdots & a_{n-1} \end{bmatrix}.$$

In this formulation, one sees that $\mathbf{x}_{n+k} = \mathrm{C}^{k+1}\mathbf{x}_{n-1}$, for $k \geq -1$. If one can find an invertible matrix P such that C is similar to a diagonal matrix D via P, then one can compute an explicit formula

$$\mathbf{x}_{n+k} = \mathrm{P}\mathrm{D}^{k+1}\mathrm{P}^{-1}\mathbf{x}_{n-1}$$

whose first entry gives an expression for x_k in terms the first n terms $x_0, x_1, \ldots x_{n-1}$ and powers of the entries of D.

In particular, one will see that the eigenvalues and eigenvectors of C build a solution to the linear homogeneous initial value recursion problem.

The Fibonacci Numbers

Definition

The Fibonacci numbers F_k are the numbers defined by the simple linear recurrence

$$F_{k+1} = F_k + F_{k-1}\,, \quad F_0 = 0\,, \quad F_1 = 1\,.$$

The Fibonacci sequence is thus the sequence starting with $0, 1, 1, 2, 3, 5, 8\ldots$ whose next term is always the sum of the preceding two terms

We can get an explicit formula for the k-th Fibonacci number using eigentheory.

From the recurrence relation $F_{k+1} = F_k + F_{k-1}$ with the initial values $F_0 = 0$ and $F_1 = 1$, we can rewrite this as a linear discrete dynamical system of the form $\mathbf{x}_k = \mathrm{C}\mathbf{x}_{k-1}$ where

$$\mathbf{x}_k = \begin{bmatrix} F_{k-1} \\ F_k \end{bmatrix}, \quad \mathrm{C} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

We will diagonalize C in order to solve to obtain an explicit formula for F_k.

The first step is to compute the characteristic polynomial $\chi_{\mathrm{C}}(\lambda)$:

$$\chi_{\mathrm{C}}(\lambda) = \lambda^2 - \lambda - 1\,.$$

We get two real, irrational eigenvalues, $\lambda_\pm = \frac{1\pm\sqrt{5}}{2}$.

We momentarily digress to mention that this polynomial and its roots have been widely studied since antiquity; the positive root λ_+ is the same as the famous *golden ration* $\phi = \frac{1+\sqrt{5}}{2}$. The negative root is just $-1/\phi$.

Observe that ϕ satisfies the useful and interesting relations

$$\phi - 1 = \frac{1}{\phi}, \quad \phi + 1 = \phi^2, \qquad 1 + \phi^2 = \sqrt{5}\phi\,,$$

as well as being given by the amusing (but less useful) formulae

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{\ddots}}} = \sqrt{1 + \sqrt{1 + \sqrt{1 + \ldots}}}\,.$$

Exercise

Show that

$$E_\phi = \mathrm{Span}\left\{\begin{bmatrix} 1 \\ \phi \end{bmatrix}\right\}, \quad E_{-1/\phi} = \mathrm{Span}\left\{\begin{bmatrix} -\phi \\ 1 \end{bmatrix}\right\}.$$

Let $\mathrm{P} = \begin{bmatrix} 1 & -\phi \\ \phi & 1 \end{bmatrix}$. Show that $\mathrm{P}^{-1} = \frac{1}{\sqrt{5}}\begin{bmatrix} 1/\phi & 1 \\ -1 & 1/\phi \end{bmatrix}$, and check that

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & -\phi \\ \phi & 1 \end{bmatrix}\begin{bmatrix} \phi & 0 \\ 0 & -1/\phi \end{bmatrix}\begin{bmatrix} 1/\phi & 1 \\ -1 & 1/\phi \end{bmatrix}$$

Then, using that

$$\begin{bmatrix} F_k \\ F_{k+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^k \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

we have that

$$\begin{aligned} \begin{bmatrix} F_k \\ F_{k+1} \end{bmatrix} &= \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & -\phi \\ \phi & 1 \end{bmatrix} \begin{bmatrix} \phi^k & 0 \\ 0 & (-1/\phi)^k \end{bmatrix} \begin{bmatrix} 1/\phi & 1 \\ -1 & 1/\phi \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{5}} \begin{bmatrix} \phi^{k-1} - (-1/\phi)^{k-1} & \phi^k - (-1/\phi)^k \\ \phi^k - (-1/\phi)^k & \phi^{k+1} - (-1/\phi)^{k+1} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= \frac{1}{\sqrt{5}} \begin{bmatrix} \phi^k - (-1/\phi)^k \\ \phi^{k+1} - (-1/\phi)^{k+1} \end{bmatrix}, \end{aligned}$$

whence

$$F_k = \frac{1}{\sqrt{5}} \left(\phi^k - (-1/\phi)^k \right) = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^k - \left(\frac{1-\sqrt{5}}{2} \right)^k \right).$$

Observe that

$$F_k = \left\lfloor \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^k \right\rceil,$$

where $\lfloor x \rceil$ is the *nearest integer function*.

More generally, n-th order linear recursions can be solved with general solutions that are linear combinations of products of powers of eigenvalues of the associated companion matrix times certain powers of the index.

Challenge Problem: Consider a general homogeneous n-th order linear recurrence of the form $x_n = a_0 x_0 + \ldots a_{n-1}x_{n-1}$.

(1) Show that the polynomial $t^n - \sum_{i=0}^{n-1} a_i t^i$ is the characteristic polynomial of the associated companion matrix C.

(2) For any eigenvalue λ which is a root of order m of the characteristic polynomial, show that $x_k = k^p\lambda^k$, is a solution of the recurrence equation for $p \in \{0, 1, \ldots m-1\}$

Challenge Problem (continued):

(3) Show that the general solution is given by linear combinations of terms of the form $k^p\lambda^k$. That is, show that any solution x_k of the recurrence has form

$$\begin{aligned} x_k &= \sum_{i=1}^{l}\sum_{j=1}^{m_l} b_{i,j}k^{j-1}\lambda_i^k \\ &= \lambda_1^k(b_{1,0} + b_{1,1}k + \dots b_{1,m_1}k^{m_1-1}) \\ &\qquad + \dots \lambda_l^k(b_{l,0} + b_{l,1}k + \dots b_{l,m_l}k^{m_l-1})\ , \end{aligned}$$

where $\lambda_1 \dots \lambda_l$ are distinct eigenvalues of C with respective algebraic multiplicities $m_1, \dots, m_l$, and $b_{i,j}$ are constants.

(4) If one specifies values for $x_0, \dots x_{n-1}$, does this uniquely determine the constants $b_{i,j}$?

An n-dimensional linear system of differential equations is a system of the form

$$\begin{cases} \frac{\mathrm{d}x_1}{\mathrm{d}t} = \sum_{j=1}^{n} a_{1,j}(t)x_j(t) \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} = \sum_{j=1}^{n} a_{2,j}(t)x_j(t) \\ \qquad\vdots \\ \frac{\mathrm{d}x_i}{\mathrm{d}t} = \sum_{j=1}^{n} a_{i,j}(t)x_j(t) \\ \qquad\vdots \\ \frac{\mathrm{d}x_n}{\mathrm{d}t} = \sum_{j=1}^{n} a_{n,j}(t)x_j(t) \end{cases}.$$

For the system to be linear in the variables x_k, we must assert that the coefficients a_{ij} are independent of x_k for all k, i.e. that $\partial a_{ij}/\partial x_k = 0$ for all i, j, and k. The system is called *autonomous* if the coefficients a_{ij} satisfy $\mathrm{d}a_{ij}/\mathrm{d}t = 0$ for all i and j, i.e., if the coefficients are also constant in time.

Such a system can be compactly described by a linear vector differential equation

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}(t) = \mathrm{A}(t)\mathbf{x}(t)\,.$$

Often one has an initial value problem, where at time $t = 0$ one is given $\mathbf{x}(0) = \mathbf{x}_0$ for some constant vector $\mathbf{x}_0 \in \mathbb{R}^n$.
In the case of autonomous systems with a constant coefficient matrix A, one can attempt to construct solutions as linear combinations of the *eigenfunctions* of the form $e^{\lambda t}\mathbf{v}_\lambda$ where λ is an eigenvalue of A and $\mathbf{v}_\lambda$ is an associated eigenvector.

It's not hard to see that such vectors furnish solutions: if λ is an eigenvalue of A and $\mathbf{v}_\lambda$ is the associated eigenvector, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{\lambda t}\mathbf{v}_\lambda\right) = \lambda e^{\lambda t}\mathbf{v}_\lambda\,,$$

while

$$\mathrm{A}\left(e^{\lambda t}\mathbf{v}_\lambda\right) = e^{\lambda t}\mathrm{A}\mathbf{v}_\lambda = e^{\lambda t}(\lambda\mathbf{v}_\lambda) = \lambda e^{\lambda t}\mathbf{v}_\lambda\,,$$

so $e^{\lambda t}\mathbf{v}_\lambda$ satisfies the differential equation $\mathbf{x}'(t) = \mathrm{A}\mathbf{x}(t)$.

The remarkable fact is that when A has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, any solution is an element of

$$\mathrm{Span}\left\{e^{\lambda_1 t}\mathbf{v}_{\lambda_1}, \ldots, e^{\lambda_n t}\mathbf{v}_{\lambda_n}\right\}\,.$$

Example

Consider the linear system of differential equations:

$$\begin{cases} \frac{\mathrm{d}x_1}{\mathrm{d}t} = -x_1 + 2x_2 \\ \frac{\mathrm{d}x_2}{\mathrm{d}t} = 3x_1 + 4x_2 \end{cases} \longleftrightarrow \quad \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) = \overbrace{\begin{bmatrix} -1 & 2 \\ 3 & 4 \end{bmatrix}}^{\mathrm{A}} \mathbf{x}(t)\,.$$

The matrix A has eigenvalues -2 and 5 with respective eigenvectors

$$\mathbf{v}_{-2} = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \text{ and } \mathbf{v}_5 = \begin{bmatrix} 1 \\ 3 \end{bmatrix},$$

whence

$$\mathbf{x}(t) = c_1 e^{-2t}\mathbf{v}_{-2} + c_2 e^{5t}\mathbf{v}_5 \longleftrightarrow \begin{cases} x_1(t) = -2c_1 e^{-2t} + c_2 e^{5t} \\ x_2(t) = c_1 e^{-2t} + 3c_2 e^{5t} \end{cases}$$

gives a general solution.

Example (An Example with Imaginary Eigenvalues)

Consider the second order linear homogeneous differential equation

$$x''(t) + x(t) = 0\,.$$

We can convert it to a first order linear system by introducing a new variable: the *velocity* $v(t) = x'(t)$. The system becomes the matrix equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ v \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} x \\ v \end{bmatrix}\,.$$

The matrix of the system has eigenvalues $\pm i$ with respective complex eigenvectors $\mathbf{v}_i = \begin{bmatrix} 1 \\ i \end{bmatrix}$ and $\mathbf{v}_{-i} = \begin{bmatrix} 1 \\ -i \end{bmatrix}$.

Example (An Example with Imaginary Eigenvalues - continued)

The general solution to the first order system is

$$\begin{bmatrix} x \\ v \end{bmatrix} = c_1 e^{it} \begin{bmatrix} 1 \\ i \end{bmatrix} + c_2 e^{-it} \begin{bmatrix} 1 \\ -i \end{bmatrix}, \quad c_1, c_2 \in \mathbb{C}\,.$$

It follows that a complex solution to the second order equation has the form $x(t) = c_1 e^{it} + c_2 e^{-it}$. Using that $e^{it} = \cos t + i \sin t$, and setting $a = c_1 + c_2$ and $b = i(c_1 - c_2)$, we obtain

$$x(t) = a\cos(t) + b\sin(t)\,.$$

One can determine real coefficients a and b given sufficient real initial conditions are provided, such as real values for $x(t_0)$ and $x'(t_0)$ for some initial time $t_0 \in \mathbb{R}$.

Final Exam Information

The final exam for *all sections* will be held

Monday 5/7/18, 10:30AM-12:30PM, in Boyden gym.