MA1522: Linear Algebra for Computing

Chapter 6: Eigenanalysis

6.1 Eigenvalues and Eigenvectors

Eigenvalues and Eigenvectors

Let **A** be a square matrix of order n.

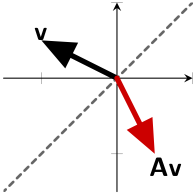
- For any vector **u** in $\mathbb{R}^n$, **Au** is also a vector in $\mathbb{R}^n$.
- So, we may think of **A** as a map $\mathbb{R}^n \to \mathbb{R}^n$, moving vectors around in $\mathbb{R}^n$, $\mathbf{u} \mapsto \mathbf{Au}$.
- That is, **A** defines a linear mapping. More on this in the next chapter.

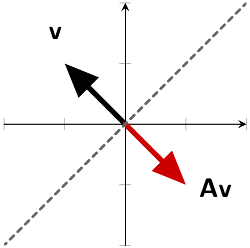
Visit https://www.geogebra.org/m/sbkscz46 to visualize how vectors are moved around by a matrix **A**.

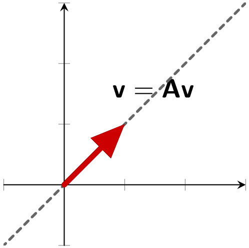
Example

Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Geometrically the matrix $\mathbf{A}$ reflects a vector along the line $x = y$.

$$\mathbf{A}\begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$





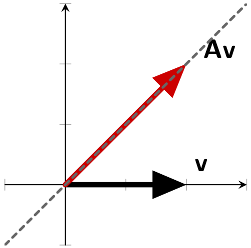


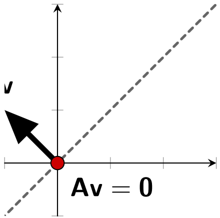
Observe that any vector on the line $x = y$ gets transform back to itself, and any vector along $x = -y$ line get transformed to the negative of itself.

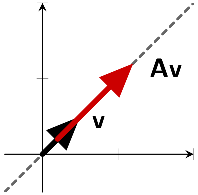
Eigenvalues and Eigenvectors

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. It takes a vector and maps it to a vector along the line $x = y$ such that both coordinates in $\mathbf{Av}$ are the sum of the coordinates in $\mathbf{v}$.

$$\mathbf{A}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \qquad \mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}.$$







Observe that any vector $\mathbf{v}$ along the line $x = y$ is mapped to twice itself, $\mathbf{Av} = 2\mathbf{v}$, and it take any vector $\mathbf{v}$ along the line $x = -y$ to the origin, $\mathbf{Av} = \mathbf{0}$.

Eigenvalues and Eigenvectors

Definition

Let $\mathbf{A}$ be a square matrix of order n. A real number λ is an *eigenvalue* of $\mathbf{A}$ if there is a nonzero vector $\mathbf{v}$ in $\mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$, such that

$$\mathbf{Av} = \lambda\mathbf{v}.$$

In this case, the nonzero vector $\mathbf{v}$ is called an *eigenvector* associated to λ.

Remark

- Geometrically, eigenvectors are the vectors that are being scaled (stretch, dilate, or reflect) when acted upon by $\mathbf{A}$, and eigenvalues are the amount to scale the eigenvectors.
- We require the eigenvector to be nonzero, $\mathbf{v} \neq \mathbf{0}$, for otherwise, the identity

  $$\mathbf{A0} = \lambda\mathbf{0}$$

  holds for every real number λ, which means that every real number is an eigenvalue of $\mathbf{A}$. This renders the definition not meaningful and uninteresting.

Examples

1. For $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\begin{pmatrix} -1 \\ 1 \end{pmatrix}$.

Eigenvalue	Eigenvector
$\lambda = 1$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\lambda = -1$	$\mathbf{v}_\lambda = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$

2. For $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $\mathbf{A}\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} = 0\begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Eigenvalue	Eigenvector
$\lambda = 2$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
$\lambda = 0$	$\mathbf{v}_\lambda = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

Characteristic Polynomial

Definition

Let $\mathbf{A}$ be a square matrix of order n, the *characteristic polynomial* of $\mathbf{A}$, denoted as $\mathrm{char}(\mathbf{A})$, is the degree n polynomial

$$\det(x\mathbf{I} - \mathbf{A}).$$

Example

1. Let $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The characteristic polynomial is $\det\left(\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix} - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right) = \begin{vmatrix} x & -1 \\ -1 & x \end{vmatrix} = x^2 - 1$.
2. Let $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. $\det\left(\begin{pmatrix} x-1 & -1 \\ -1 & x-1 \end{pmatrix}\right) = \begin{vmatrix} x-1 & -1 \\ -1 & x-1 \end{vmatrix} = (x-1)^2 - 1 = x(x-2)$.
3. Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 1 \end{pmatrix}$. $\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & 0 & 0 \\ 0 & x & -2 \\ 0 & -3 & x-1 \end{vmatrix} = (x-1)[x(x-1) - 6] = (x-1)(x+2)(x-3)$.

Finding Eigenvalues

Recall that λ is an eigenvalue if there is a nonzero $\mathbf{v}$ such that $\mathbf{Av} = \lambda\mathbf{v}$. Manipulating the equation, we have

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

which shows that $\mathbf{v}$ is a nontrivial solution to the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$.

Theorem

Let $\mathbf{A}$ be a square matrix of order n. $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathbf{A}$ if and only if the homogeneous system $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions.

Observe that $\lambda\mathbf{I} - \mathbf{A}$ is a square matrix. And so, $(\lambda\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}$ has nontrivial solutions if and only if $(\lambda\mathbf{I} - \mathbf{A})$ is singular, which can be checked using its determinant. Hence,

Theorem

Let $\mathbf{A}$ be a square matrix of order n. λ is an eigenvalue of $\mathbf{A}$ if and only if λ is a root of the characteristic polynomial $\det(x\mathbf{I} - \mathbf{A})$.

Examples

1. $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\det(x\mathbf{I} - \mathbf{A}) = x^2 - 1$. So the eigenvalues of $\mathbf{A}$ are $\lambda = \pm 1$

2. $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $\det(x\mathbf{I} - \mathbf{A}) = x(x-2)$. So the eigenvalues of $\mathbf{A}$ are $\lambda = 0, 2$

3. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 3 & 1 \end{pmatrix}$, $\det(x\mathbf{I} - \mathbf{A}) = (x+2)(x-1)(x-3)$. So the eigenvalues of $\mathbf{A}$ are $\lambda = -2, 1, 3$

Eigenvalue and Invertibility

Question: Can $\lambda = 0$ be an eigenvalue of $\mathbf{A}$?

Suppose 0 is an eigenvalue of $\mathbf{A}$. Let $\mathbf{v}$ be an eigenvector of $\mathbf{A}$ associated to eigenvalue 0. Then $\mathbf{v} \neq \mathbf{0}$ is a nonzero vector such that

$$\mathbf{A}\mathbf{v} = \mathbf{v} = \mathbf{0}.$$

This shows that the homogeneous system $\mathbf{A}\mathbf{x} = \mathbf{0}$ has nontrivial solutions, and hence, $\mathbf{A}$ is singular.

It is easy to see that the converse is true too, that is, if $\mathbf{A}$ is singular, then $\lambda = 0$ is an eigenvalue.

Theorem

A square matrix $\mathbf{A}$ *is invertible if and only if* $\lambda = 0$ *is not an eigenvalue of* $\mathbf{A}$.

We will add this to the equivalent statements of invertibility.

Equivalent Statements of Invertibility

Theorem (Equivalent Statements for Invertibility)

Let $\mathbf{A}$ be a square matrix of order n. The following statements are equivalent.

(i) $\mathbf{A}$ *is invertible.*

(ii) $\mathbf{A}^T$ *is invertible.*

(iii) *(left inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{BA} = \mathbf{I}$.

(iv) *(right inverse) There is a matrix* $\mathbf{B}$ *such that* $\mathbf{AB} = \mathbf{I}$.

(v) *The reduced row-echelon form of* $\mathbf{A}$ *is the identity matrix.*

(vi) $\mathbf{A}$ *can be expressed as a product of elementary matrices.*

(vii) *The homogeneous system* $\mathbf{Ax} = \mathbf{0}$ *has only the trivial solution.*

(viii) *For any* $\mathbf{b}$, *the system* $\mathbf{Ax} = \mathbf{b}$ *has a unique solution.*

(ix) *The determinant of* $\mathbf{A}$ *is nonzero,* $\det(\mathbf{A}) \neq 0$.

(x) *The columns/rows of* $\mathbf{A}$ *are linearly independent.*

(xi) *The columns/rows of* $\mathbf{A}$ *spans* $\mathbb{R}^n$.

(xii) $\text{rank}(\mathbf{A}) = n$ *($\mathbf{A}$ has full rank).*

(xiii) $\text{nullity}(\mathbf{A}) = 0$.

(xiv) 0 *is not an eigenvalue of* $\mathbf{A}$.

Algebraic Multiplicity

Let λ be an eigenvalue of $\mathbf{A}$. The algebraic multiplicity of λ is the largest integer r_λ such that

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda)^{r_\lambda} p(x),$$

for some polynomial $p(x)$. Alternatively, r_λ is the positive integer such that in the above equation, λ is not a root of $p(x)$.

Suppose $\mathbf{A}$ is an order n square matrix such that $\det(x\mathbf{I} - \mathbf{A})$ can be factorize into linear factors completely. Then we can write

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_k)^{r_k}$$

where $r_1 + r_2 + \cdots + r_k = n$, and $\lambda_1, \lambda_2, ..., \lambda_k$ are the distinct eigenvalues of $\mathbf{A}$. Then the algebraic multiplicity of λ_i is r_i for $i = 1, ..., k$.

Examples

1. Let $\mathbf{A} = \mathbf{0}_n$ be the order n zero matrix. Then $\det(x\mathbf{I} - \mathbf{0}) = \det(x\mathbf{I}) = x^n$. $\lambda = 0$ is the only eigenvalue of $\mathbf{A}$, with algebraic multiplicity $r_0 = n$.

2. $\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 \\ 0 & 1 & -2 \\ 0 & 0 & 3 \end{pmatrix}$. $\det(x\mathbf{I} - \mathbf{A}) = (x-1)^2(x-3)$. The eigenvalues of $\mathbf{A}$ are $\lambda = 1, 3$, with algebraic multiplicities $r_1 = 2$, $r_3 = 1$, respectively.

3. $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x-2)^2(x-4)$. The eigenvalues are $\lambda = 2, 4$, with algebraic multiplicities $r_2 = 2$, $r_4 = 1$, respectively.

4. $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$. Then $\det(x\mathbf{I} - \mathbf{A}) = (x-1)(x^2+1)$. The eigenvalue is $\lambda = 1$ only, with algebraic multiplicity $r_1 = 1$. In this case $\mathbf{A}$ has only one (real) eigenvalue.

Eigenvalues of Triangular Matrices

Theorem

The eigenvaules of a triangular matrix are the diagonal entries. The algebraic multiplicity of the eigenvalue is the number of times it appears as a diagonal entry of $\mathbf{A}$.

Proof.

We will prove for the case where $\mathbf{A}$ is an upper triangular matrix. The proof for lower triangular matrix is analogous.

Suppose $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}$. Then

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{pmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ 0 & x - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x - a_{nn} \end{pmatrix} = (x - a_{11})(x - a_{22}) \cdots (x - a_{nn}).$$

$\square$

Eigenspace

Recall that eigenvectors of $\mathbf{A}$ associated to eigenvalue λ are nontrivials solution to

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

Since the system is homogeneous, the set of all solutions is a subspace. We will call it the eigenspace of $\mathbf{A}$ associated to eigenvalue λ.

Definition

Let $\mathbf{A}$ be an order n square matrix. The *eigenspace* associated to an eigenvalue λ of $\mathbf{A}$ is

$$E_\lambda = \{\ \mathbf{v} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{v} = \lambda\mathbf{v}\ \} = \mathrm{Null}(\lambda\mathbf{I} - \mathbf{A}).$$

The *geometric multiplicity* of an eigenvalue λ is the dimension of its eigenspace,

$$\dim(E_\lambda) = \mathrm{nullity}(\lambda\mathbf{I} - \mathbf{A}).$$

Example

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. We will first find the eigenvalues.

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & -1 & 0 \\ -1 & x-1 & 0 \\ 0 & 0 & x-2 \end{vmatrix} = (x-2)((x-1)^2 - 1) = x(x-2)^2.$$

So, the eigenvalues are $\lambda = 0, 2$, with algebraic multiplicities $r_0 = 1, r_2 = 2$, respectively.

Next, we will find a basis for the eigenspaces.

For eigenvalue $\lambda = 0$: $0\mathbf{I} - \mathbf{A} = \begin{pmatrix} -1 & -1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$ So, $E_0 = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$.

for eigenvalue $\lambda = 2$: $2\mathbf{I} - \mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ So, $E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$.

Question

1. Let $\mathbf{A}$ and $\mathbf{B}$ be row equivalent order n square matrices.

   (a) If λ is an eigenvalue of $\mathbf{A}$, is it an eigenvalue of $\mathbf{B}$?

   (b) If $\mathbf{v}$ is an eigenvector of $\mathbf{A}$, is it an eigenvector of $\mathbf{B}$?

2. Can we compute the characteristic polynomial of a square matrix using row reduction instead of cofactor expansion?

Challenge

Let $\mathbf{A}$ be a $n \times n$ matrix.

1. Show that the characteristic polynomial of $\mathbf{A}$ is equal to the characteristic polynomial of $\mathbf{A}^T$. Hence $\mathbf{A}$ and $\mathbf{A}^T$ has the same eigenvalues.

2. Let λ be an eigenvalue of $\mathbf{A}$. Show that the geometric multiplicity of λ as an eigenvalue of $\mathbf{A}$ is equal to its geometric multiplicity as an eigenvalue of $\mathbf{A}^T$.

6.2 Diagonalizaton

Diagonalization

Definition

A square matrix $\mathbf{A}$ of order n is *diagonalizable* if there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$

is a diagonal matrix, OR

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

Examples

1. All zero square matrix is diagonalizable, $\mathbf{0} = \mathbf{I}\mathbf{0}\mathbf{I}^{-1}$.

2. The identity matrix is diagonalizable, $\mathbf{I} = \mathbf{P}\mathbf{I}\mathbf{P}^{-1}$ for any invertible $\mathbf{P}$.

3. Any diagonal matrix $\mathbf{D}$ is diagonalizable, $\mathbf{D} = \mathbf{I}\mathbf{D}\mathbf{I}^{-1}$.

4. $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix}$ is diagonalizable, with $\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}$.

Diagonalization

Previous example:

$$\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}.$$

- $\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. This shows that the first column of **P** is an eigenvector associated to eigenvalue 2, the (1,1) diagonal entry of **D**.
- $\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 2 \end{pmatrix} = 2 \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ This shows that the second column of **P** is an eigenvector associated to eigenvalue 2, the (2,2) diagonal entry of **D**.
- $\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 4 \\ 4 \\ 0 \end{pmatrix} = 4 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ This shows that the third column of **P** is an eigenvector associated to eigenvalue 4, the (3,3) diagonal entry of **D**.

Diagonalization

Theorem (Diagonalizability)

A $n \times n$ square matrix $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ has n linearly independent eigenvectors.

Proof.

First observe that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$ *if and only if* $\mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D}$. *Write* $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$ *and*

$$\mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$ *then*

$$\begin{pmatrix} \mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \cdots & \mathbf{A}\mathbf{u}_n \end{pmatrix} = \mathbf{A}\mathbf{P} = \mathbf{P}\mathbf{D} = \begin{pmatrix} \mu_1\mathbf{u}_1 & \mu_2\mathbf{u}_2 & \cdots & \mu_n\mathbf{u}_n \end{pmatrix},$$

and thus by comparing the columns, we have $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$. *This shows that the columns of* $\mathbf{P}$ *are eigenvectors of* $\mathbf{A}$. *Now* $\mathbf{P}$ *is invertible if and only if its columns form a basis for* $\mathbb{R}^n$. *Hence,* $\mathbf{A}$ *is diagonalizable if and only if we can find a basis of* $\mathbb{R}^n$ *consisting of eigenvectors of* $\mathbf{A}$. $\square$

Diagonalization

That is, $\mathbf{A}$ is diagonalizable if and only if we can find

$$\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where μ_i is the eigenvalue associated to eigenvector $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$.

$\mathbf{P}$ is invertible if and only if $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$.

Note that μ_i may not be distinct.

Not Diagonalizable

Not all square matrices are diagonalizable. For example, consider

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

This is a triangular matrix, with only one eigenvalue $\lambda = 0$.

$$0\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

tells us that $\mathbf{A}$ has only 1 linearly independent eigenvector associated to the only eigenvalue $\lambda = 0$. Hence, $\mathbf{A}$ is not diagonalizable.

Not Diagonalizable

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

$$\det(x\mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-2 & 1 \\ -1 & x \end{vmatrix} = (x-1)^2;$$

$\mathbf{A}$ has only one eigenvalue $\lambda = 1$.

$$\mathbf{I} - \mathbf{A} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$$

tell us that $\mathbf{A}$ has only 1 linearly independent eigenvector. Hence, $\mathbf{A}$ is not diagonalizable.

Notice that in both examples above, the algebraic multiplicities are greater than the geometric multiplicities.

Independence of Eigenspaces

Suppose λ_1 and λ_2 are distinct eigenvalues. Let $\mathbf{v}_1$ be an eigenvector associated to eigenvalue λ_1. Then since $\lambda_1 \neq \lambda_2$,

$$\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \neq \lambda_2\mathbf{v}_1,$$

$\mathbf{v}_1$ cannot be in the eigenspace associated to λ_2. This demonstrates that vectors from different eigenspaces are linearly independent. The proof of the following theorem is given in the appendix.

Theorem (Eigenspaces are linearly independent)

Let $\mathbf{A}$ be a $n \times n$ square matrix. Let λ_1 and λ_2 are distinct eigenvalues of $\mathbf{A}$, $\lambda_1 \neq \lambda_2$. Suppose $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ is a linearly independent subset of eigenspace associated to eigenvalue λ_1, and $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$ is a linearly independent subset of of eigenspace associated to eigenvalue λ_2. Then the union $\{\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{v}_1, ..., \mathbf{v}_m\}$ is linearly independent.

Visualization of Eigenspaces

Click on the following link, https://www.geogebra.org/3d/u87k4uah, to visualize the independence of the he eigenspaces.

1. Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & 2 \end{pmatrix}$. It is a diagonalizable matrix.

2. The eigenvalues are $\lambda = 2, 3$, with algebraic and thus geometric multiplicities $r_2 = \dim(E_2) = 2$ and $r_3 = \dim(E_3) = 1$, respectively.

3. At the side, if we let $c_3 = 0$, then for any scalars c_1, c_2, $\mathbf{w}$ is a vector in the blue plane and $\mathbf{Aw} = 2\mathbf{w}$. If we let $c_1 = c_2 = 0$, $\mathbf{w}$ is a vector in alone the red line and $\mathbf{Aw} = 3\mathbf{w}$.

4. This shows that blue plane is the eigenspace E_2, and the red line is the eigenspace E_3.

5. It is clear from the plot that vectors in the blue plane and the red line are independent.

Question

Suppose $\mathbf{A}$ is a square matrix of order n with distinct eigenvalues $\lambda_1, ..., \lambda_p$, and algebraic multiplicities $r_1, ..., r_p$, respectively. What can you conclude about the sum

$$r_1 + r_2 + \cdots + r_p?$$

Hint: Suppose $\mathbf{A}$ is a square matrix of order n such that the characteristic polynomial splits into linear factors

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_1}(x - \lambda_2)^{r_2} \cdots (x - \lambda_p)^{r_p}.$$

What can you conclude about the sum

$$r_1 + r_2 + \cdots + r_p?$$

Question

Is it possible for the geometric multiplicity to be 0, $\dim(E_\lambda) = 0$?

Equivalent Statements for Diagonalizability

The geometric multiplicity is bounded above by the algebraic multiplicity. The proof can be found in the appendix.

Theorem (Geometric Multiplicity is no greater than Algebraic multiplicity)

The geometric multiplicity of an eigenvalue λ of a square matrix $\mathbf{A}$ is no greater than the algebraic multiplicity, that is,

$$1 \leq \dim(E_\lambda) \leq r_\lambda.$$

Equivalent Statements for Diagonalizability

Let $\mathbf{A}$ be a $n \times n$ matrix. Let $\lambda_1, \lambda_2, ..., \lambda_p$ be the distinct eigenvalues of $\mathbf{A}$ with algebraic multiplicities $r_1, r_2, ..., r_p$ respectively. Let

$$\{\mathbf{v}_{i,1}, ..., \mathbf{v}_{i,k_i}\} \text{ be a basis for } E_{\lambda_i},$$

the eigenspace associated to eigenvalue λ_i, i.e. $\dim(E_{\lambda_i}) = k_i$. Collect the bases together, we get the set $\{\mathbf{v}_{1,1}, ..., \mathbf{v}_{1,k_1}, \mathbf{v}_{2,1}, ..., \mathbf{v}_{2,k_2}, ... \; ..., \mathbf{v}_{p,1}, ..., \mathbf{v}_{p,k_p}\}$. Now

$$\begin{array}{ccccccccc}
k_1 & + & k_2 & + & \cdots & + & k_p & & \\
\| & & \| & & \cdots & & \| & & \\
\dim(E_{\lambda_1}) & + & \dim(E_{\lambda_2}) & + & \cdots & + & \dim(E_{\lambda_k}) & & \\
|\wedge & & |\wedge & & \cdots & & |\wedge & & \\
r_1 & + & r_2 & + & \cdots & + & r_k & \leq & n
\end{array}$$

If $\mathbf{A}$ is diagonalizable, then $k_1 + k_2 + \cdots + k_p = n$. Thus, necessarily $r_1 + r_2 + \cdots + r_p = n$ and $\dim(E_{\lambda_i}) = r_i$. On the other hand, if $r_1 + r_2 + \cdots + r_p = n$ and $\dim(E_{\lambda_i}) = r_i$, then $k_1 + k_2 + \cdots + k_p = n$ and since $\{\mathbf{v}_{1,1}, ..., \mathbf{v}_{1,k_1}, \mathbf{v}_{2,1}, ..., \mathbf{v}_{2,k_2}, ... \; ..., \mathbf{v}_{p,1}, ..., \mathbf{v}_{p,k_p}\}$ is linearly independent, $\mathbf{A}$ has n linearly independent eigenvectors, and is thus diagonalizable.

Equivalent Statements for Diagonalizability

Theorem

Let $\mathbf{A}$ *be a square matrix of order* n. *The following statements are equivalent.*

(i) $\mathbf{A}$ *is diagonalizable.*

(ii) *There exists a basis* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *of* $\mathbb{R}^n$ *of eigenvectors of* $\mathbf{A}$.

(iii) *The characteristic polynomial of* $\mathbf{A}$ *splits into linear factors,*

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

where r_{λ_i} *is the algebraic multiplicity of* λ_i, *for* $i = 1, ..., k$, *and the eigenvalues are distinct,* $\lambda_i \neq \lambda_j$ *for all* $i \neq j$, *and the geometric multiplicity is equal to the algebraic multiplicity for each eigenvalue* λ_i,

$$\dim(E_{\lambda_i}) = r_{\lambda_i}.$$

Example

Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix}$. The characteristic polynomial is

$$x\mathbf{I} - \mathbf{A} = \begin{vmatrix} x-3 & -1 & 1 \\ -1 & x-3 & 1 \\ 0 & 0 & x-2 \end{vmatrix} = (x-2)[(x-3)^2 - 1] = (x-2)(x^2 - 6x + 8) = (x-2)(x-2)(x-4).$$

Find a basis for the eigenspaces.

$$2\mathbf{I} - \mathbf{A} = \begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \text{ So, } \left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\} \text{ is a basis for } E_2, \text{ and } \dim(E_2) = 2 = r_2.$$

$$4\mathbf{I} - \mathbf{A} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}. \text{ So, } \left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\} \text{ is a basis for } E_4, \text{ and } \dim(E_4) = 2 = r_4.$$

$\mathbf{A}$ is diagonalizable with

$$\begin{pmatrix} 3 & 1 & -1 \\ 1 & 3 & -1 \\ 0 & 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1}$$

Not Diagonalizable

A square matrix $\mathbf{A}$ is diagonalizable if

(i) The characteristic polynomial splits into linear factors,

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}},$$

(ii) and the algebraic multiplicity is equal to the geometric multiplicity,

$$r_\lambda = \dim(E_\lambda),$$

for every eigenvalue λ of $\mathbf{A}$.

To show that a square matrix $\mathbf{A}$ of order n is not diagonalizable, show that either

(i) $\det(x\mathbf{I} - \mathbf{A})$ does not split into linear factors, or

(ii) there exists an eigenvalue λ such that $\dim(E_\lambda) < r_\lambda$.

Example

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

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-1 & 0 & 0 \\ 0 & x & -1 \\ 0 & 1 & x \end{vmatrix} = (x-1)(x^2+1).$$

The characteristic polynomial do not split into linear factors, it has only 1 real root. Hence, $\mathbf{A}$ is not diagonalizable.

Example

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

Since $\mathbf{A}$ is a triangular matrix, 1 is the only eigenvalue of $\mathbf{A}$ with algebraic multiplicity $r_1 = 2$. Compute the eigenspace.

$$\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0 & -2 \\ 0 & 0 \end{pmatrix} \quad \Rightarrow \quad E_1 = \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\} \quad \Rightarrow \quad \dim(E_1) = 1 < r_1 = 2.$$

Hence, $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}$ is not diagonalizable.

Exercise

Suppose $\mathbf{A}$ is a $n \times n$ matrix with $n > 1$. Show that if $\mathbf{A}$ has only 1 eigenvalue λ, then $\mathbf{A}$ is diagonalizable if and only if $\mathbf{A}$ is the scalar matrix, $\mathbf{A} = \lambda \mathbf{I}_n$.

Hence, all non-scalar matrix with only 1 eigenvalue is not diagonalizable.

Example

$$\mathbf{A} = \begin{pmatrix} 2 & -2 & 2 \\ 1 & -3 & 5 \\ 1 & -3 & 5 \end{pmatrix}$$

$$\det(\lambda \mathbf{I} - \mathbf{A}) = \begin{vmatrix} x-2 & 2 & -2 \\ -1 & x+3 & -5 \\ -1 & 3 & x-5 \end{vmatrix} = x(x-2)^2.$$

Eigenvalues are $\lambda = 0, 2$ with multiplicities $r_0 = 1$, $r_2 = 2$. Compute eigenspace associated to eigenvalue 2.

$$2\mathbf{I} - \mathbf{A} = \begin{pmatrix} 0 & 2 & -2 \\ -1 & 5 & -5 \\ -1 & 3 & -3 \end{pmatrix} \xrightarrow{rref} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \quad \Rightarrow \quad \dim(E_2) = 1 < r_2 = 2.$$

Hence, $\mathbf{A} = \begin{pmatrix} 2 & -2 & 2 \\ 1 & -3 & 5 \\ 1 & -3 & 5 \end{pmatrix}$ is not diagonalizable.

Distinct Eigenvalues

Theorem

*If **A** is a square matrix of order n with n distinct eigenvalues, then **A** is diagonalizable.*

Sketch of Proof.

Follows from

$$
\begin{array}{ccccccccccc}
n & = & 1 & + & 1 & + & \cdots & + & 1 & & \\
 & & \rotatebox[origin=c]{-90}{$\leq$} & & \rotatebox[origin=c]{-90}{$\leq$} & & \cdots & & \rotatebox[origin=c]{-90}{$\leq$} & & \\
 & & \dim(E_{\lambda_1}) & + & \dim(E_{\lambda_2}) & + & \cdots & + & \dim(E_{\lambda_n}) & & \\
 & & \rotatebox[origin=c]{-90}{$\leq$} & & \rotatebox[origin=c]{-90}{$\leq$} & & \cdots & & \rotatebox[origin=c]{-90}{$\leq$} & & \\
 & & r_1 & + & r_2 & + & \cdots & + & r_n & \leq & n
\end{array}
$$

$\square$

Example

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix}. \text{ Eigenvalues are: } \lambda = 1, 2, 3.$$

$$\lambda = 1: \begin{pmatrix} 0 & -1 & -1 \\ 0 & -1 & -2 \\ 0 & 0 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow E_1 = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$$

$$\lambda = 2: \begin{pmatrix} 1 & -1 & -1 \\ 0 & 0 & -2 \\ 0 & 0 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \right\}$$

$$\lambda = 3: \begin{pmatrix} 2 & -1 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 2 & 0 & -3 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow E_3 = \text{span}\left\{ \begin{pmatrix} 3 \\ 4 \\ 2 \end{pmatrix} \right\}$$

$$\Rightarrow \mathbf{A} = \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix} \begin{pmatrix} 1 & 1 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{pmatrix}^{-1}$$

Algorithm to Diagonalization

Let $\mathbf{A}$ be an order n square matrix.

1. Compute the characteristic polynomial

$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}}.$$

   If the characteristic polynomial do not split into linear factors, $\mathbf{A}$ is not diagonalizable.

2. For each eigenvalue λ_i of $\mathbf{A}$, $i = 1, ..., k$, find a basis S_{λ_i} for the eigenspace, that is, find a basis S_{λ_i} for the solution space of the following linear system,

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

   Compute first the eigenspace associated to eigenvalues with algebraic multiplicity greater than 1. If $\dim(E_\lambda) < r_\lambda$, $\mathbf{A}$ is not diagonalizable.

3. Let $S = \bigcup_{i=1}^{k} S_{\lambda_i}$. Then $S = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is a basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$.
4. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, and $\mathbf{D} = \mathrm{diag}(\mu_1, \mu_2, ..., \mu_n)$, where μ_i is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i\mathbf{u}_i$. Then

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

Example

Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$.

The characteristic polynomial is $\begin{vmatrix} x-1 & -1 & 0 \\ -1 & x-1 & 0 \\ 0 & 0 & x-2 \end{vmatrix} = x(x-2)^2$. So, the eigenvalues are $\lambda = 0, 2$ with algebraic multiplicities , $r_0 = 1, r_2 = 2$, respectively. Now find a basis for the eigenspaces.

- $2\mathbf{I} - \mathbf{A} = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. So, $E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$
- $0\mathbf{I} - \mathbf{A} = \begin{pmatrix} -1 & -1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. So, $E_0 = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} \right\}$

Hence, $\mathbf{A}$ is diagonalizable, with

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

Question

Suppose $\mathbf{A}$ is diagonalizable. Which of the following statement(s) is/are true?

(i) If the diagonal matrix $\mathbf{D}$ is fixed, then the invertible matrix $\mathbf{P}$ is fixed.

(ii) If the invertible matrix $\mathbf{P}$ is fixed, then the diagonal matrix $\mathbf{D}$ is fixed.

6.3 Orthogonally Diagonalizable

Question

Fill in the blank.

Suppose $\mathbf{P}$ is a square matrix such that $\mathbf{P}^T = \mathbf{P}^{-1}$. Then $\mathbf{P}$ is a ____________________ matrix.

Orthogonally Diagonalizable

Definition

An order n square matrix $\mathbf{A}$ is orthogonally diagonalizable if

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T$$

for some orthogonal matrix $\mathbf{P}$ and diagonal matrix $\mathbf{D}$.

Remark

Note that since $\mathbf{P}$ is orthogonal, $\mathbf{P}^T = \mathbf{P}^{-1}$, then $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. That is, orthogonally diagonalizable matrices are also diagonalizable, except we need $\mathbf{P}$ to not only be invertible, but also an orthogonal matrix.

The Spectral Theorem

Theorem (Spectral theorem)

Let $\mathbf{A}$ be a $n \times n$ square matrix. $\mathbf{A}$ is orthgonally diagonalizable if and only if $\mathbf{A}$ is symmetric.

Suppose $\mathbf{A}$ is orthogonally diagonalizable. Write $\mathbf{A} = \mathbf{PDP}^T$. Then

$$\mathbf{A}^T = (\mathbf{PDP}^T)^T = (\mathbf{P}^T)^T\mathbf{D}^T\mathbf{P}^T = \mathbf{PDP}^T = \mathbf{A},$$

where we used the fact that diagonal matrices are symmetric.

The proof of the converse if beyond the scope of the syllabus.

Equivalent Statements for Orthogonally Diagonalizable

Theorem

Let $\mathbf{A}$ *be a square matrix of order* n. *The following statements are equivalent.*

(i) $\mathbf{A}$ *is orthogonally diagonalizable.*

(ii) *There exists an orthonormal basis* $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ *of* $\mathbb{R}^n$ *of eigenvectors of* $\mathbf{A}$.

(iii) $\mathbf{A}$ *is a symmetric matrix.*

Example

- $$\begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{pmatrix}$$

- $$\begin{pmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{pmatrix}$$

Orthogonally Diagonalization

$\mathbf{A}$ orthogonally diagonalizable if and only if

$$\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}, \text{ and } \mathbf{D} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix},$$

where μ_i is the eigenvalue associated to eigenvector $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$.

Now $\mathbf{P}$ is orthogonal if and only if its columns $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.

However, in the algorithm to diagonalization, there is no guarantee that the basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ obtained is orthonormal. Does this mean that we need to apply the Gram-Schmidt process to all n vectors in the basis of $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}$?

Example

$\mathbf{A} = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$. Characteristic polynomial: $\begin{vmatrix} x-5 & 1 & 1 \\ 1 & x-5 & 1 \\ 1 & 1 & x-5 \end{vmatrix} = (x-3)(x-6)^2$. Eigenvalues: $\lambda = 3, 6$.

$$3\mathbf{I} - \mathbf{A} = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \text{ So, } E_3 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$$

$$6\mathbf{I} - \mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ So, } E_6 = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Check that the eigenspaces are orthogonal.

$$\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = 0, \quad \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} = 0 \quad \Rightarrow \quad E_3 \perp E_6.$$

Visualization of Eigenspace

Visit the following link, https://www.geogebra.org/m/uqksb8h5, to visualize the eigenspaces.

- Here $\mathbf{A} = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$, from the previous example.
- The eigenvalues are $\lambda = 3, 6$.
- The blue plane is E_6, the eigenspace associated to eigenvalue $\lambda = 6$ and the red line is E_3, the eigenspace associated to eigenvalue $\lambda = 3$.
- When $c_1 = 0$, the purple vector $\mathbf{w}$ is a vector in E_6, and the green vector is $\mathbf{A}\mathbf{w} = 6\mathbf{w}$.
- When $c_2 = c_3 = 0$, the purple vector $\mathbf{w}$ is a vector in E_3, and the green vector is $\mathbf{A}\mathbf{w} = 3\mathbf{w}$.
- Observe that E_3 is orthogonal to E_6. However, $\mathbf{u}_2$ is not orthogonal to $\mathbf{u}_3$. The set $\{\mathbf{u}_2, \mathbf{v}_3\}$ is orthogonal. $\mathbf{v}_3$ was obtained via Gram-Schmidt process, see later.

Eigenspaces of a Symmetric Matrix are Orthogonal

Theorem (Eigenspaces of a symmetric matrix is orthogonal)

If $\mathbf{A}$ is a symmetric matrix, then the eigenspaces are orthogonal to each other. That is, suppose λ_1 and λ_2 are distinct eigenvalues of a symmetric matrix $\mathbf{A}$, $\lambda_1 \neq \lambda_2$, and $\mathbf{v}_i$ is an eigenvector associated to eigenvalue λ_i, for $i = 1, 2$. Then $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

Proof.

Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvalues associated to eigenvalues λ_1 and λ_2 of the symmetric matrix $\mathbf{A}$, respectively. Then

$$\lambda_1 \mathbf{v}_1 \cdot \mathbf{v}_2 = (\mathbf{A}\mathbf{v}_1) \cdot \mathbf{v}_2 = (\mathbf{A}\mathbf{v}_1)^T \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{A}^T \mathbf{v}_2 = \mathbf{v}_1^T \mathbf{A}\mathbf{v}_2 = \mathbf{v}_1^T (\lambda_2 \mathbf{v}_2) = \lambda_2 \mathbf{v}_1 \cdot \mathbf{v}_2$$

$\Rightarrow (\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0 \Rightarrow \mathbf{v}_1 \cdot \mathbf{v}_2 = 0$ since $\lambda_1 \neq \lambda_2$. $\square$

This means that, vectors belonging to different eigenspaces are orthogonal to each other.

Hence, we only need to perform Gram-Schmidt process to the eigenvectors within the same eigenspace.

Algorithm to orthogonal diagonalization

Let $\mathbf{A}$ be an order n symmetric matrix. Since $\mathbf{A}$ is symmetric, it is orthogonally diagonalizable.

1. Compute the characteristic polynomial
$$\det(x\mathbf{I} - \mathbf{A}) = (x - \lambda_1)^{r_{\lambda_1}}(x - \lambda_2)^{r_{\lambda_2}} \cdots (x - \lambda_k)^{r_{\lambda_k}}.$$
2. For each eigenvalue λ_i of $\mathbf{A}$, $i = 1, ..., k$, find a basis S_{λ_i} for the eigenspace, that is, find a basis S_{λ_i} for the solution space of the following linear system,
$$(\lambda_i \mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{0}.$$
3. Apply Gram-Schmidt process to each basis S_{λ_i} of the eigenspace E_{λ_i} to obtain an orthonormal basis T_{λ_i}. Let $T = \bigcup_{i=1}^{k} T_{\lambda_i}$. Then $T = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is an orthonormal basis for $\mathbb{R}^n$.
4. Let $\mathbf{P} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, and $\mathbf{D} = \text{diag}(\mu_1, \mu_2, ..., \mu_n)$, where μ_i is the eigenvalue associated to $\mathbf{u}_i$, $i = 1, ..., n$, $\mathbf{A}\mathbf{u}_i = \mu_i \mathbf{u}_i$. Then $\mathbf{P}$ is an orthogonal matrix, and
$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^T.$$

Example

Let $\mathbf{A} = \begin{pmatrix} 5 & -1 & -1 \\ -1 & 5 & -1 \\ -1 & -1 & 5 \end{pmatrix}$. We have found a basis for the eigenspaces.

$$E_3 = \text{span}\left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}, \quad E_6 = \text{span}\left\{ \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$$

Perform Gram-Schmidt process to the vectors in the basis of E_6.

$$\mathbf{v}_1 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} - \frac{1}{2}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}$$

After normalizing the vectors, put them as columns of the matrix $\mathbf{P}$.

$$\mathbf{A} = \begin{pmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \end{pmatrix}.$$

Example

$$\mathbf{A} = \begin{pmatrix} 3 & 0 & -1 \\ 0 & 2 & 0 \\ -1 & 0 & 3 \end{pmatrix}. \text{ Characteristic polynomial: } \begin{vmatrix} x-3 & 0 & 1 \\ 0 & x-2 & 0 \\ 1 & 0 & x-3 \end{vmatrix} = (x-2)^2(x-4). \text{ Eigenvalues: } \lambda = 2, 4.$$

Find a basis for the eigenspaces.

$$2\mathbf{I} - \mathbf{A} = \begin{pmatrix} -1 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & -1 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow E_2 = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$$

$$4\mathbf{I} - \mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow E_4 = \text{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}$$

In this case, there is no need to perform Gram-Schmidt process, just normalize the vectors.

$$\mathbf{A} = \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \end{pmatrix}$$

6.4 Application of Diagonalization: Markov Chain

Powers of Diagonalizable Matrices

Theorem

Suppose $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}$. *Then* $\mathbf{A}^m = \mathbf{P}\mathbf{D}^m\mathbf{P}^{-1}$.

This can be proved by induction; the proof is left as an exercise. Note that we do not require **D** to be diagonal in the theorem.

Theorem

Let $\mathbf{D} = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{pmatrix}$ *be a diagonal matrix. Then for any positive integer* m, $\mathbf{D}^m = \begin{pmatrix} d_1^m & 0 & \cdots & 0 \\ 0 & d_2^m & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & d_n^m \end{pmatrix}$.

That is, the positive powers of a diagonal matrix is a diagonal matrix with entries the powers of the diagonal entries.

Powers of Diagonalizable Matrices

Corollary

Suppose $\mathbf{A}$ *is diagonalizable. Write* $\mathbf{A} = \mathbf{P} \begin{pmatrix} \mu_1 & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n \end{pmatrix} \mathbf{P}^{-1}$. *Then for any positive integer* $k > 0$,

$$\mathbf{A}^k = \mathbf{P} \begin{pmatrix} \mu_1^k & 0 & \cdots & 0 \\ 0 & \mu_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_n^k \end{pmatrix} \mathbf{P}^{-1}.$$

Moreover, if $\mathbf{A}$ *is invertible, then the identity above holds for any integer* $k \in \mathbb{Z}$.

Markov Chain

Definition

(i) A vector $\mathbf{v} = (v_i)_n$ with nonnegative coordinates that add up to 1, $\sum_{i=1}^{n} v_i = 1$, is called a *probability vector*.

(ii) A *stochastic* matrix is a square matrix whose columns are probability vectors.

(iii) A *Markov chain* is a sequence of probability vectors $\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_k, ...$, together with a stochastic matrix $\mathbf{P}$ such that

$$\mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \mathbf{x}_2 = \mathbf{P}\mathbf{x}_1, \quad ... \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}, ...$$

When a Markov chain of vectors in $\mathbb{R}^n$ describes a system or a sequence of experiments, the entries in $\mathbf{x}_k$ list, respectively, the probabilities that the system is in each of n possible states, or the probabilities that the outcome of the experiment is one of n possible outcomes. For this reason, $\mathbf{x}_k$ is often called a state vector.

Observe that $\mathbf{x}_k = \mathbf{P}^k\mathbf{x}_0$.

Example

Sheldon only patronizes three stalls in the school canteen, the mixed rice, noodle, and mala hotpot stall for lunch everyday. He never buys from same stall two days in a row. If he buys from the mixed rice stall on a certain day, there is a 40% chance he will patronize the noodles stall the next day. If he buys from the noodle stall on a certain day, there is a 50% chance he will eat mala hotpot the next day. If he eats mala hotpot on a certain day, there is a 60% chance he will patronize the mixed rice the next day.

Let a_n, b_n, c_n be the probability that Sheldon patronizes the mixed rice, noodles, and mala hotpot stall for lunch after n days. Let $\mathbf{x}_n = \begin{pmatrix} a_n \\ b_n \\ c_n \end{pmatrix}$, then $\mathbf{x}_n$ is a state vector. Let $\mathbf{P}$ be the stochastic matrix. Suppose Sheldon patronizes the mixed rice, noodles, mala hotpot stalls today, his state vector is $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$, respectively, where $e_i \in \mathbb{R}^3$ is the i-th vector in the standard basis. Then by the given hypothesis,

$$\begin{pmatrix} 0 \\ 0.4 \\ 0.6 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \mathbf{Pe}_1, \quad \begin{pmatrix} 0.5 \\ 0 \\ 0.5 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \mathbf{Pe}_2, \quad \begin{pmatrix} 0.6 \\ 0.4 \\ 0 \end{pmatrix} = \mathbf{P}\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \mathbf{Pe}_3 \quad \Rightarrow \mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0.6 \\ 0.4 & 0 & 0.4 \\ 0.6 & 0.5 & 0 \end{pmatrix}$$

Example

By construction, $\mathbf{P} = \begin{pmatrix} 0 & 0.5 & 0.6 \\ 0.4 & 0 & 0.4 \\ 0.6 & 0.5 & 0 \end{pmatrix}$ is a stochastic matrix. The state vector after n days will be $\mathbf{x}_n = \mathbf{P}^n \mathbf{x}_0$.

To compute the powers of $\mathbf{P}$, we may diagonalize $\mathbf{P}$. Performing the algorithm to diagonalization, we obtain

$$\mathbf{P} = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.6 & 0 \\ 0 & 0 & -0.4 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1}.$$

Suppose Sheldon had noodles today. The probability that he patronizes each of the stalls 3 days later is

$$\mathbf{x}_3 = \mathbf{P}^3 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & (-0.6)^3 & 0 \\ 0 & 0 & (-0.4)^3 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.38 \\ 0.24 \\ 0.38 \end{pmatrix}.$$

Example

Recall that if $-1 < r < 1$, then $r^k \to 0$ as $k \to \infty$; that is for very big k, r^k is approximately 0. Hence, in the long run, Sheldon's state vector is

$$\mathbf{P}^k\mathbf{x}_0 = \mathbf{P}^k \begin{pmatrix} a \\ b \\ c \end{pmatrix} \xrightarrow[k\to\infty]{} \mathbf{x}_\infty = \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \\ 0.8 & 0 & -2 \\ 1 & 1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

$$= \frac{1}{14} \begin{pmatrix} 5 & 5 & 5 \\ 4 & 4 & 4 \\ 5 & 5 & 5 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \frac{1}{14} \begin{pmatrix} 5(a+b+c) \\ 4(a+b+c) \\ 5(a+b+c) \end{pmatrix} = \frac{1}{14} \begin{pmatrix} 5 \\ 4 \\ 5 \end{pmatrix},$$

where the last equality follows from the fact that $\mathbf{x}_0$ is a probability vector. That is, he will most probability patronize the mixed rice or mala hotpot stall with equal probability $\frac{5}{14}$ in the long run.

Observe that $\frac{1}{14}\begin{pmatrix} 5 \\ 4 \\ 5 \end{pmatrix}$ is an probability state vector and an eigenvector associated to eigenvalue 1. Meaning, regardless of the choice to the starting state vector $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$, in the long run, the resultant state vector is the same!

Challenge

Definition

A *steady-state vector*, or *equilibrium vector* for a stochastic matrix $\mathbf{P}$ is a probability vector that is an eigenvector associated to eigenvalue 1.

Theorem

Let $\mathbf{P}$ be a $n \times n$ stochastic matrix and

$$\mathbf{x}_0, \quad \mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, \quad \dots \quad , \quad \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k-1}$$

be a Markov chain for some probability vector $\mathbf{x}_0$. If the Markov chain converges, it will converge to an equilibrium vector.

Proof.

Exercise. Hint:

(i) Show that 1 is always an eigenvalue of a stochastic matrix.

(ii) Show that if $\mathbf{v}$ is a probability vector and $\mathbf{P}$ a stochastic matrix, then $\mathbf{P}\mathbf{v}$ is also a probability vector.

(iii) Show that if the Markov chain do converge, then the state vectors will converge to an equilibrium vector.

□

Google PageRank Algorithm

- Assume a set S of sites contain key words on a topic of common interest.
- Need an algorithm to rank the sites, so that the sites with the highest rank appear first.
- In 1996, a new search engine "Google" was developed by Larry Page and Sergey Brin. This engine is based on the PageRank algorithm, which involves the use of a dominant eigenvector of some matrix.
- The underlying assumption is that more important websites are likely to receive more links from other websites.

Adjacency Matrix and Probability Transition Matrix

Suppose the set S contains n sites. We define the adjacency matrix for S for be the order n square matrix $\mathbf{A} = (a_{ij})$ where

$$a_{ij} = \begin{cases} 1 & \text{if site } j \text{ has an outgoing link to site } i; \\ 0 & \text{if site } j \text{ does not have an outgoing link to site } i. \end{cases}$$

$$\mathbf{A} = \begin{array}{c} \\ \\ \\ \\ \end{array} \begin{array}{cc} \begin{array}{cccc} s_1 & s_2 & s_3 & s_4 \end{array} & \\ \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} & \begin{array}{c} s_1 \\ s_2 \\ s_3 \\ s_4 \end{array} \end{array}$$

- s_1 references s_2, s_3 and s_4,
- s_2 references s_4 only,
- s_3 references s_1 and s_4,
- s_4 references s_1 and s_3.

Observe that

- The sum of the entries in the i-th row of is the number of incoming links to site i from other sites.
- The sum of the entries of the j-th column is the number of outgoing links on site j-th to other sites.

Adjacency Matrix and Probability Transition Matrix

From the adjacency matrix $\mathbf{A}$, we define the probability transition matrix $\mathbf{P} = (p_{ij})$ by dividing each entry of $\mathbf{A}$ by the sum of the entries in the same column; that is

$$p_{ij} = \frac{a_{ij}}{\sum_{k=1}^{n} a_{kj}}.$$

Using $\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$, we have

$$\mathbf{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix}.$$

Observe that this is a stochastic matrix.

Example

This matrix incorporates the probability information for advancing randomly from one site to the next with a mouse click. For example, suppose $\mathbf{x}_0 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, that is, suppose the surfer begins surfing from site 2. Then the state vector after 2 subsequent mouse click will be

$$\mathbf{x}_2 = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix}^2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{pmatrix}.$$

This shows that indeed, after 2 random clicks the surfer would have probability $\frac{1}{2}$ of landing on s_1 or s_3.

Discussion

- We may diagonalize the probability transition matrix $\mathbf{P}$ to obtain the outcome in the long run for the different starting state vectors.
- However, since $\mathbf{P}$ is a stochastic matrix, and the starting state vectors are probability vectors, if the Markov chain converges, it will converge to an equilibrium vector.
- Moreover, if the probability transition matrix $\mathbf{P}$ in the Google PageRank algorithm is a regular stochastic matrix (see below for definition), it will always converge to a unique equilibrium vector.

Definition

A stochastic matrix is *regular* if for some positive integer $k > 0$, the matrix power $\mathbf{P}^k$ has positive entries,

$$\mathbf{P}^k = (a_{ij})_n, \quad a_{ij} > 0 \text{ for all } i, j = 1, ..., n.$$

Algorithm to Compute Equilibrium vector

Let $\mathbf{P}$ be a $n \times n$ stochastic matrix.

1. Find an eigenvector $\mathbf{u}$ associate to eigenvalue $\lambda = 1$, that is, find a nontrivial solution to the homogeneous system $(\mathbf{I} - \mathbf{P})\mathbf{x} = \mathbf{0}$.

2. Write $\mathbf{u} = (u_i)$. Then
$$\mathbf{v} = \frac{1}{\sum_{k=1}^{n} u_k}\mathbf{u}$$
will be an equilibrium vector. Indeed, the i-th coordinate of $\mathbf{v}$ is $\frac{u_i}{\sum_{k=1}^{n} u_k}$ and hence, the sum of the coordinates of $\mathbf{v}$ is
$$\sum_{i=1}^{n} \frac{u_i}{\sum_{k=1}^{n} u_k} = \frac{\sum_{i=1}^{n} u_i}{\sum_{k=1}^{n} u_k} = 1.$$

Example

Find an eigenvector $\mathbf{P} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/3 & 0 & 0 & 0 \\ 1/3 & 0 & 0 & 1/2 \\ 1/3 & 1 & 1/2 & 0 \end{pmatrix}$ associated to eigenvalue 1.

$$\mathbf{I} - \mathbf{P} = \begin{pmatrix} 1 & 0 & -1/2 & -1/2 \\ -1/3 & 1 & 0 & 0 \\ -1/3 & 0 & 1 & -1/2 \\ -1/3 & -1 & -1/2 & 1 \end{pmatrix} \xrightarrow{RREF} \begin{pmatrix} 1 & 0 & 0 & -9/10 \\ 0 & 1 & 0 & -3/10 \\ 0 & 0 & 1 & -4/5 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Then $\mathbf{u} = \begin{pmatrix} 9/10 \\ 3/10 \\ 4/5 \\ 1 \end{pmatrix}$ is an eigenvector associate to eigenvalue 1. Hence, the equilibrium vector $\mathbf{v}$ of $\mathbf{P}$ is

$$\mathbf{v} = \left(\frac{9}{10} + \frac{3}{10} + \frac{4}{5} + 1\right)^{-1} \begin{pmatrix} 9/10 \\ 3/10 \\ 4/5 \\ 1 \end{pmatrix} = \begin{pmatrix} 3/10 \\ 1/10 \\ 4/15 \\ 1/3 \end{pmatrix}.$$

Example

This is the probability of visiting the various sites of a random surfer, starting at any random site. This can also be interpreted as the proportion all the surfers for each sites. Therefore, one may ranked according to their probability, that is, ranking the sites as follows

1. site s_4
2. site s_1
3. site s_3
4. site s_2

Exercise

Let $\mathbf{P}$ be a $n \times n$ stochastic matrix. Show that $\mathbf{v}$ is an equilibrium vector of $\mathbf{P}$ if and only if it is a solution to the system

$$\begin{pmatrix} & \mathbf{P} - \mathbf{I}_n & & \\ 1 & 1 & \cdots & 1 \end{pmatrix} \mathbf{x} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix. Here $\begin{pmatrix} & \mathbf{P} - \mathbf{I}_n & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}$ is the $(n+1) \times n$ matrix whose first n rows are the matrix $\mathbf{P} - \mathbf{I}_n$, and the last row has all entries 1.

6.5 Application of Orthogonal Diagonalization: Singular Value Decomposition

Introduction

All non-square matrices are non-diagonalizable, much less orthogonally diagonalizable. However, we can still factorize any $m \times n$ $\mathbf{A}$ into

$$\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T,$$

where $\mathbf{U}$ is an order m orthogonal matrix, $\mathbf{V}$ an order n orthogonal matrix, and the matrix $\boldsymbol{\Sigma}$ has the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix},$$

for some diagonal matrix $\mathbf{D}$ of order r, where $r \leq \min\{m, n\}$.

Example

1. $$\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix} \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{pmatrix}.$$

2. $$\begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} 1/3 & 2/\sqrt{5} & -2/\sqrt{45} \\ -2/3 & 1/\sqrt{5} & 4/\sqrt{5} \\ 2/3 & 0 & 5/\sqrt{45} \end{pmatrix} \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Singular Values

Let $\mathbf{A}$ be a $m \times n$ matrix. Then since $\mathbf{A}^T\mathbf{A}$ is an order n symmetric matrix, we may orthogonally diagonalize it. Let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}^T\mathbf{A}$.

Let μ_i be the eigenvalue associated to $\mathbf{v}_i$, for $i = 1, ..., n$, not necessarily distinct.

Lemma

The eigenvalue μ_i of $\mathbf{A}^T\mathbf{A}$ is nonnegative.

Proof.

$$\|\mathbf{A}\mathbf{v}_i\|^2 = (\mathbf{A}\mathbf{v}_i)^T(\mathbf{A}\mathbf{v}_i) = \mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v}_i = \mu_i\mathbf{v}_i^T\mathbf{v}_i = \mu_i,$$

where third equality follows from the fact that μ_i is an eigenvalue of $\mathbf{A}^T\mathbf{A}$, and the forth equality follows from the fact that $\mathbf{v}_i$ is a unit vector. $\square$

Singular Values

Reordering if necessary, we may assume that

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n \geq 0.$$

The singular values of $\mathbf{A}$ are

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0,$$

where $\sigma_i = \sqrt{\mu_i}$, i=1,...,n, is the square root of the eigenvalues of $\mathbf{A}^T\mathbf{A}$, arranged in decreasing order. Let r be the largest integer such that $1 \leq r \leq n$ and $\sigma_i > 0$ for all $i \leq r$, that is

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > \sigma_{r+1} = 0 = \cdots = \sigma_m = 0$$

Define the matrix $m \times n$ matrix $\mathbf{\Sigma}$ to be

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}, \text{ where } \mathbf{D} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

Example

Let $\mathbf{A} = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$.

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 4 & 8 \\ 11 & 7 \\ 14 & -2 \end{pmatrix} \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 80 & 100 & 40 \\ 100 & 170 & 140 \\ 40 & 140 & 200 \end{pmatrix}.$$

$$\det(x\mathbf{I} - \mathbf{A}^T\mathbf{A}) = \begin{vmatrix} x-80 & -100 & -40 \\ -100 & x-170 & -140 \\ -40 & -140 & x-200 \end{vmatrix} = x(x-90)(x-360).$$

So, the singular values are

$$\sigma_1 = \sqrt{360} = 6\sqrt{10}, \quad \sigma_2 = \sqrt{90} = 3\sqrt{10}, \quad \sigma_3 = 0,$$

and

$$\mathbf{\Sigma} = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}$$

Exercise

1. Show that

$$\mathbf{\Sigma}^T\mathbf{\Sigma} = \begin{pmatrix} \mu_1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & \mu_2 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mu_r & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & \mu_n \end{pmatrix},$$

where μ_i, $i = 1, ..., n$, is the eigenvalues of $\mathbf{A}^T\mathbf{A}$; that is $\mathbf{A}^T\mathbf{A} = \mathbf{P}\mathbf{\Sigma}^T\mathbf{\Sigma}\mathbf{P}^T$, where $\mathbf{P} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}$ and $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ of eigenvectors of $\mathbf{A}^T\mathbf{A}$.

2. Show that $\mathbf{A}\mathbf{v}_i \neq \mathbf{0}$ for all $i \leq r$ and $\mathbf{A}\mathbf{v}_i = \mathbf{0}$ for all $i > r$.

Singular Value Decomposition

Suppose $\mathbf{A}$ is a $m \times n$ matrix. Let $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ be an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of $\mathbf{A}^T\mathbf{A}$. Let

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$$

be the nonzero singular values of $\mathbf{A}$. Define

$$\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i, \quad i = 1, .., r.$$

Lemma

$\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is an orthonormal basis for the column space of $\mathbf{A}$, and $\operatorname{rank}(\mathbf{A}) = r$.

Proof.

By construction, $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is an orthonormal set,

$$\mathbf{u}_i \cdot \mathbf{u}_j = \frac{1}{\sigma_i\sigma_j}(\mathbf{A}\mathbf{v}_i)^T(\mathbf{A}\mathbf{v}_j) = \frac{1}{\sigma_i\sigma_j}\mathbf{v}_i^T\mathbf{A}^T\mathbf{A}\mathbf{v_j} = \frac{\mu_j}{\sigma_i\sigma_j}\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 0 & \text{if } i \neq j \\ \frac{\mu_i}{\sigma_i\sigma_i} = 1 & \text{if } i = j \end{cases}$$

since $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n\}$ is an orthonormal set.

Singular Value Decomposition

Continue of proof.

Recall that $\mathrm{Col}(\mathbf{A}) = \{\mathbf{Av} \mid \mathbf{v} \in \mathbb{R}^n\}$. Hence, by construction, $\mathbf{u}_i \in \mathrm{Col}(\mathbf{A})$, and hence $\mathrm{span}\{\mathbf{u}_1, ..., \mathbf{u}_r\} \subseteq \mathrm{Col}(\mathbf{A})$. Now given any $\mathbf{v} \in \mathbb{R}^n$, write $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n$. Then

$$\begin{aligned} \mathbf{Av} &= c_1\mathbf{Av}_1 + c_2\mathbf{Av}_2 + \cdots + c_n\mathbf{Av}_n \\ &= \sigma_1 c_1 \left(\frac{1}{\sigma_1}\mathbf{Av}_1\right) + \sigma_2 c_2 \left(\frac{1}{\sigma_2}\mathbf{Av}_2\right) + \cdots + \sigma_r c_r \left(\frac{1}{\sigma_r}\mathbf{Av}_r\right) + \mathbf{0} + \cdots + \mathbf{0} \\ &= \sigma_1 c_1 \mathbf{u}_1 + \sigma_2 c_2 \mathbf{u}_2 + \cdots \sigma_r c_r \mathbf{u}_r \end{aligned}$$

This shows that $\mathrm{Col}(\mathbf{A}) \subseteq \mathrm{span}\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ too, therefore there are equal. Hence, $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ is an orthonormal basis for the column space of $\mathbf{A}$, which proves that $\mathrm{rank}(\mathbf{A}) = r$. $\square$

Singular Value Decomposition

Using the notations from above, extend $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_m\}$ for $\mathbb{R}^m$ (if $r \neq m$). Define

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_m \end{pmatrix},$$

it is an order m orthogonal matrix. Define

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix},$$

then $\mathbf{V}$ is an order n orthogonal matrix. Let $\mathbf{\Sigma}$ be the matrix defined by the nonzero singular values $\sigma_1, \sigma_2, ..., \sigma_r$. Then

$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T.$$

Proof.

Since $\mathbf{V}$ is orthogonal, suffice to show that $\mathbf{AV} = \mathbf{U\Sigma}$, but by construction,

$$\mathbf{AV} = \begin{pmatrix} \mathbf{Av}_1 & \cdots & \mathbf{Av}_r & \mathbf{Av}_{r+1} & \cdots & \mathbf{Av}_n \end{pmatrix} = \begin{pmatrix} \sigma_1\mathbf{u}_1 & \cdots & \sigma_r\mathbf{u}_r & \mathbf{0} & \cdots & 0 \end{pmatrix} = \mathbf{U\Sigma}.$$

$\square$

Algorithm to Singular Value Decomposition

Let $\mathbf{A}$ be a $m \times n$ matrix with $\text{rank}(\mathbf{A}) = r$.

1. Find the eigenvalues of $\mathbf{A}^T\mathbf{A}$. Arrange the nonzero eigenvalues in descending order (counting multiplicity)

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_r > 0 = \mu_{r+1} = \cdots = \mu_n,$$

   and let $\sigma_i = \sqrt{\mu_i}$, $i = 1, ..., r$. Set

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{D} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{pmatrix}, \text{ where } \mathbf{D} = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix}.$$

2. Find an orthogonal basis for each eigenspace, and let $\mathbf{v}_i$ be the unit vector associated to μ_i. Set

$$\mathbf{V} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{pmatrix}.$$

3. Let $\mathbf{u}_i = \frac{1}{\sigma_i}\mathbf{A}\mathbf{v}_i$ for $i = 1, ..., r$. Extend $\{\mathbf{u}_1, ..., \mathbf{u}_r\}$ to an orthonormal basis $\{\mathbf{u}_1, ..., \mathbf{u}_r, \mathbf{u}_{r+1}, ..., \mathbf{u}_m\}$ of $\mathbb{R}^m$, that is, solve for $\begin{pmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_r \end{pmatrix}^T \mathbf{x} = \mathbf{0}$ and find an orthonormal basis for the solution space. Let

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$$

Example

Let $\mathbf{A} = \begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}$.

Eigenvalues are $\mu_1 = 360, \mu_2 = 90, \mu_3 = 0$. Here $\text{rank}(A) = 2$, so,

$$\mathbf{\Sigma} = \begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}.$$

$$\mu_1 = 360: \quad 360\mathbf{I} - \mathbf{A}^T\mathbf{A} = \begin{pmatrix} 280 & -100 & -40 \\ -100 & 190 & -140 \\ -40 & -140 & 160 \end{pmatrix} \xrightarrow{rref} \begin{pmatrix} 1 & 0 & -1/2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \mathbf{v}_1 = \begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix}$$

$$\mu_2 = 90: \quad 90\mathbf{I} - \mathbf{A}^T\mathbf{A} = \begin{pmatrix} 10 & -100 & -40 \\ -100 & -80 & -140 \\ -40 & -140 & -110 \end{pmatrix} \xrightarrow{rref} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1/2 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \mathbf{v}_2 = \begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix}$$

$$\mu_2 = 0: \quad -\mathbf{A}^T\mathbf{A} = \begin{pmatrix} -80 & -100 & -40 \\ -100 & -170 & -140 \\ -40 & -140 & -200 \end{pmatrix} \xrightarrow{rref} \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{pmatrix} \Rightarrow \mathbf{v}_3 = \begin{pmatrix} 2/3 \\ -2/3 \\ 1/3 \end{pmatrix}$$

Example

Set $\mathbf{V} = \begin{pmatrix} 1/3 & -2/3 & 2/3 \\ 2/3 & -1/3 & -2/3 \\ 2/3 & 2/3 & 1/3 \end{pmatrix}$. Finally,

$$\mathbf{u}_1 = \frac{1}{6\sqrt{10}}\mathbf{A}\mathbf{v}_1 = \frac{1}{6\sqrt{10}}\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}\begin{pmatrix} 1/3 \\ 2/3 \\ 2/3 \end{pmatrix} = \frac{1}{6\sqrt{10}}\begin{pmatrix} 18 \\ 6 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{pmatrix},$$

$$\mathbf{u}_2 = \frac{1}{3\sqrt{10}}\mathbf{A}\mathbf{v}_2 = \frac{1}{3\sqrt{10}}\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix}\begin{pmatrix} -2/3 \\ -1/3 \\ 2/3 \end{pmatrix} = \frac{1}{3\sqrt{10}}\begin{pmatrix} 3 \\ -9 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{10} \\ -3/\sqrt{10} \end{pmatrix}.$$

$\{\mathbf{u}_1, \mathbf{u}_2\}$ is already an orthonormal basis for $\mathbb{R}^2$. Set

$$\mathbf{U} = \begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix}.$$

Then

$$\begin{pmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{pmatrix} = \begin{pmatrix} 3/\sqrt{10} & 1/\sqrt{10} \\ 1/\sqrt{10} & -3/\sqrt{10} \end{pmatrix}\begin{pmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{pmatrix}\begin{pmatrix} 1/3 & 2/3 & 2/3 \\ -2/3 & -1/3 & 2/3 \\ 2/3 & -2/3 & 1/3 \end{pmatrix}$$

Example

$\mathbf{A} = \begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix}$. Observe that $\operatorname{rank}(\mathbf{A}) = 1$

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} 9 & -9 \\ -9 & 9 \end{pmatrix}. \quad \det(x\mathbf{I} - \mathbf{A}^T\mathbf{A}) = \begin{vmatrix} x-9 & 9 \\ 9 & x-9 \end{vmatrix} = x(x-18) \Rightarrow \mu_1 = 18, \mu_2 = 0 \Rightarrow \mathbf{\Sigma} = \begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

$$\mu_1 = 18: \quad 18\mathbf{I} - \mathbf{A}^T\mathbf{A} = \begin{pmatrix} 9 & 9 \\ 9 & 9 \end{pmatrix} \quad \Rightarrow \quad \mathbf{v}_1 = \begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix},$$

$$\mu_2 = 0: \quad -\mathbf{A}^T\mathbf{A} = \begin{pmatrix} -9 & 9 \\ 9 & -9 \end{pmatrix} \quad \Rightarrow \quad \mathbf{v}_2 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

$$\Rightarrow \mathbf{V} = \begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}.$$

Example

$$\mathbf{u}_1 = \frac{1}{3\sqrt{2}}\mathbf{A}\mathbf{v}_1 = \frac{1}{3\sqrt{2}}\begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix}\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = \begin{pmatrix} -1/3 \\ 2/3 \\ -2/3 \end{pmatrix}.$$

Extend $\{\mathbf{u}_1\}$ to an orthonormal basis for $\mathbb{R}^3$. Solve for

$$-\frac{1}{3}x + \frac{2}{3}y - \frac{2}{3}z = 0. \text{ General Solution: } \quad s\begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + t\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \quad s, t \in \mathbb{R}.$$

Performing Gram-Schmidt process, we get

$$\mathbf{u}_2 = \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \\ 0 \end{pmatrix}, \quad \mathbf{u}_3 = \begin{pmatrix} -2/(3\sqrt{5}) \\ 4/(3\sqrt{5}) \\ 5/(3\sqrt{5}) \end{pmatrix} \quad \Rightarrow \quad \mathbf{U} = \begin{pmatrix} -1/3 & 2/\sqrt{5} & -2/(3\sqrt{5}) \\ 2/3 & 1/\sqrt{5} & 4/(3\sqrt{5}) \\ -2/3 & 0 & 5/(3\sqrt{5}) \end{pmatrix}$$

So,

$$\begin{pmatrix} 1 & -1 \\ -2 & 2 \\ 2 & -2 \end{pmatrix} = \begin{pmatrix} -1/3 & 2/\sqrt{5} & -2/(3\sqrt{5}) \\ 2/3 & 1/\sqrt{5} & 4/(3\sqrt{5}) \\ -2/3 & 0 & 5/(3\sqrt{5}) \end{pmatrix}\begin{pmatrix} 3\sqrt{2} & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}$$

Challenge

Let $\mathbf{A}$ be a $m \times n$ matrix. Prove the following statements.

1. $\text{rank}(\mathbf{A}) = n$ if and only if all the singular values of $\mathbf{A}$ are positive.

2. $\text{rank}(\mathbf{A}) = m$ if and only if all the singular values of $\mathbf{A}^T$ are positive.

Appendix

Independence of Eigenspaces

Theorem

Let $\mathbf{A}$ *be a* $n \times n$ *matrix. Let* λ_1 *and* λ_2 *are distinct eigenvalues of* $\mathbf{A}$, $\lambda_1 \neq \lambda_2$. *Suppose* $\{\mathbf{u}_1, ..., \mathbf{u}_k\}$ *is a linearly independent subset of eigenspace associated to eigenvalue* λ_1, *and* $\{\mathbf{v}_1, ..., \mathbf{v}_m\}$ *is a linearly independent subset of of eigenspace associated to eigenvalue* λ_2. *Then* $\{\mathbf{u}_1, ..., \mathbf{u}_k, \mathbf{v}_1, ..., \mathbf{v}_m\}$ *is linearly independent.*

Sketch of Proof.

Suppose $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m = \mathbf{0}$. Multiply both sides of the equation by $\mathbf{A}$, λ_1, and λ_2, respectively, we have

$$\begin{aligned} \mathbf{0} &= \mathbf{A}(c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m) = c_1\lambda_1\mathbf{u}_1 + \cdots + c_k\lambda_1\mathbf{u}_k + d_1\lambda_2\mathbf{v}_1 + \cdots + d_m\lambda_2\mathbf{v}_m && (1) \\ \mathbf{0} &= \lambda_1(c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m) = c_1\lambda_1\mathbf{u}_1 + \cdots + c_k\lambda_1\mathbf{u}_k + d_1\lambda_1\mathbf{v}_1 + \cdots + d_m\lambda_1\mathbf{v}_m && (2) \\ \mathbf{0} &= \lambda_2(c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k + d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m) = c_1\lambda_2\mathbf{u}_1 + \cdots + c_k\lambda_2\mathbf{u}_k + d_1\lambda_2\mathbf{v}_1 + \cdots + d_m\lambda_2\mathbf{v}_m && (3) \end{aligned}$$

Take equation(1) - equation(2), we have

$$\mathbf{0} = (\lambda_2 - \lambda_1)(d_1\mathbf{v}_1 + \cdots + d_m\mathbf{v}_m).$$

Since $(\lambda_2 - \lambda_1) \neq 0$, we can conclude that $d_1 = \cdots = d_m = 0$. Take equation(1)-equation(3), we too can conclude that $c_1 = \cdots = c_k = 0$. $\square$

Similar Matrices

Definition

Two square matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be *similar* if there exists an invertible matrix $\mathbf{P}$ such that

$$\mathbf{A} = \mathbf{P}\mathbf{B}\mathbf{P}^{-1}.$$

Example

Diagonalizable matrices are similar to diagonal matrices.

Similar Matrices

Lemma

Suppose $\mathbf{A}$ *and* $\mathbf{B}$ *are similar matrices, then they have the same characteristic polynomial,*

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

Proof.

Let $\mathbf{P}$ be such that $\mathbf{A} = \mathbf{PBP}^{-1}$. Then

$$\begin{aligned} \det(x\mathbf{I} - \mathbf{B}) &= \det(x\mathbf{I} - \mathbf{B}) \det(\mathbf{P}) \det(\mathbf{P})^{-1} \\ &= \det(\mathbf{P}) \det(x\mathbf{I} - \mathbf{B}) \det(\mathbf{P}^{-1}) \\ &= \det(\mathbf{P}(x\mathbf{I} - \mathbf{B})\mathbf{P}^{-1}) \\ &= \det(\mathbf{P}x\mathbf{I}\mathbf{P}^{-1} - \mathbf{PBP}^{-1}) \\ &= \det(x\mathbf{I} - \mathbf{A}) \end{aligned}$$

$\square$

Geometric multiplicity is no greater than algebraic multiplicity

Theorem (Geometric multiplicity is no greater than algebraic multiplicity)

The geometric multiplicity of an eigenvalue λ of a square matrix $\mathbf{A}$ is no greater than the algebraic multiplicity, that is,

$$1 \leq \dim(E_\lambda) \leq r_\lambda.$$

Proof.

Let $\mathbf{A}$ be a $n \times n$ matrix. Let λ be an eigenvalue of $\mathbf{A}$ and E_λ be the associated eigenspace. Suppose $\dim(E_\lambda) = k$. Let $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k\}$ be a basis for the eigenspace E_λ. Extend this set to be a basis $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_k, \mathbf{u}_{k+1}, ..., \mathbf{u}_n\}$ of $\mathbb{R}^n$. Let $\mathbf{Q} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix}$, it is an invertbile matrix. Note that

$$\begin{aligned} \begin{pmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{pmatrix} &= \mathbf{I} = \mathbf{Q}^{-1}\mathbf{Q} = \mathbf{Q}^{-1}\begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{Q}^{-1}\mathbf{u}_1 & \mathbf{Q}^{-1}\mathbf{u}_2 & \cdots & \mathbf{Q}^{-1}\mathbf{u}_n \end{pmatrix}, \end{aligned}$$

that is, $\mathbf{Q}^{-1}\mathbf{u}_i = \mathbf{e}_i$ for all $i = 1, ..., n$. Let $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}$.

Geometric multiplicity is no greater than algebraic multiplicity

Continue of Proof.
Then

$$
\begin{aligned}
\mathbf{B} &= \mathbf{Q}^{-1}\mathbf{A}\begin{pmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{pmatrix} = \mathbf{Q}^{-1}\begin{pmatrix}\mathbf{A}\mathbf{u}_1 & \mathbf{A}\mathbf{u}_2 & \cdots & \mathbf{A}\mathbf{u}_n\end{pmatrix} \\
&= \mathbf{Q}^{-1}\begin{pmatrix}\lambda\mathbf{u}_1 & \cdots & \lambda\mathbf{u}_k & \mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{A}\mathbf{u}_n\end{pmatrix} \\
&= \begin{pmatrix}\lambda\mathbf{Q}^{-1}\mathbf{u}_1 & \cdots & \lambda\mathbf{Q}^{-1}\mathbf{u}_k & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n\end{pmatrix} \\
&= \begin{pmatrix}
\lambda & 0 & \cdots & 0 & & \\
0 & \lambda & \cdots & 0 & & \\
\vdots & & \ddots & \vdots & & \\
0 & 0 & \cdots & \lambda & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_{k+1}\cdots & \mathbf{Q}^{-1}\mathbf{A}\mathbf{u}_n \\
\vdots & \vdots & & \vdots & & \\
0 & 0 & \cdots & 0 & &
\end{pmatrix}.
\end{aligned}
$$

Geometric multiplicity is no greater than algebraic multiplicity

Continue of Proof.

This means that $\det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x)$ for some polynomial $p(x)$. By since $\mathbf{A}$ and $\mathbf{B}$ are similar matrices,

$$\det(x\mathbf{I} - \mathbf{A}) = \det(x\mathbf{I} - \mathbf{B}) = (x - \lambda)^k p(x).$$

This means that the algebraic multiplicity of the eigenvalue λ of $\mathbf{A}$ is no less than k, that is,

$$r_\lambda \geq k = \dim(E_\lambda).$$

□

Regular Stochastic Matrix

Definition

A stochastic matrix is *regular* if for some positive integer $k > 0$, the matrix power $\mathbf{P}^k$ has positive entries,

$$\mathbf{P}^k = (a_{ij})_n, \quad a_{ij} > 0 \text{ for all } i, j = 1, ..., n.$$

Lemma

Let $\mathbf{A} = (a_{ij})_n$ be a $n \times n$ stochastic matrix with positive entries, $a_{ij} > 0$ for all $i, j = 1, ..., n$. Then geometric multiplicity of eigenvalue 1 is 1, $\dim(E_1) = 1$.

Proof.

Write $\mathbf{A} = (a_{ij})_n$. We will show that the geometric multiplicity of $\lambda = 1$ as an eigenvalue of $\mathbf{A}^T$ is 1. Let $\mathbf{x} = (x_i)$ be an eigenvector of $\mathbf{A}^T$ associated to eigenvalue 1. By taking a multiple of $\mathbf{x}$ if necessary, we may assume that $\mathbf{x}$ has some coordinates with positive entries. Let $1 \leq m \leq n$ be the coordinate of $\mathbf{x}$ such that x_m is the largest, $x_m \geq x_i$ for all $i = 1, ..., n$. Now comparing the m-th coordinate of $\mathbf{A}^T\mathbf{x} = \mathbf{x}$, we have

$$a_{1m}x_1 + a_{2m}x_2 + \cdots + a_{nm}x_n = x_m.$$

Regular Stochastic Matrix

Continue of Proof.

Note that necessarily $x_m \neq 0$ and hence, dividing by x_m, we have

$$a_{1m}\frac{x_1}{x_m} + a_{2m}\frac{x_2}{x_m} + \cdots + a_{nm}\frac{x_n}{x_m} = 1.$$

Note that $\frac{x_i}{x_m} \leq 1$ for all $i = 1, ..., n$ and so $a_{im}\frac{x_i}{x_m} \leq a_{im}$. Suppose $x_m > x_j$ for some $j = 1, ..., n$. Then since $a_{jm} > 0$, $a_{jm}\frac{x_j}{x_m} < a_{jm}$, and thus

$$1 = a_{1m}\frac{x_1}{x_m} + a_{2m}\frac{x_2}{x_m} + \cdots + a_{jm}\frac{x_j}{x_m} + \cdots + a_{nm}\frac{x_n}{x_m} < a_{1m} + a_{2m} + \cdots + a_{jm} + \cdots + a_{nm} = 1,$$

a contradiction. Hence, $x_i = x_m$ for all $i = 1, ..., n$, that is $\mathbf{x} = \alpha \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ for some real number α, which shows that the geometric multiplicity of eigenvalue 1 as an eigenvalue of $\mathbf{A}^T$ is 1. The result therefore follows from the fact that the geometric multiplicity of 1 as an eigenvalue of $\mathbf{A}^T$ is equal to its geometric multiplicity as an eigenvalue of $\mathbf{A}$. $\square$

Regular Stochastic Matrix

Lemma

Let $\mathbf{A}$ *be a* $n \times n$ *square matrix and* $\mathbf{v}$ *an eigenvector of* $\mathbf{A}$ *associated to eigenvalue* λ*. Then for any positive integer* k*,* $\mathbf{v}$ *is an eigenvector of* $\mathbf{A}^k$ *associated to eigenvalue* λ^k*.*

The proof is left as an exercise. Together with the previous lemma, this shows that if $\mathbf{P}$ is a regular stochastic matrix, then the geometric multiplicity of eigenvalue 1 is 1.

Lemma

Suppose $\mathbf{P}$ *is a regular stochastic matrix. Then for any probability vector* $\mathbf{x}_0$ *the Markov chain* $\{\mathbf{x}_0, \mathbf{x}_1 = \mathbf{P}\mathbf{x}_0, ..., \mathbf{x}_k = \mathbf{P}\mathbf{x}_{k+1}\}$ *will converge.*

The proof of the lemma regulars knowledge of Jordan block form, which is beyond the syllabus of the course.

Regular Stochastic Matrix

Theorem

If $\mathbf{P}$ *is an* $n \times n$ *regular stochastic matrix, then* $\mathbf{P}$ *has a unique equilibrium vector. Moreover, if* $\mathbf{x}_0$ *is any probability vector and* $\mathbf{x}_{k+1} = \mathbf{P}\mathbf{x}_k$ *for* $k = 0, 1, \ldots$*, then the Markov chain* $\{\mathbf{x}_k\}$ *converges to the unique equilibrium vector.*

Proof.

Since $\mathbf{P}$ is a regular stochastic matrix, the geometric multiplicity of the eigenvalue 1 is 1, and thus the equilibrium vector is unique. Also, since the Markov chain will converge, it will converge to the unique equilibrium vector. $\square$