

Chapter 7

Special matrices, Markov chains & PageRank

Contents (class version)

7.0 Introduction	7.3
7.1 Companion matrices	7.3
Vandermonde matrices and diagonalizing a companion matrix	7.11
Using companion matrices to check for common roots of two polynomials	7.14
7.2 Circulant matrices	7.16
7.3 Toeplitz matrices	7.23
7.4 Power iteration	7.25
Geršgorin disk theorem	7.30
7.5 Nonnegative matrices and graphs	7.34
Weighted directed graphs	7.38
Strongly connected graphs	7.42
Irreducible matrix	7.43
Matrix period	7.44
7.6 Nonnegative matrices and Perron-Frobenius theorems	7.50
PF for square nonnegative matrices	7.50

PF for nonnegative irreducible matrices	7.53
PF for primitive matrices	7.54
PF for stochastic matrices	7.55
7.7 Markov chains	7.56
Equilibrium distribution(s) of a Markov chain	7.60
Limiting distribution(s) of a Markov chain	7.63
Markov chains with strongly connected graphs	7.66
Google's PageRank method	7.70
7.8 Summary	7.74

7.0 Introduction

This chapter contains topics related to matrices with special structures that arise in many applications, including in the analysis of **Markov chains**. The chapter ends with Google's **PageRank** method as an application.

The sections on nonnegative matrices benefited greatly from Caroline Crockett's feedback.

7.1 Companion matrices

L§10.4

Previously we defined the **eigenvalues** of a square matrix A to be the roots of the **characteristic polynomial**:

$$\det\{A - zI\} = 0.$$

Here we work somewhat in the reverse direction by starting with a polynomial and then defining a matrix from it. For $n \in \mathbb{N}$, consider the **monic polynomial**

$$p(z) = z^n + c_{n-1}z^{n-1} + \cdots + c_1z + c_0, \quad z \in \mathbb{C},$$

and now define the following $n \times n$ matrix called a **companion matrix** of that polynomial:

$$A \triangleq \begin{bmatrix} -c_{n-1} & -c_{n-2} & \dots & -c_1 & -c_0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

1. Which matrix class is this?

- A: Lower triangular B: Upper triangular C: Lower Hessenberg D: Upper Hessenberg E: None of these

??

Sparse

Example. For $n = 3$ we have

$$\mathbf{A} = \begin{bmatrix} -c_2 & -c_1 & -c_0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

To examine the eigenvalues of \mathbf{A} , determine the **characteristic polynomial** by evaluating the **determinant** using the usual minors:

$$\begin{aligned} \det\{z\mathbf{I} - \mathbf{A}\} &= \det \left\{ \begin{bmatrix} z + c_2 & c_1 & c_0 \\ -1 & z & 0 \\ 0 & -1 & z \end{bmatrix} \right\} \\ &= (z + c_2) \det \left\{ \begin{bmatrix} z & 0 \\ -1 & z \end{bmatrix} \right\} - c_1 \det \left\{ \begin{bmatrix} -1 & 0 \\ 0 & z \end{bmatrix} \right\} + c_0 \det \left\{ \begin{bmatrix} -1 & z \\ 0 & -1 \end{bmatrix} \right\} \\ &= (z + c_2)(z^2 + 1 \cdot 0) - c_1(-z - 0^2) + c_0((-1)^2 - 0z) \\ &= z^3 + c_2 z^2 + c_1 z + c_0 = p(z). \end{aligned}$$

So the eigenvalues of the companion matrix \mathbf{A} are exactly the roots of the monic polynomial whose (negative) coefficients correspond to the first row of \mathbf{A} . This is not a coincidence; it is by design.

Explore 7.0.1: Examine the case $n = 1$.

For the general $n \times n$ case the same process yields:

$$\begin{aligned} \det\{zI - A\} &= \det \left\{ \begin{bmatrix} z + c_{n-1} & c_{n-2} & \dots & c_1 & c_0 \\ -1 & z & 0 & \dots & 0 \\ 0 & -1 & z & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & z \end{bmatrix} \right\} = (z + c_{n-1}) \det \underbrace{\left\{ \begin{bmatrix} z & 0 & \dots & 0 \\ -1 & z & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -1 & z \end{bmatrix} \right\}}_{(n-1) \times (n-1)} \\ &\quad - c_{n-2} \det \underbrace{\left\{ \begin{array}{c|ccccc} -1 & 0 & 0 & \dots & 0 \\ \hline 0 & z & 0 & \dots & 0 \\ 0 & -1 & z & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & z \end{array} \right\}}_{(n-1) \times (n-1)} + c_{n-3} \det \underbrace{\left\{ \begin{array}{c|ccccc} -1 & z & 0 & 0 & \dots & 0 \\ 0 & -1 & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & z & 0 & \dots & 0 \\ 0 & 0 & -1 & z & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 & z \end{array} \right\}}_{(n-1) \times (n-1)} + c_{n-4} \dots \\ &= (z + c_{n-1})z^{n-1} + c_{n-2}z^{n-2} + c_{n-3}z^{n-3} + \dots \end{aligned}$$

The general pattern is that we have a **block diagonal** matrix for the c_{n-k} term, where the first block is $(k-1) \times (k-1)$ **upper triangular** and has determinant $(-1)^{k-1}$, and the second block is **lower triangular** and

has determinant z^{n-k} , leading to a final overall contribution to the determinant of $c_{n-k}z^{n-k}$, for $k = 1, \dots, n$, because $\det\left\{\begin{bmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}\right\} = \det\{\mathbf{B}\} \det\{\mathbf{C}\}$. Thus

$$\det\{z\mathbf{I} - \mathbf{A}\} = c_0 + c_1 z + \cdots + c_{n-1} z^{n-1} + z^n = p(z).$$

In words: the polynomial $p(z)$ is the **characteristic polynomial** of \mathbf{A} .

It is also its **minimal polynomial**. (See next page.)

Due to this property, the matrix \mathbf{A} so defined is called a **companion matrix** of the monic polynomial $p(z)$.

The **eigenvalues** of the **companion matrix** \mathbf{A} are exactly the **roots** of the **monic polynomial** whose (negative) coefficients correspond to the first row of \mathbf{A} .

There are other rearrangements of \mathbf{A} (transposes and other permutations) that have the same property.

Practical implementation

To match the arrangement given above, use this one-line JULIA function:

```
compan = c -> [-transpose(reverse(c)); [I zeros(length(c)-1)]]
```

$\mathbf{A} = \text{compan}(\mathbf{c})$

where \mathbf{c} is the (column) vector with elements $(c_0, c_1, \dots, c_{n-1})$.

(This code needs $n \geq 1$ to work.)

However, in practice this matrix is perhaps used more for analysis than for implementation.

Polynomial matrix functions

$$A \in \mathbb{F}^{N \times N}$$

$$\begin{aligned} p(z) &= z^3 - 5z \\ p(A) &= A^3 - 5A \end{aligned}$$

The **Cayley-Hamilton theorem** says that the **characteristic polynomial** $p(z)$ of any matrix $N \times N$ matrix A has the property that

$$\underbrace{p(A)}_{\text{sum of powers}} = \sum_{k=0}^N c_k A^k = c_0 I_N + \sum_{k=1}^N c_k A^k = 0_{N \times N},$$

$$p(z) = \det(zI - A)$$

where we use the matrix power property (see Ch. 8) that $\boxed{A^0 = I}$ for a square matrix A .

Here is a proof of this property for the case of **diagonalizable** matrices. If $\underbrace{A = V \Lambda V^{-1}}$ then

$$p(A) = \sum_{k=0}^N c_k \underbrace{V \Lambda^k V^{-1}}_A = V \left(\sum_{k=0}^N c_k \Lambda^k \right) V^{-1} = V \operatorname{Diag} \left\{ \sum_{k=0}^N c_k \lambda_i^k \right\} V^{-1} = V \operatorname{Diag} \{p(\lambda_i)\} V^{-1} = 0_{N \times N}$$

because each eigenvalue λ_i is a root of the characteristic polynomial.

The proof for non-diagonalizable matrices uses the **Jordan form**.

$$\begin{aligned} \tilde{A} &= A \\ A &= V \Lambda V^{-1} \\ \tilde{A} &= V \Lambda^2 V^{-1} \end{aligned}$$

Define. The **minimal polynomial** is the monic polynomial $\mu(z)$ having *least degree* for which $\mu(A) = 0_{N \times N}$

A matrix is **diagonalizable** iff its **minimal polynomial** is a product of distinct factors.

The **minimal polynomial** of a **companion matrix** for $p(z)$ is $p(z)$, the **characteristic polynomial**.

\Rightarrow A **companion matrix** for $p(z)$ is **diagonalizable** iff $p(z)$ has distinct roots. ★ ★

In general, the **characteristic polynomial** for a $N \times N$ matrix has the form $p(z) = (z - z_1)^{p_1} \cdots (z - z_K)^{p_K}$ where K is the number of distinct eigenvalues (roots) and where $p_1 + \cdots + p_K = N$.

The corresponding **minimal polynomial** for a $N \times N$ matrix has the form $p(z) = (z - z_1)^{q_1} \cdots (z - z_K)^{q_K}$ where $1 \leq q_k \leq p_k$ and $1 \leq q_1 + \cdots + q_K \leq N$.

$$P(A) = 0$$

Example. Consider the matrix $A = \beta I_N$. The **characteristic polynomial** is $p(z) = (z - \beta)^N$, whereas the **minimal polynomial** is $\mu(z) = z - \beta$. In this example:

$$\mu(z) = 1 \cdot z^1 + (-\beta)z^0$$

$$\mu(A) = 1 \cdot A^1 + (-\beta)A^0 = A - \beta I = 0_{N \times N}$$

2. The degree of the minimal polynomial of a $N \times N$ matrix is always in the set $\{1, \dots, N\}$.

A: True

even for $A = 0$

B: False

??

Exercise 7-2-1. Given a polynomial $p(z)$, determine another very simple matrix for which the eigenvalues $\{\lambda_i\}$ of that matrix are exactly the roots $\{z_i\}$ of that polynomial.

$$p(z) = \underbrace{z^2 - 7z + 10}_{\text{factored}} = \underbrace{(z-2)(z-5)}_{\text{factored}}$$

$$C = \begin{bmatrix} 7 & -10 \\ 1 & 0 \end{bmatrix} = V \Lambda V^{-1} \quad \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix}$$

Example. Consider the (non-diagonalizable) matrix $\mathbf{A} = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix}$ that has eigenvalues $(\lambda, \lambda, \lambda)$ and characteristic polynomial $p(z) = (z - \lambda)^3$. Clearly $(\mathbf{A} - \lambda \mathbf{I}) \neq 0$ but one can verify that $(\mathbf{A} - \lambda \mathbf{I})^2 = 0$. Thus the **minimal polynomial** for this matrix is $\mu(z) = (z - \lambda)^2$.

$$\mu(z) = (z - \lambda)^2 \quad q \in \{1, 2, 3\}$$

$$(\mathbf{A} - \lambda \mathbf{I})^3 = p(\mathbf{A}) = 0$$

(Read)

Example. Consider the monic polynomial $p(z) = z^2 - 6z + 9 = (z - 3)^3$ with corresponding **companion matrix** $\mathbf{A} = \begin{bmatrix} 6 & -9 \\ 1 & 0 \end{bmatrix}$ for which $\mathbf{A} - 3\mathbf{I} = \begin{bmatrix} 3 & -9 \\ 1 & -3 \end{bmatrix}$.

The only options for the **minimal polynomial** are $\mu(z) = (z - 3)^2$ and $\mu(z) = (z - 3)$.

Because $\mathbf{A} - 3\mathbf{I} \neq 0$, the minimal polynomial must be $\mu(z) = (z - 3)^2 = p(z)$.

This is expected because $\mu(z) = p(z)$ for all companion matrices.

Both this example and the preceding one illustrate that a minimal polynomial can have repeated roots.

Eigenvectors of companion matrices

One can find **eigenvectors** of any **companion matrix** by inspection (*i.e.*, by guess and check).

$$\text{If } \mathbf{v} = \begin{bmatrix} t^{n-1} \\ \vdots \\ t \\ 1 \end{bmatrix}, t \in \mathbb{C}, \text{ then } A\mathbf{v} = \begin{bmatrix} -c_{n-1} & -c_{n-2} & \dots & -c_1 & -c_0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} t^{n-1} \\ \vdots \\ t \\ 1 \end{bmatrix}$$

$= \begin{bmatrix} -c_{n-1}t^{n-1} - \dots - c_1t - c_0 \\ t^{n-1} \\ \vdots \\ t \end{bmatrix} = ? \begin{bmatrix} t^n \\ t^{n-1} \\ \vdots \\ t \\ 1 \end{bmatrix} = t \begin{bmatrix} t^{n-1} \\ \vdots \\ t \\ 1 \end{bmatrix} = t\mathbf{v},$

where the inner equality holds iff

$$-c_{n-1}t^{n-1} - \dots - c_1t - c_0 = t^n, \text{ i.e., } 0 = t^n + c_{n-1}t^{n-1} + \dots + c_1t + c_0 = p(t).$$

In other words, if t is a root of the polynomial $p(z)$, then the corresponding vector \mathbf{v} is an eigenvector of A with eigenvalue t . Note that \mathbf{v} is a nonzero vector by definition.

Vandermonde matrices and diagonalizing a companion matrix

For each root of $p(z)$, the **companion matrix** A has a corresponding (unit norm) eigenvector.

A **companion matrix** A for $p(z)$ is **diagonalizable** iff $p(z)$ has no repeated roots.

(Recall a matrix A is **diagonalizable** iff its **minimal polynomial** is a product of distinct factors.)

If the roots $\{z_1, \dots, z_n\}$ of $p(z)$ are *distinct*, then the corresponding companion matrix A is diagonalizable:

$$V^{-1}AV = \text{Diag}\{z_1, \dots, z_n\},$$

$$V = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \\ z_1^{n-1} & z_2^{n-1} & \cdots & z_n^{n-1} \\ z_1^{n-2} & z_2^{n-2} & \cdots & z_n^{n-2} \\ \vdots & \vdots & \vdots & \vdots \\ z_1 & z_2 & \cdots & z_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}.$$

where the eigenvector matrix V is an $n \times n$ **Vandermonde matrix** of the form on the right.

$$A = V \text{Diag}(z_i) V^{-1}$$

Fact. A **Vandermonde matrix** is **invertible** iff the z_k values are distinct.

Vandermonde matrices are often transposed or reverse versions of the matrix shown above.

Example. In signal processing, the most important Vandermonde matrix is the N -point **DFT matrix**, which is a transposed and flipped version of the above matrix with $z_k = e^{-i2\pi k/N}$.

Rectangular Vandermonde matrices have been used as **frames** [1].



$$p(z) = (z-2)(z-2)$$

7.12

$$V = \alpha \begin{bmatrix} z \\ 1 \end{bmatrix}$$

3. The companion matrix for the polynomial $p(z) = z^2 - 4z + 4$ is diagonalizable.

A: True

B: False

??

Example. The companion matrix for the polynomial $p(z) = z^2 - 6z + 9$ is $A = \begin{bmatrix} 6 & -9 \\ 1 & 0 \end{bmatrix}$.

The two eigenvalues are both at $z = 3$ and $A - 3I = \begin{bmatrix} 3 & -9 \\ 1 & -3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \end{bmatrix} [1 \ -3]$, so all eigenvectors are in

$\text{span}^\perp \left(\begin{bmatrix} 1 \\ -3 \end{bmatrix} \right)$, i.e., multiples of $v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$. This companion matrix is not diagonalizable. Try this in JULIA:

 $A = [6 \ -9; 1 \ 0]$

eigenvalues:

 $(l, V) = \text{eigen}(A)$

2-element ArrayComplexFloat64, 1:

 $3.0 - 3.725290298461914\text{e-8im}$ $3.0 + 3.725290298461914\text{e-8im}$

eigenvectors:

2x2 ArrayComplexFloat64, 2:

 $0.948683-0.0\text{im} \quad 0.948683+0.0\text{im}$ $\rightarrow 0.316228+3.9268\text{e-9im} \quad 0.316228-3.9268\text{e-9im}$

$$\begin{bmatrix} 3/\sqrt{10} \\ 1/\sqrt{10} \end{bmatrix}$$

The two columns of V are almost identical to $v_1 = \begin{bmatrix} 3 \\ 1 \end{bmatrix} / \sqrt{10}$. Oddly, $\text{rank}(V)$ returns 2.

Even more oddly, $V * \text{Diagonal}(l) * \text{inv}(V)$ returns a matrix that is quite close to A .

$\underbrace{\text{no warning}! ?}$

$$V = \begin{bmatrix} 3 & 3 \\ 1 & 1 \end{bmatrix} / \sqrt{10}$$

not
invertible

But `cond(V)` returns `2.68e8` and `svdvals(V)` gives `1.414, 5.27e-9`, showing that `V` is essentially singular.

[Explore 7.0.2](#): What is happening here?

Using companion matrices to check for common roots of two polynomials

(Read)

First we need another tool.

Kronecker sum

Define. The **Kronecker sum** of $M \times M$ matrix \mathbf{A} with a $N \times N$ matrix \mathbf{B} is the following $MN \times MN$ matrix [2, p.143]:

$$\underbrace{\mathbf{A} \oplus \mathbf{B}}_{= (\mathbf{I}_N \otimes \mathbf{A}) + (\mathbf{B} \otimes \mathbf{I}_M)} = (\mathbf{I}_N \otimes \mathbf{A}) + (\mathbf{B} \otimes \mathbf{I}_M).$$

L§13.2

Wikipedia uses this definition:

$$\mathbf{A} \oplus \mathbf{B} = (\mathbf{A} \otimes \mathbf{I}_N) + (\mathbf{I}_M \otimes \mathbf{B}).$$

The two definitions are the same to within a permutation, *i.e.*, there is a $MN \times MN$ permutation matrix \mathbf{P} such that

$$\mathbf{P} ((\mathbf{I}_N \otimes \mathbf{A}) + (\mathbf{B} \otimes \mathbf{I}_M)) \mathbf{P}' = (\mathbf{A} \otimes \mathbf{I}_N) + (\mathbf{I}_M \otimes \mathbf{B}).$$

Because eigenvalues are invariant to similarity transforms, both definitions have the same eigenvalues so the following Fact holds for both definitions.

Properties

- Scaling: $\alpha(\mathbf{A} \oplus \mathbf{B}) = (\alpha\mathbf{A}) \oplus (\alpha\mathbf{B})$
- Associative: $(\mathbf{A} \oplus \mathbf{B}) \oplus \mathbf{C} = \mathbf{A} \oplus (\mathbf{B} \oplus \mathbf{C})$
- *Not* commutative: $\mathbf{A} \oplus \mathbf{B} \neq \mathbf{B} \oplus \mathbf{A}$ in general.

Explore 7.0.3: Find a counter-example.

Fact. [2, Thm. 13.16]

If \mathbf{A} has eigenvalues $\{\lambda_m, m = 1, \dots, M\}$ and \mathbf{B} has eigenvalues $\{\mu_n, n = 1, \dots, N\}$, then the MN eigenvalues of $\mathbf{A} \oplus \mathbf{B}$ are

$$\lambda_m + \mu_n, \quad m = 1, \dots, M, \quad n = 1, \dots, N.$$

Proof sketch. If $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{B}\mathbf{y} = \mu\mathbf{y}$, then

$$\begin{aligned} (\mathbf{A} \oplus \mathbf{B})(\mathbf{y} \otimes \mathbf{x}) &= (\mathbf{I}_N \otimes \mathbf{A})(\mathbf{y} \otimes \mathbf{x}) + (\mathbf{B} \otimes \mathbf{I}_M)(\mathbf{y} \otimes \mathbf{x}) = (\mathbf{y} \otimes (\mathbf{A}\mathbf{x})) + ((\mathbf{B}\mathbf{y}) \otimes \mathbf{x}) \\ &= (\mathbf{y} \otimes (\lambda\mathbf{x})) + ((\mu\mathbf{y}) \otimes \mathbf{x}) = (\lambda + \mu)(\mathbf{y} \otimes \mathbf{x}). \end{aligned}$$

Application: checking for common roots of two polynomials

If $p_1(z)$ and $p_2(z)$ are two monic polynomials (possibly of different degrees) having corresponding companion matrices \mathbf{A} and \mathbf{B} , then $p_1(z)$ and $p_2(z)$ share a common root iff \mathbf{A} and \mathbf{B} have a common eigenvalue, i.e., iff there exists some λ_m and μ_n such that $\lambda_m = \mu_n$. In other words, $p_1(z)$ and $p_2(z)$ share a common root iff the matrix $\mathbf{A} \oplus (-\mathbf{B})$ has a zero eigenvalue, i.e., is singular. Thus we can determine if two polynomials have a common root *without* performing any eigendecomposition. This property is explored in HW.

Another application is finding zeros of a univariate equation $f(x) = 0$ [3].

$$\mathbf{x} \rightarrow \boxed{G_N} \rightarrow \mathbf{y} = G_N \mathbf{x}$$

$$y[n] = x[n - 1 \bmod N]$$

7.2 Circulant matrices

A special case of the **companion matrix** considered above corresponds to the simple monic polynomial

$$p(z) = z^N - 1.$$

$$y[n] = x[n] \otimes s[n-1]$$

Circular shift operation

Here $c_0 = -1$ and all other coefficients are 0, so the corresponding **companion matrix** is simply:

$$\text{"Generator"} \quad G_N \triangleq \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

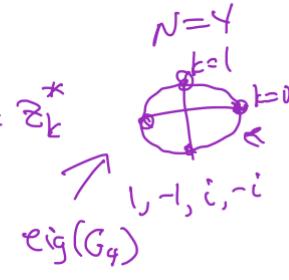
$h[n] \leftrightarrow e^{-j\frac{2\pi}{N}nk}$
 $\uparrow \text{DFT}$
 $e^{-j\frac{2\pi}{N}k}$
 $k=0, \dots, N-1$

This special case is a **circulant matrix**. (Indeed it is the parent of all circulant matrices, as shown below.) $z^N = 1$

What are the eigenvectors of this matrix G_N ? The roots of $p(z)$ here are solutions to $z^N - 1 = 0$, i.e., the N th roots of unity: $z_k = e^{-j2\pi k/N}$ $k = 0, \dots, N - 1$. An eigenvector corresponding to the k th root z_k is

$$v_k \triangleq \frac{1}{\sqrt{N}} z_k^{1-N} \begin{bmatrix} z_k^{N-1} \\ z_k^{N-2} \\ \vdots \\ z_k \\ 1 \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} z_k^0 \\ z_k^{-1} \\ \vdots \\ z_k^{2-N} \\ z_k^{1-N} \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ w_k \\ \vdots \\ w_k^{N-2} \\ w_k^{N-1} \end{bmatrix}, \quad w_k \triangleq e^{j2\pi k/N} = z_k^*$$

$\underbrace{\quad \quad \quad}_{\text{eigvec of } G_N}$



with corresponding eigenvalue z_k .

Note that w_k for $k = 0, \dots, N - 1$ are also (distinct!) roots of $p(z)$, spaced equally around the unit circle in the complex plane.

The $N \times N$ matrix formed from all N of the eigenvectors is a scaled Vandermonde-type matrix:

$$\mathbf{Q} \triangleq [\mathbf{v}_0 \ \dots \ \mathbf{v}_{N-1}] = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & w_k & \dots & w_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & w_k^{N-2} & \dots & w_{N-1}^{N-2} \\ 1 & w_k^{N-1} & \dots & w_{N-1}^{N-1} \end{bmatrix} \quad \text{orthonormal DFT matrix} \quad (7.1)$$

Because the roots $\{w_k\}$ are distinct, this Vandermonde-like matrix is invertible. But here we can say much more: the columns of \mathbf{Q} here are **orthonormal**, so \mathbf{Q} is a **unitary matrix**. Thus

$$\mathbf{G}_N = \mathbf{Q} \operatorname{Diag}\{z_0, \dots, z_{N-1}\} \mathbf{Q}' \text{, or equivalently: } \mathbf{Q}' \mathbf{G}_N \mathbf{Q} = \operatorname{Diag}\{z_0, \dots, z_{N-1}\}.$$

Multiplying matrix \mathbf{Q}' by a vector \mathbf{x} of length N corresponds to taking the N -point (orthonormal) **DFT** of \mathbf{x} .

Explore 7.0.4: Verify that the columns of \mathbf{Q} are orthonormal vectors, i.e., $\langle \mathbf{v}_k, \mathbf{v}_l \rangle = 0$ for $k \neq l$.

Explore 7.0.5: Verify that \mathbf{Q} is a symmetric (not Hermitian!) matrix.

Now consider a general **circulant matrix**:

$$C = \begin{bmatrix} c_0 & c_{n-1} & \cdots & c_2 & c_1 \\ c_1 & c_0 & c_{n-1} & \cdots & c_2 \\ \vdots & c_1 & c_0 & \ddots & \vdots \\ c_{n-2} & c_{n-1} & \ddots & \ddots & c_{n-1} \\ c_{n-1} & c_{n-2} & \cdots & c_1 & c_0 \end{bmatrix}.$$

"generator"

Because \mathbf{G}_N is a circulant permutation matrix, powers of \mathbf{G}_N provide other shifts:

$$\mathbf{G}_N^2 = \mathbf{G}_N \mathbf{G}_N = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 1 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \end{bmatrix}.$$

(circular)
shift by
2 samples

\mathbf{G}_N^2

So by defining $\mathbf{G}_N^0 = \mathbf{I}$, we write the general circulant matrix C in terms of powers of \mathbf{G}_N as follows:

$$C = c_0 \mathbf{I} + c_1 \mathbf{G}_N + c_2 \mathbf{G}_N^2 + \dots + c_{N-1} \mathbf{G}_N^{N-1} = c_0 \mathbf{I} + \sum_{n=1}^{N-1} c_n \mathbf{G}_N^n = \sum_{n=0}^{N-1} c_n \mathbf{G}_N^n.$$

first column of C

The matrix \mathbf{G}_N^n has the same (orthonormal) eigenvectors as \mathbf{G}_N because $\mathbf{G}_N^2 \mathbf{v}_k = \mathbf{G}_N (\mathbf{G}_N \mathbf{v}_k) = z_k^2 \mathbf{v}_k$.

$$A v = \lambda v \Rightarrow A^2 v = \lambda^2 v$$

Thus we can **diagonalize** C as follows:

$$\begin{aligned} Q' C Q &= Q' \left(\sum_{n=0}^{N-1} c_n G_N^n \right) Q = \sum_{n=0}^{N-1} c_n (Q' G_N^n Q) = \sum_{n=0}^{N-1} c_n \text{Diag}\{z_0^n, \dots, z_{N-1}^n\} \\ &= \text{Diag} \left\{ \sum_{n=0}^{N-1} c_n z_0^n, \dots, \sum_{n=0}^{N-1} c_n z_{N-1}^n \right\} = \Lambda = \text{Diag}\{\lambda_k\}, \quad \lambda_k \triangleq \sum_{n=0}^{N-1} c_n z_k^n. \end{aligned}$$

DFT of
[first] column!

Rearranging, a unitary eigendecomposition of C is

$$\underline{C = Q \Lambda Q'}, \quad \lambda_k \triangleq \sum_{n=0}^{N-1} c_n z_k^n. \quad (7.2)$$

This decomposition holds for *any* circulant matrix C . In other words, all circulant matrices have the same unitary eigenvectors (the orthonormal DFT basis).

The k th eigenvalue of C is

$$\lambda_k = \sum_{n=0}^{N-1} c_n z_k^n = \sum_{n=0}^{N-1} c_n e^{-i2\pi nk/N}, \quad k = 0, \dots, N-1,$$

which is the (ordinary, not orthonormal) DFT of the first column of C .

One can use the **fast Fourier transform (FFT)** to find all N eigenvalues with $O(N \log N)$ operations,

Relationship to DFT properties from DSP

(Read)

- The eigendecomposition (7.2) is the linear algebra expression of the DFT property that circular convolution corresponds to multiplying spectra:

$$h[n] * x[n] \xrightarrow{\text{DFT}} H[k]X[k].$$

- The fact that all circulant matrices have the same unitary basis, means that all circulant matrices **commute**, (HW) which is the linear algebra expression of the DFT property that circular convolution operations commute:

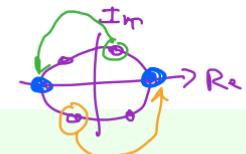
$$h[n] * x[n] = x[n] * h[n] \xrightarrow{\text{DFT}} H[k]X[k] = X[k]H[k].$$

- The associative property of circular convolution is also related to the fact all circulant matrices commute:

$$h[n] * (g[n] * x[n]) = (h[n] * g[n]) * x[n].$$

- The identity matrix \mathbf{I} is a circulant matrix, so $\mathbf{I} = \mathbf{Q}\mathbf{I}\mathbf{Q}'$ where \mathbf{Q} is the orthonormal DFT matrix defined in (7.1). The fact that the DFT of the Kronecker impulse signal $x[n] = [1 \ 0 \ \dots \ 0]$ is a flat spectrum $X[k] = 1$ is equivalent to the fact that the eigenvalues of \mathbf{I} are all unity.

$$e^{\pm i \frac{2\pi}{6} k}$$



4. The number of distinct eigenvalues of the matrix G_6^3 is:

A: 1

B: 2

C: 3

D: 4

E: 6

??

$$\text{eigs}(G_6^3) = \left\{ \left(e^{\pm i \frac{2\pi}{6} k} \right)^3 \right\} = \left\{ e^{\pm i \pi k} \right\} \quad \times$$

$$G_6^3$$

Practical implementation

A **circulant matrix** is a special case of a **Toeplitz matrix**. To construct a circulant matrix in JULIA:

```
using ToeplitzMatrices
```

```
A = Circulant(1:3)
```

In JULIA, the resulting variable is a special type of matrix (somewhat like `Diagonal` or `Sparse`) that performs matrix-vector multiplication efficiently.

Directly multiplying a $N \times N$ matrix C by a length- N vector, i.e., $y = Cx$ via $y = C * x$, would require $O(N^2)$ operations. When C is circulant, we can compute the same product in $O(N \log_2 N)$ operations [4] using the **fast Fourier transform (FFT)** as follows:

 $y = \text{ifft}(\underline{\text{fft}(c)} \ .\star \ \text{fft}(x))$ cf $y = \underline{Q(\Lambda(Q'x))}$,

where c is the first column of C .

The resulting y will be the same as if we did $y = C * x$ except for some small numerical differences due to finite precision. These differences are rarely (if ever) important in practice.

One cautionary note is that if C and x are both real then of course $y = Cx$ is also real. However, the result of the above set of `fft` operations may end up with some very small imaginary part, depending on the implementation, so you might need to take the real part at the end.

The JULIA type `Circulant` in the `ToeplitzMatrices.jl` package takes care of this automatically by overloading the `*` operation. Other operations implemented elegantly include `inv` and `pinv`.

Spectral properties of a circulant matrices*opnorm(CC, 2)*Every circulant matrix is a **normal matrix**.

(HW)

5. If C is a circulant matrix, then its **spectral norm** is $\sigma_1 = \frac{\text{maximum}(\text{abs.}(\text{fft}(C[:, 1])))}{\lambda_k}$ (?)

A: True

B: False

??

Assume that using FFTW was invoked first.

$$C = Q \Lambda Q' = \underbrace{Q}_{U} \underbrace{\text{Diag}(\text{sign}(\lambda_k))}_{\substack{\text{almost } \Sigma \\ (\text{permute})}} \underbrace{\text{Diag}(|\lambda_k|)}_{V} \underbrace{Q'}_{V'}$$

Inverting a circulant matrix

Any circulant matrix C has eigendecomposition $C = Q \Lambda Q'$. If the eigenvalues are nonzero, then C is invertible and its inverse is

$$C^{-1} = Q \Lambda^{-1} Q'.$$

In other words, multiplying the inverse of C times a vector x , as in $C^{-1}x$, can be done with $O(N \log N)$ FLOPS, which is quite fast compared to a general invertible matrix. Specifically: $C^{-1}x = Q(\Lambda^{-1}(Q'x))$.

7.3 Toeplitz matrices

Circulant matrices are special cases of **Toeplitz** matrices. The broader family of Toeplitz matrices is of great interest in signal processing because Toeplitz matrix structure corresponds to linear shift invariant operations.

Toeplitz matrix multiplication with a vector

(Read)

As discussed earlier, convolution of a finite-length signal with a finite-length filter is equivalent to multiplying a Toeplitz matrix by a vector with the signal samples. Roughly:

$$h * x \equiv \mathbf{H}x$$

for an appropriate matrix \mathbf{H} defined from the filter h .

(HW)

If $x \in \mathbb{F}^N$ and the filter length is L then ordinary convolution requires about $O(LN)$ multiplies whereas the simplest implementation of $\mathbf{H}x$ would require $O((L + N - 1)N) = O(N^2)$ multiplies. When $L \ll N$ we can store \mathbf{H} as a sparse matrix in which case $\mathbf{H}x$ requires $O(LN)$ multiplies.

When L is not small, it is more efficient to use zero-padding and perform convolution using an FFT-based approach. The `fftfilt` function in the `DSP.jl` package does this. The approach requires roughly $O(N \log N)$ multiplies.

In terms of matrices, it is basically like extending \mathbf{H} to make a bigger circulant matrix, zero-padding x as well, and then doing circulant matrix-vector multiplication using FFT operations and then trimming the result to the correct output size.

Inverting a Toeplitz matrix

(Read)

There are also efficient methods for inverting Toeplitz matrices [5–10].

Factoring a Toeplitz matrix

If \mathbf{V} is a $N \times K$ **Vandermonde** matrix for which $V_{nk} = z_k^{n-1} = e^{\imath 2\pi(n-1)f_k}$, where the $\{f_k\}$ are distinct, and $\mathbf{D} = \text{Diag}\{d_1, \dots, d_K\} \in \mathbb{F}^{K \times K}$, then the following is a $N \times N$ rank- K **Toeplitz** matrix:

$$\mathbf{T} = \mathbf{V} \mathbf{D} \mathbf{V}' = \sum_{k=1}^K d_k \mathbf{v}_k \mathbf{v}_k'$$

$$\text{because } T_{nm} = \sum_{k=1}^K d_k V_{nk} V_{mk}^* = \sum_{k=1}^K d_k e^{\imath 2\pi(n-1)f_k} e^{-\imath 2\pi(m-1)f_k} = \sum_{k=1}^K d_k e^{\imath 2\pi(n-m)f_k}.$$

This fact is useful in some sampling/interpolation problems [11] and some imaging applications [12].

There is a partial converse of this fact.

If \mathbf{T} is a **positive semidefinite Toeplitz** matrix with rank K , then there exists \mathbf{V} and \mathbf{D} of the above form such that $\mathbf{T} = \mathbf{V} \mathbf{D} \mathbf{V}'$ where the elements of \mathbf{D} are positive [13]. WLOG we can sort them from largest to smallest.

6. This form is a compact SVD factorization of \mathbf{T} .

A: True

B: False

??

7.4 Power iteration

(Read)

For large matrices, performing a full eigendecomposition is impractical. Fortunately, sometimes we need only the largest magnitude eigenvalue (and/or corresponding eigenvector). The **power iteration** is an iterative method for such tasks. It is a classic method with numerous applications, including recent training of a CNN for image denoising [14].

First consider powers of a square diagonalizable matrix $\mathbf{A} \in \mathbb{F}^{N \times N}$:

$$\underbrace{\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}}_{\text{Diagonalizable}} \implies \mathbf{A}^2 = \mathbf{A}\mathbf{A} = (\mathbf{V}\Lambda\mathbf{V}^{-1})(\mathbf{V}\Lambda\mathbf{V}^{-1}) = \mathbf{V}\Lambda^2\mathbf{V}^{-1}.$$

Proceeding by induction shows more generally that

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1} \implies \mathbf{A}^k = \mathbf{V}\Lambda^k\mathbf{V}^{-1}, \quad \forall k \in \mathbb{N}.$$

Thus for a vector \mathbf{x}_0 of appropriate size:

$$\mathbf{A}^k \mathbf{x}_0 = \mathbf{V}\Lambda^k \overbrace{\mathbf{V}^{-1} \mathbf{x}_0}^{\triangleq \mathbf{z}} = \sum_{n=1}^N (\lambda_n^k z_n) \mathbf{v}_n. \tag{7.3}$$

Now make some assumptions:

- WLOG we order the eigenvalues with decreasing magnitudes: $|\lambda_1| \geq |\lambda_2| \geq \dots \geq |\lambda_N|$.
- Assume $\lambda_1 \neq 0$ to avoid the trivial case of $\mathbf{A} = 0$.

- Assume $z_1 = [\mathbf{V}^{-1}\mathbf{x}_0]_1 \neq 0$, i.e., the initial vector \mathbf{x}_0 has a nonzero coordinate in the direction associated with the dominant eigenvalue λ_1 . Choosing \mathbf{x}_0 at random ensures $z_1 \neq 0$ with probability 1.

Then it follows from the linear independence of the columns of \mathbf{V} and the summation (7.3) that $\mathbf{A}^k \mathbf{x}_0 \neq \mathbf{0}$ for all $k \in \mathbb{N}$, so we can normalize it to have unit norm and define:

$$\mathbf{x}_k \triangleq \frac{\mathbf{A}^k \mathbf{x}_0}{\|\mathbf{A}^k \mathbf{x}_0\|_2}, \quad k = 1, 2, \dots \quad (7.4)$$

This definition leads to the following simple recursion:

$$\mathbf{x}_{k+1} = \frac{\mathbf{A}^{k+1} \mathbf{x}_0}{\|\mathbf{A}^{k+1} \mathbf{x}_0\|_2} = \frac{\mathbf{A}(\mathbf{A}^k \mathbf{x}_0)}{\|\mathbf{A}(\mathbf{A}^k \mathbf{x}_0)\|_2} = \frac{\mathbf{A}(\mathbf{A}^k \mathbf{x}_0)/\|\mathbf{A}^k \mathbf{x}_0\|_2}{\|\mathbf{A}(\mathbf{A}^k \mathbf{x}_0)/\|\mathbf{A}^k \mathbf{x}_0\|_2\|_2} = \frac{\mathbf{A}\mathbf{x}_k}{\|\mathbf{A}\mathbf{x}_k\|_2}, \quad k = 0, 1, \dots$$

For implementation we use this recursive form that is called the **power iteration**:

$$\boxed{\mathbf{x}_{k+1} = \frac{\mathbf{A}\mathbf{x}_k}{\|\mathbf{A}\mathbf{x}_k\|_2}, \quad k = 0, 1, \dots} \quad (7.5)$$

Returning to the summation form (7.3), one can show with some algebra [wiki] that

$$\mathbf{x}_k \propto \mathbf{A}^k \mathbf{x}_0 = z_1 \lambda_1^k \left(\mathbf{v}_1 + \sum_{n=2}^N \left(\left(\frac{\lambda_n}{\lambda_1} \right)^k \frac{z_n}{z_1} \right) \mathbf{v}_n \right),$$

where \mathbf{v}_n denotes a (unit-norm) eigenvector of \mathbf{A} corresponding to λ_n , i.e., the n th column of \mathbf{V} .

Example. Consider the 1×1 matrix $\mathbf{A} = i$. Here $\mathbf{A}^k \mathbf{x}_0 = z_1 i^k$ so $\mathbf{x}_k = \mathbf{A}^k \mathbf{x}_0 / \|\mathbf{A}^k \mathbf{x}_0\| = z_1 i^k / |z_1 i^k| = e^{i\angle z_1} i^k$, which does not converge due to the oscillations of the i^k term! Casually speaking, people say “the power iteration converges to \mathbf{v}_1 ” but this 1×1 example shows such words are imprecise. In general one cannot claim that $\|\mathbf{x}_k - \alpha \mathbf{v}_1\| \rightarrow 0$, no matter how we might pick α .

Convergence of the power iteration

However, if the first eigenvalue magnitude dominates all others, *i.e.*, if

$$|\lambda_n| < |\lambda_1|, \quad n = 2, \dots, N,$$

then one can show that $\|e^{-i\angle z_1} (e^{-i\angle \lambda_1})^k \mathbf{x}_k - \mathbf{v}_1\| \rightarrow 0$, because $e^{-i\angle z_1} z_1 = |z_1|$ so

$$e^{-i\angle z_1} (e^{-i\angle \lambda_1})^k \mathbf{x}_k = \frac{|z_1| |\lambda_1|^k \left(\mathbf{v}_1 + \sum_{n=2}^N \left(\left(\frac{\lambda_n}{\lambda_1} \right)^k \frac{z_n}{z_1} \right) \mathbf{v}_n \right)}{\left\| z_1 \lambda_1^k \left(\mathbf{v}_1 + \sum_{n=2}^N \left(\left(\frac{\lambda_n}{\lambda_1} \right)^k \frac{z_n}{z_1} \right) \mathbf{v}_n \right) \right\|_2} \rightarrow \mathbf{v}_1.$$

In other words, as k increases, \mathbf{x}_k is approximately $e^{i\angle z_1} e^{ik\angle \lambda_1} \mathbf{v}_1$, meaning that it is approximately \mathbf{v}_1 times a unit-magnitude complex number.

Another way of saying this is $\|P_{\mathbf{v}_1}^\perp \mathbf{x}_k\| \rightarrow 0$ as $k \rightarrow \infty$.

Under all the conditions assumed above, one can also show that the **eigenvalue** converges:

$$\mathbf{x}_k' \mathbf{A} \mathbf{x}_k \rightarrow \lambda_1.$$

Here we do not need to worry about the extra phase term $e^{i\angle z_1} e^{ik\angle \lambda_1}$ because it cancels out due to \mathbf{x}'_k and \mathbf{x}_k :

$$(e^{i\phi} \mathbf{v}_1)' \mathbf{A} (e^{i\phi} \mathbf{v}_1) = e^{-i\phi} e^{i\phi} \mathbf{v}'_1 (\mathbf{A} \mathbf{v}_1) = \mathbf{v}'_1 (\lambda_1 \mathbf{v}_1) = \lambda_1.$$

- In words, if the magnitude of one eigenvalue magnitude dominates all others, and if $z_1 \neq 0$, then for large k the output of the power iteration \mathbf{x}_k is approximately a unit-norm eigenvector corresponding to that largest magnitude eigenvalue.
- Casually speaking we say “the power iteration converges to the eigenvector corresponding to the largest magnitude eigenvalue” but it is imprecise to say “the” eigenvector since we can always scale by -1 (or $e^{i\phi}$ more generally), and to make rigorous claims about convergence we must also think about the phase.
- The convergence rate of the power iteration is governed by $|\lambda_2/\lambda_1|$, so the bigger the “gap” between the first and second largest eigenvalue magnitudes, the faster the convergence.
- We have assumed \mathbf{A} is diagonalizable here, but the **power iteration convergence analysis** generalizes to non-diagonalizable matrices using the **Jordan form**. So sufficient conditions are:
 - \mathbf{A} is square
 - $z_1 = [\mathbf{V}^{-1} \mathbf{x}_0]_1 \neq 0$, where \mathbf{V} denotes the (linearly independent) **generalized eigenvectors** in the Jordan form, ordered so that $|\lambda_1|$ is largest.
 - $|\lambda_1|$ dominates the other eigenvalue magnitudes
- For analyzing a **Markov chain** (soon), we will see that the dominant eigenvalue of the transition matrix \mathbf{P} is real and positive. For any matrix where λ_1 is real and positive, we have that $e^{i\angle \lambda_1} = 1$ so there is no problematic oscillation term. In this case, under the sufficient conditions above, we can say that the power

iteration converges:

$$\mathbf{x}_k \rightarrow e^{i\angle z_1} \mathbf{v}_1.$$

For Markov chains, we will also see that \mathbf{v}_1 can be a nonnegative vector, so we can normalize it to *sum* to one, and if we initialize with a nonnegative vector \mathbf{x}_0 that sums to one then $e^{i\angle z_1} = 1$ and we can simply let $\mathbf{x}_{k+1} = \mathbf{P}\mathbf{x}_k$ and conclude that $\mathbf{x}_k \rightarrow \mathbf{v}_1$.

$$\mathbf{A} = \mathbf{U} \Sigma \mathbf{V}' \quad \mathbf{A} \mathbf{A}' = \mathbf{U} \Sigma \Sigma' \mathbf{V}'$$

7. To determine the principal left singular vector \mathbf{u}_1 , we apply the power iteration to
 A: \mathbf{A}' B: \mathbf{A} C: $\mathbf{A}\mathbf{A}'$ D: $\mathbf{A}'\mathbf{A}$ E: None of these

??

Twitter uses the **power iteration** to show users recommendations of whom to follow. See their [WTF paper](#).

Example. Here is an illustration of a representative case where the power iteration fails converge when the condition $[\mathbf{V}^{-1}\mathbf{x}_0]_1 \neq 0$ is not met, i.e., when $[\mathbf{V}^{-1}\mathbf{x}_0]_1 = 0$.

Consider $\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 2 \end{bmatrix}$, an asymmetric matrix having the following eigendecomposition:

$$\mathbf{A} = \mathbf{V} \Lambda \mathbf{V}^{-1}, \quad \mathbf{V} = \begin{bmatrix} 0 & 1/\sqrt{2} \\ 1 & -1/\sqrt{2} \end{bmatrix} = [\mathbf{v}_1 \ \mathbf{v}_2], \quad \Lambda = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{V}^{-1} = \begin{bmatrix} 1 & 1 \\ \sqrt{2} & 0 \end{bmatrix}.$$

Now initialize the power iteration with $\mathbf{x}_0 = \sqrt{2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$, for which $[\mathbf{V}'\mathbf{x}_0]_1 = -1 \neq 0$, but $[\mathbf{V}^{-1}\mathbf{x}_0]_1 = 0$. In this case we have $\mathbf{A}\mathbf{x}_0 = \mathbf{x}_0$, so the power iteration does not converge to \mathbf{v}_1 for this choice of \mathbf{x}_0 .

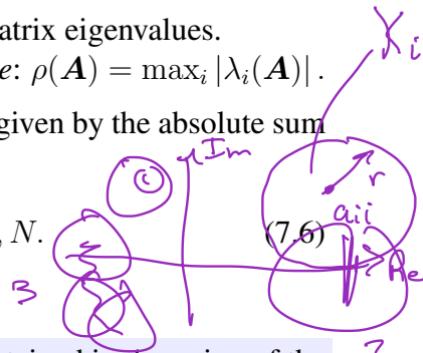
Geršgorin disk theorem

Before proceeding to further special matrices, we state an important result about matrix eigenvalues.

Recall that the **spectral radius** of a square matrix is its largest eigenvalue *magnitude*: $\rho(\mathbf{A}) = \max_i |\lambda_i(\mathbf{A})|$.

For a $N \times N$ matrix, the i th **Geršgorin (row) disk** is centered at a_{ii} and has radius given by the absolute sum of the off-diagonal elements of the i th row:

$$\mathcal{K}_i \triangleq \left\{ \lambda \in \mathbb{C} : |\lambda - a_{ii}| \leq r_i = \sum_{j \neq i} |a_{ij}| \right\}, \quad i = 1, \dots, N.$$



Fact. The **Geršgorin disk theorem** states that all eigenvalues of a matrix \mathbf{A} are contained in the union of the Geršgorin disks:

$$\lambda_i \in \bigcup_{i=1}^N \mathcal{K}_i$$

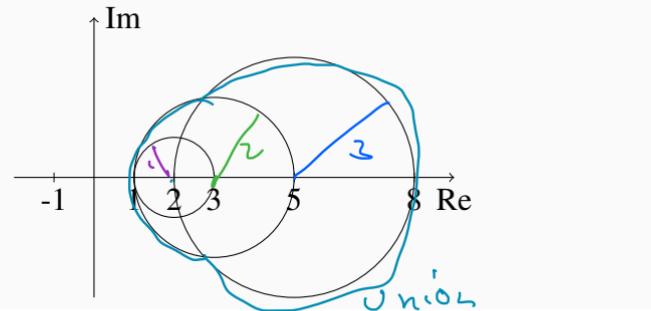
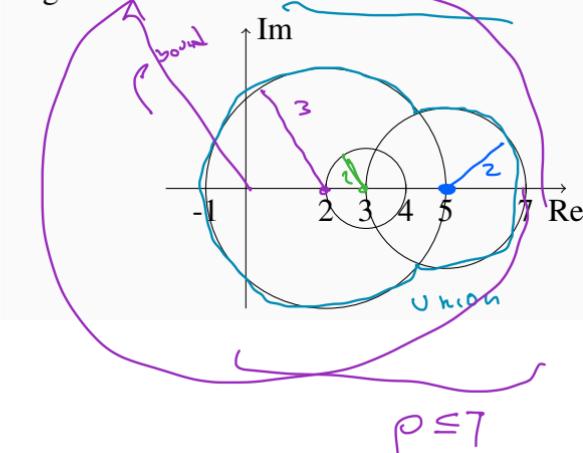
More generally, if any union of J disks is disjoint from all the other $N - J$ discs, then that union contains exactly J eigenvalues of \mathbf{A} .

Example. If \mathbf{P} is a permutation matrix, then each row has one element that is 1 and all others are zero. Thus either $\mathcal{K}_i = \{1\}$ or \mathcal{K}_i is the unit disk. So all the eigenvalues of a permutation matrix lie inside the unit disk. (In fact they lie on the unit circle, because all singular values of \mathbf{P} are 1 and \mathbf{P} is normal so its singular values are the sorted absolute values of its eigenvalues.)

Row and column Geršgorin disks

The disks in (7.6) use the off-diagonal row sums. The eigenvalues of a matrix and its transpose (not Hermitian transpose!) are the same, so one can also define (“column”) disks using the off-diagonal column sums. The union of those column disks must also contain all the eigenvalues, so the eigenvalues must be in the intersections of those two unions.

Example. For the matrix $A = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 5 & 0 \\ 0 & 1 & 3 \end{bmatrix}$, the row disks are shown below left; they contain the origin, so the matrix looks like it could be **singular**. The column disks are shown below right; they do not contain the origin so the matrix must be **invertible**.



$\rho \leq 8$ A invertible!

(lowest)

8. What is the best upper bound for the spectral radius of A ?

A: -1

B: 1

C: 5

D: 7

E: 8

??

Example. Consider the 2×2 permutation matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. This matrix is invertible because $P^2 = I_2$.

Each **Geršgorin disk** is simply the unit disk centered at the origin. Thus for a matrix to be **invertible** it is *sufficient* that the disk union does not cover the origin, but that is not a *necessary* condition for invertibility.

General spectral radius bounds

If $\lambda \in \mathcal{K}_i$, then applying the **reverse triangle inequality** to the Geršgorin disk theorem yields:

$$|\lambda| - |a_{ii}| \leq ||\lambda| - |a_{ii}|| \leq |\lambda - a_{ii}| \leq r_i \implies |\lambda| \leq |a_{ii}| + r_i = |a_{ii}| + \sum_{j \neq i} |a_{ij}| = \sum_j |a_{ij}|.$$

This inequality leads to the following upper bound on the **spectral radius** of any square matrix:

$$\rho(\mathbf{A}) = \max_i |\lambda_i| \leq \max_i \sum_j |a_{ij}| = \|\mathbf{A}\|_\infty.$$

Because \mathbf{A} and \mathbf{A}' have the same eigenvalues:

$$\rho(\mathbf{A}) \leq \min \left(\max_i \sum_j |a_{ij}|, \max_j \sum_i |a_{ij}| \right) = \min(\|\mathbf{A}\|_\infty, \|\mathbf{A}\|_1).$$

$$\boxed{\rho(\mathbf{A}) \leq \|\mathbf{A}\|_\infty}$$

$$\rho(\mathbf{A}) = \rho(\mathbf{A}') \leq \|\mathbf{A}'\|_\infty \geq \|\mathbf{A}\|_1$$

Ch. 5 also provided these upper bounds because $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ for any induced norm.

Similarly, again using the **reverse triangle inequality**:

$$|a_{ii}| - |\lambda| \leq ||\lambda| - |a_{ii}|| \leq |\lambda - a_{ii}| \leq r_i \implies |a_{ii}| - r_i \leq |\lambda| \implies \min_i \{ \max(|a_{ii}| - r_i, 0) \} \leq |\lambda|,$$

from which we could obtain a (often loose) lower bound on ρ .

Challenge. find a useful lower bound for general square matrices [15] [16] [17], Section 6.2].

We turn now to matrices with quite simple upper and lower bounds.

7.5 Nonnegative matrices and graphs

We now turn to the special category of matrices whose elements are all real and nonnegative. These matrices have special eigenvalue and eigenvector properties (p. 7.50) that are important in their many applications.

- Such matrices can describe transition probabilities for **Markov chains** that model many random processes.
- Google's **PageRank** algorithm for ranking web sites is built on a certain nonnnegative matrix.
- Other applications include power control, commodity pricing, networks, and population growth [18].

We begin with some definitions.

Define. A matrix is called a **positive matrix** iff all of its elements are real and positive.

Define. A matrix is called a **nonnegative matrix** iff all of its elements are real and nonnegative.

Define. A square **nonnegative matrix** is called a (left) **stochastic matrix** iff each of its columns sum to 1.

Mathematically, P is a **stochastic matrix** iff $P \in [0, \infty)^{N \times N}$ and $\mathbf{1}'_N P = \mathbf{1}'_N$.

Such a matrix typically is used to describe the **transition matrix** of a **Markov chain**.

Define. A matrix A is called a **primitive matrix** iff it is a **square nonnegative matrix** and A^m is a **positive matrix** for some $m \in \mathbb{N}$.

A^2 is positive

9. The matrix $A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$ is (choose most specific answer):

A: Square B: Nonnegative C: Primitive D: Positive E: None of these

??

Positive versus positive definite

The preceding definitions are unrelated to the terms **positive definite** and **positive semidefinite** matrices, except in some special cases like **diagonal** matrices.



10. (Any positive semidefinite diagonal matrix is a nonnegative matrix.) $\neg T$

Any positive definite diagonal matrix is a positive matrix.) (?)

A: (T,T) B: (T,F) C: (F,T) D: (F,F)

??

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \geq 0$$

11. (Every stochastic matrix is a square positive matrix.) $\neg F$

Every square positive matrix is a primitive matrix.) (?)

A: (T,T) B: (T,F) C: (F,T) D: (F,F)

??

 F $m=1$

12. The circulant generator matrix G_N on p. 7.16 for $N > 1$ is (choose most specific answer):

A: Square B: Nonnegative C: Primitive D: Positive E: None of these

??

2021-11-09

Venn diagram (for square matrices only):

Square &
Nonnegative

Primitive **Positive**

Exercise 7-6-2. Add **stochastic** matrices.

(There are rectangular matrices that are nonnegative and/or positive, but those are not of interest here.)

Power test for primitive matrices

(Read)

Fact. A $N \times N$ **nonnegative matrix** A is a **primitive matrix** iff A^{N^2-2N+2} is a **positive matrix** [19].

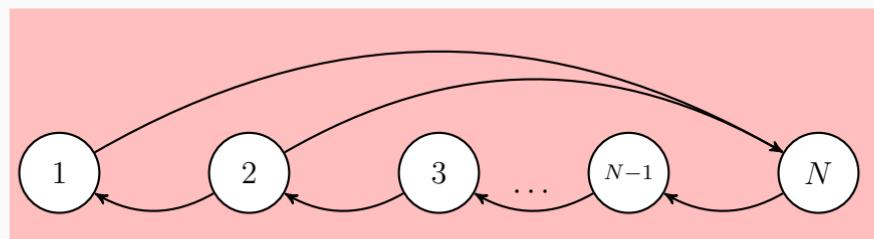
(It suffices to use the binary matrix that has 1's where A is positive and 0's elsewhere.)

Unfortunately, this test of whether a matrix is **primitive** is impractical for large N .

Example. The following code constructs an example matrix A (due to **Helmut Wielandt**) where A^{N^2-2N+1} is not positive yet A^{N^2-2N+2} is. This example is key to the “only if” part of the above fact.

```
N = 8
p = N^2 - 2N + 2
A = diagm(1 => ones(Int, N-1)); A[N, 1:2] .= 1 # Wielandt's example
@assert !all(A^(p-1) .> 0) && all(A^p .> 0)
```

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & \dots \\ & & \ddots & & \\ 0 & \dots & 0 & 0 & 1 \\ 1 & 1 & 0 & \dots & 0 \end{bmatrix}$$



Fact. If \mathbf{A} is **primitive** with \mathbf{A}^m being positive for some $m \in \mathbb{N}$, then \mathbf{A}^n is positive $\forall n \geq m$.

Proof. When \mathbf{A}^m is positive, then, because \mathbf{A} is nonnegative, element i, j of $\mathbf{A}^{m+1} = \mathbf{A}\mathbf{A}^m$ is positive unless the i th row of \mathbf{A} were entirely zero, because $[\mathbf{A}^{m+1}]_{i,j} = \mathbf{A}_{i,:} [\mathbf{A}^m]_{:,j}$. But if the i th row of \mathbf{A} were entirely zero then so would be the i th row of every power of \mathbf{A} , contradicting the assumption that \mathbf{A}^m is positive. Thus \mathbf{A}^{m+1} is positive, and by induction so is every higher power of \mathbf{A} .

Explore 7.0.6: After studying the next section, express the above matrix property in terms of a graph property.

Weighted directed graphs

Square positive, nonnegative, stochastic and primitive matrices have various special properties in terms of their eigenvalues and eigenvectors that are important in applications. We focus on those properties soon, but first we relate these matrices to directed graphs.

We can associate any **square nonnegative matrix** A with a corresponding **weighted directed graph** $\mathcal{G}(A)$ that provides an intuitive visual depiction of the positive elements of A .

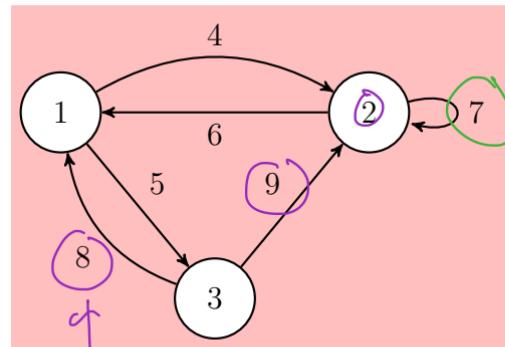
A directed graph has **nodes** that are connected by **arrows** (or **directed edges**) that are labeled by weights. Our convention is that a_{ij} denotes the weight from node j to node i , with arrows *only* for positive weights. A **node** is also called a **vertex** or **state**, and is represented by a circle.

Example. For the square, nonnegative matrix

$$A = \begin{bmatrix} 0 & 6 & 8 \\ 4 & 7 & 9 \\ 5 & 0 & 0 \end{bmatrix}, \text{ the corresponding weighted directed graph } \mathcal{G}(A) \text{ is shown here.}$$

We do not draw an arrow from node j to node i if $a_{ij} = 0$.

We allow non-zero diagonal elements, so technically this is a **weighted directed graph with loops**.



$a_{12} = 4$
 $a_{21} = 6$
 $a_{23} = 9$
 $a_{32} = 7$
 $a_{31} = 5$
 $a_{22} = 7$
 $a_{13} = 8$
 $a_{32} = 8$

from 2 to 3

Explore 7.0.7: Think about what the graph would look like for a **positive matrix**.

Define. The weighted directed graph $\mathcal{G}(\mathbf{A})$ associated with a $N \times N$ square nonnegative matrix \mathbf{A} has nodes numbered $1, \dots, N$, and a set of *distinct* weighted directed edges (arrows) consisting of at most N^2 **tuples** of the form (i, j, a_{ij}) where $i, j \in \{1, \dots, N\}$ and $a_{ij} > 0$.

Example. For the \mathbf{A} shown on the previous page, the set of 6 distinct tuples is $\{(2, 1, 4), (3, 1, 5), (1, 2, 6), (2, 2, 7), (1, 3, 8), (2, 3, 9)\}$

A directed graph looks visually like a “map” and we can imagine traveling from node to node around the diagram by following paths defined by the arrows.

Define. We say we can **reach** or **access** node i from node j in n steps if there is a path along n arrows (including loops) starting at j and ending at i (possibly passing through j or i along the way).

Example. In the previous diagram, we can reach node 3 from node 2 in two steps: $2 \rightarrow 1 \rightarrow 3$.

13. In the preceding example, we can reach node 3 from node 2 in 4 steps in how many distinct ways?
 A: 0 B: 1 C: 2 D: 3 E: 4 ??

??

Reachability or accessibility

To determine mathematically (in matrix form) which nodes in $\mathcal{G}(\mathbf{A})$ we can reach from node j in 1 step, simply compute the matrix-vector product $\mathbf{A}\mathbf{e}_j = \mathbf{A}_{\cdot,j}$ and examine which elements are positive.

Example. $\mathbf{A}\mathbf{e}_2 = \begin{bmatrix} 0 & 6 & 8 \\ 4 & 7 & 9 \\ 5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 6 \\ 7 \\ 0 \end{bmatrix}$, so from node 2 we can reach nodes 1 and 2 in one step.

Fact. To determine which nodes we can reach in $\mathcal{G}(\mathbf{A})$ from node j in m steps, simply compute this product

$$\mathbf{A}^m \mathbf{e}_j \quad (7.7)$$

and then examine which elements are positive.

This is the mathematical version of what we do graphically when we travel along arrows in the diagram. The fact that \mathbf{A} is square and nonnegative is key to establishing this property.

Example. $\mathbf{A}^2 \mathbf{e}_1 = \mathbf{A}(\mathbf{A}\mathbf{e}_1) = \begin{bmatrix} 0 & 6 & 8 \\ 4 & 7 & 9 \\ 5 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 64 \\ 73 \\ 0 \end{bmatrix}$, so from node 1 we can reach nodes 1 or 2 in two steps, but not node 3.

Fact. If A is a **primitive matrix** then $\exists m \in \mathbb{N}$ s.t. A^m is a positive matrix, so in $\mathcal{G}(A)$ one can reach any node i from any node j in m steps.

$$A^m e_j = \begin{bmatrix} >0 \\ >0 \\ \vdots \\ >0 \end{bmatrix}$$

(Read)

Adjacency matrix

How we reach from one node to another node does not depend on the graph weights (other than the fact that $a_{ij} = 0$ is excluded from the graph). Only the presence or absence of an arrow matters.

Thus, often we only need to consider the **adjacency matrix** of a (directed) graph.

For directed graphs, an **adjacency matrix** need not be symmetric.



Example. For the previous example, the **adjacency matrix** is $M = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix}$. Note that $M_{ij} = \mathbb{I}_{\{a_{ij} \neq 0\}}$.

When illustrating an **adjacency matrix** using a (non-weighted) **directed graph**, we need not label the arrows because zero weights are not shown so all arrows correspond to a unity weight, like those illustrated on the next page.

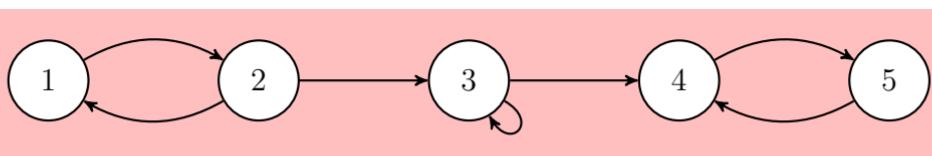
Strongly connected graphs

Now we begin to relate matrix properties and graph properties.

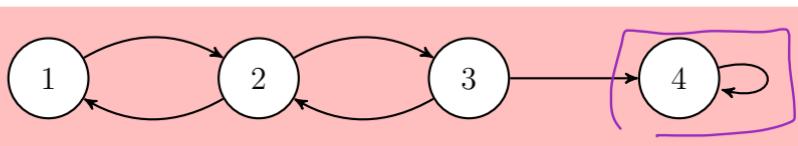
Define. A **directed graph** is **strongly connected** if it is possible to reach any vertex/state/node from any other (in one or more steps).

Example. The directed graphs on p. 7.38 and p. 7.36 are **strongly connected**.

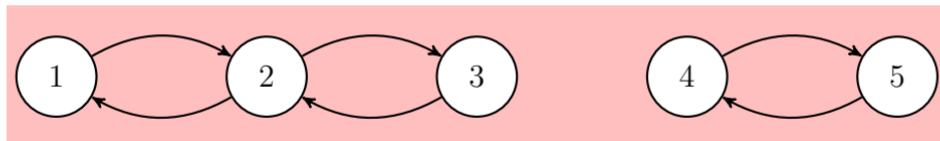
A **strongly connected** graph cannot have any transient or absorbing states, nor can it be partitioned into two or more separate graphs.



Node 3 is **transient**: after you leave it, you never return to it.



Node 4 is **absorbing**: after you enter it, you can never leave.



The adjacency matrix can be permuted to be block diagonal.

$$\begin{bmatrix} \text{I} & & \\ & \ddots & \\ & & 0 \\ \vdots & & \\ & & \text{I} & \text{I} & \end{bmatrix}$$

$$\overset{A^m \text{ is positive}}{\overset{\nearrow}{\text{A}}} \quad \overset{\nearrow}{\text{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}} = G_2 \quad \overset{\nearrow}{\text{A}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}} \quad 7.43$$

$\overset{\nearrow}{\text{A}^3 = \text{A}}$
 $\overset{\nearrow}{\text{A}^4 = \text{A}^2}$
 $\uparrow \text{periodic}$

14.

(For any primitive matrix A , the corresponding directed graph $\mathcal{G}(A)$ is strongly connected. If $\mathcal{G}(A)$ is strongly connected, then the corresponding matrix A is primitive.) (?)

A: (T,T)

B: (T,F)

C: (F,T)

D: (F,F)

??

Strongly connect graphs are especially important, so we want a suitable corresponding matrix type.

Irreducible matrix

Define. A square matrix A is called irreducible iff

$$\forall i, j, \exists m \in \mathbb{N} \text{ such that } [A^m]_{i,j} > 0. \quad (7.8)$$

(In general, m depends on i, j .) Otherwise, the matrix is called reducible.

" m_{ij} "

There are other equivalent definitions of irreducible matrices.

15.

(Any primitive matrix is an irreducible matrix. $\leftarrow T$)

The circulant generator matrix G_N is irreducible.) (?)

A: (T,T)

B: (T,F)

C: (F,T)

D: (F,F)

??

not primitive

Properties of irreducible matrices

Fact. (Taussky's 1948 theorem [20, Thm. 1.11, p. 14]) If $A \in \mathbb{C}^{N \times N}$ is **irreducible** and **diagonally dominant**, then A is **nonsingular**. This property extends the sufficient condition for invertibility provided by Geršgorin's theorem.

Fact. If a $N \times N$ **nonnegative matrix** A is **irreducible**, then $(I + A)^{N-1}$ is a **positive matrix** [21, § 8.2].

Fact. For a square nonnegative matrix A , the corresponding graph $\mathcal{G}(A)$ is strongly connected iff A (or equivalently its **adjacency matrix**) is irreducible.



The proof essentially is due to (7.7).

In words, $[A^m]_{i,j} > 0$ means that we can reach node i from state j in m steps in the directed graph $\mathcal{G}(A)$.

We have seen that any primitive matrix is **irreducible**, but not the other way around in general. To fully describe how these two families of matrices are related, we need one more definition.

Matrix period

(Read)

Define. For a square nonnegative matrix, the period of the i th index is the **greatest common divisor** of all $m \in \mathbb{N}$ such that $[A^m]_{i,i} > 0$, if any such m exists.

GCD

(if set nonempty)

Define. If the period of every index exists, and the periods all equal some $m > 1$, then we call m the **period** of A .

However, we will *not* call such a matrix “periodic” because that has a different meaning. (See below.)



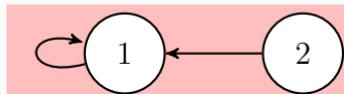
Define. If the period of every index exists and is 1, then A is called **aperiodic**.

Otherwise, we will call it **not aperiodic**. Again, we will not call it periodic!

$$\begin{aligned} [A^m]_{1,1} &= (1, 1, 1, \dots)^T \\ m &= 1, 2, 3, \dots \\ [A^n]_{2,2} &= (0, 0, \dots, 0) \\ \text{GCD} &= 1 \\ m = [A^m]_{2,2} &> 0 \\ \text{is empty} \end{aligned}$$

Example. Consider $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, for which $\mathcal{G}(A)$ is:

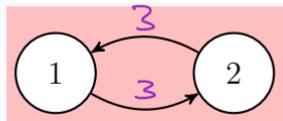
idempotent
We have $A^m = A$, $\forall m \in \mathbb{N}$, so the period of index 1 is simply 1, and index 2 does not have a period.
Thus, this matrix is **not aperiodic**.



Example. Consider $A = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}$, for which $\mathcal{G}(A)$ is:

We have $A^m = \begin{bmatrix} 0 & 3^m \\ 3^m & 0 \end{bmatrix}$ for m odd and $A^m = \begin{bmatrix} 3^m & 0 \\ 0 & 3^m \end{bmatrix}$ for m even. Thus $[A^m]_{i,i} > 0$, for $m = 2, 4, 6, \dots$ for both $i = 1$ and $i = 2$, so the period is 2 for both indexes. We say this matrix has **period 2**.

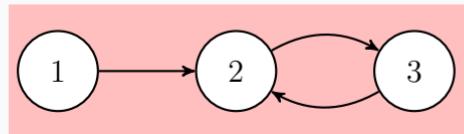
We also say that this matrix is **not aperiodic**.



$$\begin{aligned} [A^m]_{1,1} &= (0, \underset{2}{\cancel{3}}, 0, \underset{4}{\cancel{3^2}}, 0, \underset{6}{\cancel{3^4}}, \dots)^T \\ \text{GCD} &= 2 \end{aligned}$$

$$[\mathbf{A}^m]_{2,2} \approx (0, 3^2, 0, 3^4) \quad \text{--} \quad \text{GCD} = 2$$

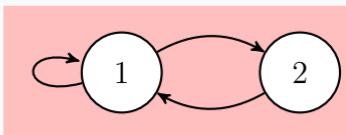
Example. Consider $\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$, for which $\mathcal{G}(\mathbf{A})$ is:



We have $\mathbf{A}^m = \mathbf{A}$ for m odd and $\mathbf{A}^m = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ for m even.

Thus index 1 does not have a period, so this matrix is not aperiodic.

Example. Consider $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$, for which $\mathcal{G}(\mathbf{A})$ is:



We have $\mathbf{A}^2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ and \mathbf{A}^m is positive for all $m > 1$. So the period of index 1 is $\text{gcd}(1, 2, 3, \dots) = 1$ and the period of index 2 is $\text{gcd}(2, 3, \dots) = 1$. This matrix is aperiodic.

16. Any square **positive** matrix is **aperiodic**. (?)

A: True

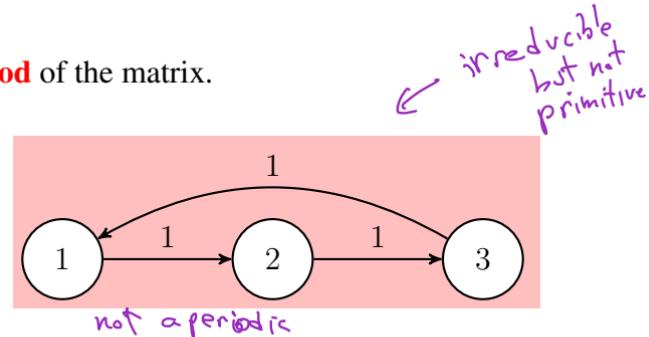
B: False

??

Fact. For a nonnegative irreducible matrix, the period of the i th index always exists and is the same for all i .

As noted before, if the value exceeds 1, then it is called the **period** of the matrix.

Example. The adjacency matrix corresponding to a loop of 3 states is $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, for which $P^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$, $P^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $P^4 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$, and $\mathcal{G}(P)$ is shown here. This matrix is nonnegative and irreducible and it has period 3.



Fact. A square nonnegative matrix is **primitive** iff it is **irreducible** and **aperiodic**.

If A is primitive, then there is $K \in \mathbb{N}$ such that A^K is **positive** and hence A is **irreducible**. Furthermore, A^m is **positive** for all $m \geq K$ so the GCD must be 1 and hence A is **aperiodic**.

The converse is more subtle. See:

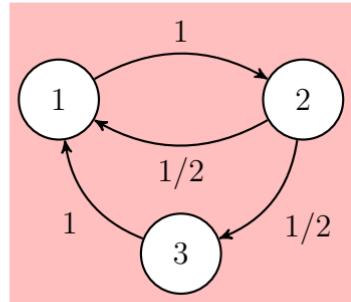
[\(sufficient\)](https://math.stackexchange.com/questions/172313/why-is-every-irreducible-matrix-with-period-1-primitive?rq=1)

Fact. If an nonnegative **irreducible** matrix A has at least one non-zero diagonal element, then A is **primitive** [21, p. 678].

Because then from any state i we can return to that state in some m_i steps and also in $m_i + 1$ steps, so the i th period is 1 for all i , and hence the matrix is **aperiodic**.

Example. However, having a non-zero diagonal element is not a necessary condition for a matrix P to be **primitive**. Consider the (transition) matrix $P = \begin{bmatrix} 0 & 1/2 & 1 \\ 1 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix}$ with $\mathcal{G}(P)$ as shown.

Here P^5 is positive, so P is **irreducible** and **aperiodic**, and hence **primitive**, even though P has all zero diagonals.



Some texts define a square matrix A to be **periodic** iff $A^{k+1} = A$ for some $k \in \mathbb{N}$, e.g.,
<http://mathworld.wolfram.com/PeriodicMatrix.html>

If such a k exists and is the least such k , then k is called “the period” of the matrix.

That definition is not equivalent to our definition, even for nonnegative matrices.

Example. Consider $B = \begin{bmatrix} 0 & 2 \\ 2 & 0 \end{bmatrix}$ that has period 2 but all powers of B differ.

Challenge. prove or disprove whether the two definitions are equivalent for stochastic matrices.



Summary of relationships between square nonnegative matrix types and corresponding graph types –

Matrix type	Graph type	notes
nonnegative	weighted directed graph	weights are positive matrix elements
adjacency	directed graph	no weights, just arrows (allow loops)
transition / stochastic	Markov chain	columns sum to 1
irreducible	strongly connected	can reach any node from any node
primitive	(?!)	$\exists m \in \mathbb{N}$ s.t. can reach any node from any other node in m steps
positive	fully connected	1 step between any nodes

7.6 Nonnegative matrices and Perron-Frobenius theorems

Now we examine eigen properties, starting with the broadest category: square **nonnegative matrices**.

PF for square nonnegative matrices

Fact. If \mathbf{A} is **square** and **nonnegative**, then we have the following bounds on its **spectral radius** [18]:

- Row sums provide bounds on the spectral radius: $\min_i \sum_j a_{ij} \leq \rho(\mathbf{A}) \leq \max_i \sum_j a_{ij} = \|\mathbf{A}\|_\infty$
- Column sums provide bounds on the spectral radius: $\min_j \sum_i a_{ij} \leq \rho(\mathbf{A}) \leq \max_j \sum_i a_{ij} = \|\mathbf{A}\|_1$

Combining yields the following bounds on the **spectral radius** of a **nonnegative matrix**:

$$\max\left(\min_j \sum_i a_{ij}, \min_i \sum_j a_{ij}\right) \leq \rho(\mathbf{A}) \leq \min\left(\max_j \sum_i a_{ij}, \max_i \sum_j a_{ij}\right). \quad (7.9)$$

- The upper bounds follow from the fact that $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$ for any **induced** matrix norm.
- Challenge: [18] claims the bounds are easy to show. The proof of the lower bound given in [22, p. 52], citing [23], uses nonnegative Perron vector, whereas [24, p.37] uses a messy determinant.

Example. For a permutation matrix, the row and column sums are all 1 so we know $\underline{\rho} = 1$.

If \mathbf{P} is a **stochastic matrix**, i.e., $\underline{1'P = 1}$, then $\rho(\mathbf{P}) = 1$.

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



Proof. Clearly one eigenvalue of P is 1, so we know $1 \leq \rho(P)$.

Yet, the upper bound in (7.9) is also 1, because the column sums are all 1. \square

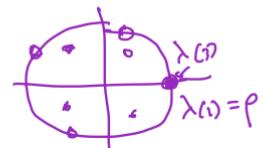
The **Perron-Frobenius theorem** for any square, nonnegative matrix A states the following:

- $r \triangleq \rho(A) \geq 0$ is an eigenvalue of A .

In other words, the largest magnitude eigenvalue is a real, nonnegative eigenvalue.

Think $\lambda_{(1)} = r \geq 0$, where $\lambda_{(1)}$ denotes the first eigenvalue when sorted by magnitude.

We call r the **Perron root** or **Perron-Frobenius eigenvalue**.



- A has a (right) eigenvector v with nonnegative elements corresponding to that eigenvalue, i.e., $Av = rv$.
- There likewise exists a left eigenvector w having all nonnegative elements for which $w^T A = rw^T$.
We refer to such v and w as right and left **Perron vectors**, respectively.

Remarks.

- Because A and A' have the same eigenvalues, the existence of the nonnegative left eigenvector follows directly from the existence of the nonnegative right eigenvector, because A' is also a nonnegative matrix.
- The fact that the largest magnitude eigenvalue is real and nonnegative takes some work to show, and does not follow from the **Geršgorin disk theorem**. One version uses the **proof for positive matrices** based on **Gelfand's formula** (5.23), followed by noting that any nonnegative matrix is a **limit of positive matrices**.
- By definition of spectra radius: $|\lambda_i(A)| \leq \rho(A)$, so for a square nonnegative matrix $|\lambda_i(A)| \leq r = \lambda_{(1)}$.

- This inequality is not strict and in general there can be multiple eigenvalues with the same magnitude. In fact there can be multiple eigenvalues all equal to ρ , e.g., for $\underline{A = I}$. This matters because our convergence theory for the **power method** assumed there was one eigenvalue with dominant magnitude. So we cannot say much about powers of nonnegative matrices.
- Likewise, there can be multiple non-co-linear right nonnegative eigenvectors having an eigenvalue where $|\lambda_i| = r$. In other words, we cannot say anything about uniqueness (or **multiplicity**) of an eigenvector having nonnegative elements. This means we cannot say much about the limiting behavior of Markov chains having merely nonnegative transition matrices.

Example. For the nonnegative matrix $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ the eigenvalues are 1, 1, so $\rho = 1$. Any nonzero vector of the form $\begin{bmatrix} x \\ y \end{bmatrix}$ with $x, y \geq 0$ is a (right and left) Perron vector.



Example. For $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ the eigenvalues are $+1, -1$, so $\rho = 1$. The vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ is a (right and left) Perron vector; the vector $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ is also an eigenvector whose eigenvalue, -1 , also has magnitude 1. Powers A^k do not converge. In this case, $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ is the unique nonnegative unit-norm eigenvector corresponding to the eigenvalue $\lambda = 1 = \rho$.



PF for nonnegative irreducible matrices

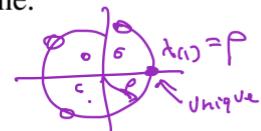
The **Perron-Frobenius theorem** for any (square) nonnegative irreducible matrix A states that:

- All the conclusions for (square) nonnegative matrices hold.
- $r \triangleq \rho(A) > 0$ is a (real) eigenvalue of A .

In other words, the largest magnitude eigenvalue is a real, *positive* eigenvalue. Think $\lambda_{(1)} = r > 0$.

We call r the **Perron root** or **Perron-Frobenius eigenvalue**.

- That eigenvalue is simple (aka unique), i.e., has multiplicity one, meaning both its algebraic multiplicity is one (non-repeated root of the characteristic polynomial) and geometric multiplicity is one.
- If A has period m , then there are exactly m eigenvalues for which $|\lambda| = r = \rho(A)$.



- There exists a unit-norm eigenvector v with eigenvalue r , i.e., $Av = rv$, where v has all positive elements.
- There likewise exists a unit-norm left eigenvector w having all positive elements for which $w'A = rw'$. We call this v and w the right and left **Perron vectors**, respectively
- Uniqueness: All other (left and right) unit-norm eigenvectors have negative and/or non-real elements. So v and w are unique in that sense.

Recall that for a nonnegative irreducible matrix, the period of the i th index exists and is the same for all i .

- If that period is $m > 1$, then the matrix has period m and there are m eigenvalues where $|\lambda_i| = r$.
- If that period is $m = 1$, then the matrix is **aperiodic** and hence **primitive**, the case discussed next.

PF for primitive matrices (includes positive matrices)

To make stronger statements about uniqueness, we make stronger assumptions.

The **Perron-Frobenius theorem** for any (square) primitive matrix A states that:

- All the conclusions for (square) nonnegative irreducible matrices hold.
- $r \triangleq \rho(A) > 0$ is a (real) eigenvalue of A .

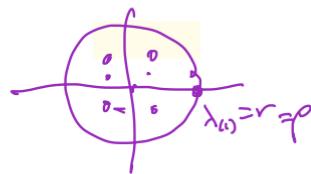
In other words, the largest magnitude eigenvalue is a real, *positive* eigenvalue. Think $\lambda_{(1)} = r > 0$.

We call r the **Perron root** or **Perron-Frobenius eigenvalue**.

- That eigenvalue has **multiplicity one**, both **algebraic multiplicity** and **geometric multiplicity**.
- Uniqueness of *magnitude*: For any other eigenvalue λ , $|\lambda| < r = \rho(A)$, i.e., there is no other λ such that $|\lambda| = r$.
- There exists a unit-norm eigenvector v with eigenvalue r , i.e., $Av = rv$, where v has all *positive* elements.
- There likewise exists a unit-norm left eigenvector w having all positive elements for which $w^T A = rw^T$. We call this v and w the right and left **Perron vectors**, respectively
- Uniqueness: All other (left and right) unit-norm eigenvectors have negative and/or non-real elements. So v and w are unique in that sense. (proof)

Because a primitive matrix A has one eigenvalue whose magnitude dominates all others, the **power iteration converges** to the **Perron vector** (with no sign issues if initialized with a positive vector).

All of these properties hold for any (square) **positive matrix** because any such matrix is also **primitive**.



PF for stochastic matrices

(Read)

We will apply all the above definitions and properties to study **Markov chains**. The **transition matrix P** for a **Markov chain** is a **stochastic matrix**, meaning that its columns sum to one. In other words $\underline{1'P = 1}$.

As explained on p. 7.34, a **stochastic matrix** has $\rho(P) = 1$. All of the preceding Perron-Frobenius theory (for nonnegative, irreducible and primitive matrices) applies readily to any **stochastic matrix** simply by taking the special case that $\rho(P) = 1$.

$$G_4$$

$$e^{i \frac{\pi}{4} k}$$

17.

(There exists a 4×4 stochastic matrix having the four eigenvalues $\{\pm 1, \pm i\}$.

There exists a 4×4 stochastic matrix for which every nonnegative vector in \mathbb{R}^4 is an eigenvector.)

(?)

A: (T,T)

B: (T,F)

C: (F,T)

D: (F,F)

??

 $T - I_4$

Explore 7.0.8: Give an example of an **aperiodic** stochastic matrix whose graph has a **transient** state.

Next we apply these definitions and properties to study Markov chains.

2021-11-11

7.7 Markov chains

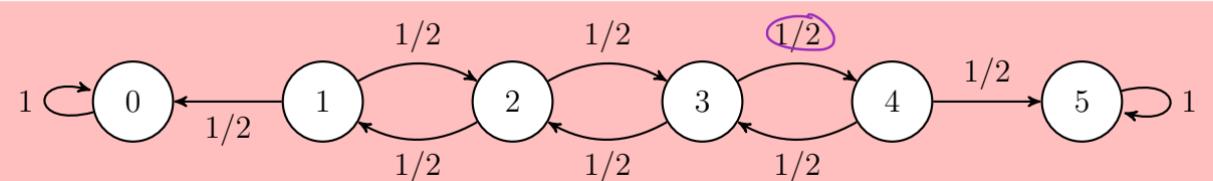
Define. A (first-order) **Markov chain** is a random process (or random sequence, *i.e.*, a sequence of random variables) X_1, X_2, \dots whose joint distribution is such that the conditional distribution of X_{k+1} given the values (states) of all preceding variables depends only on the preceding variable's value (state) and not how the process reached that state:

$$\underbrace{p(X_{k+1} = x | X_k = x_k, X_{k-1} = x_{k-1}, \dots, X_1 = x_1)}_{\text{next}} \underbrace{=}_{\text{curr}} \underbrace{p(X_{k+1} = x | X_k = x_k)}_{\text{prev}}$$

Example. You start with \$2 and bet \$1 on the outcome of a fair coin toss. If you win you then have \$3 and if you lose you then have \$1. And you continue until (w.p.1) you eventually lose all your money or you stop betting because you reach your goal of, say, \$5. If at time k point you have, say \$3, then

$$p(X_{k+1} = x | X_k = 3, X_{k-1} = x_{k-1}, \dots, X_1 = x_1) = p(X_{k+1} = x | X_k = 3) = \begin{cases} 1/2, & x = 4 \\ 1/2, & x = 2 \\ 0, & \text{otherwise.} \end{cases}$$

In words, the next state will be \$2 or \$4, (with probability 1/2), regardless of how you reached the state of \$3.



We focus on cases (like the preceding example) where the random variables take a finite set of discrete values. In this case, WLOG we can use the values $\{1, \dots, N\}$, where N is the number of **states** (i.e., values).

In this typical case, the statistical properties are governed by the **transition probabilities**:

$$p(X_{k+1} = i \mid X_k = j), \quad \underset{\text{next}}{i}, \underset{\text{current}}{j} = 1, \dots, N.$$

In the typical case where the distributions are independent of “time” index k , we call it a **time-homogeneous Markov chain**, and we often describe the Markov chain by a **weighted directed graph** that illustrates the transition probabilities between states. (See examples on previous and on next page.)

- Circles represent states.
- Arrows are labeled with transition probabilities.
- We draw arrows in such graphs only for nonzero transition probabilities.

Here is a demo showing many realizations of sequences $\{X_k\}$; it also illustrates the limiting distribution discussed later:

https://web.eecs.umich.edu/~fessler/course/551/julia/demo/07_markov_chain1.html

https://web.eecs.umich.edu/~fessler/course/551/julia/demo/07_markov_chain1.ipynb

18.

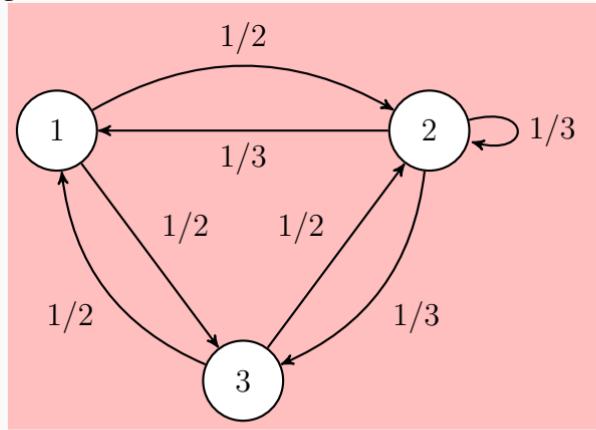
The transition matrix P in the demo is (choose most specific correct answer):

- A: Square B: Nonnegative C: Irreducible D: Primitive E: Positive

??

??

Example.



Because our focus is matrix methods, not graphs or random processes, naturally we use a $N \times N$ **matrix** to describe these probabilities, called the **transition matrix**:

$$\underline{\boldsymbol{P}_{i,j} \triangleq p(X_{k+1} = i | X_k = j), \quad i, j = 1, \dots, N.}$$

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1/3 & 1/2 \\ 1/2 & 1/3 & 1/2 \\ 1/2 & 1/3 & 0 \end{bmatrix}$$

$$= \left[\begin{array}{c} p(X_{k+1} = i | X_k = j) \end{array} \right]$$

In some texts the transpose of \boldsymbol{P} is also called the **transition matrix**.

Note that the sum of the elements of each column is one, *i.e.*, $\mathbf{1}'_3 \boldsymbol{P} = \mathbf{1}'_3$. So \boldsymbol{P} is a **stochastic matrix**.



Properties of transition matrices

Any $N \times N$ transition matrix \mathbf{P} has useful properties.

- All transition matrices are **square** and **nonnegative**.
- Generalizing the previous example, the columns always sum to unity:

$$\boxed{1'_N \mathbf{P} = 1'_N} \text{ because } \boxed{\sum_{n=1}^N p(X_{k+1} = n | X_k = j) = 1.}$$

In words, the transition probabilities for leaving the j th state sum to one, for every j .

- Due to (7.10), a $N \times N$ transition matrix \mathbf{P} always has $\mathbf{1}_N$ as a left eigenvector with eigenvalue 1.
- Because all columns sum to unity: $\rho(\mathbf{P}) \leq 1$ so $|\lambda_i(\mathbf{P})| \leq 1$. (No eigenvalues are outside the unit disk.)
- Thus $\rho(\mathbf{P}) = 1$.

In general, there can be multiple nonnegative eigenvectors having eigenvalue equal to unity.

Example. The following transition matrix has two distinct unit-sum, nonnegative eigenvectors with eigenvalue 1:

$$e^{i\pi}(\mathbf{P}) = (1, 1, -1) \quad \mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = P^3 \quad \mathbf{P}\mathbf{v}_1 = \mathbf{v}_1 \quad \mathbf{P}\mathbf{v}_2 = \mathbf{v}_2$$



19.

(This \mathbf{P} is **irreducible**. This \mathbf{P} is **primitive**) (?)

A: (T,T)

B: (T,F)

C: (F,T)

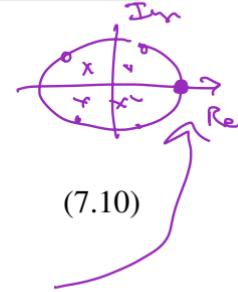
D: (F,F)



③ ↗ 1

not strongly connected

$$\mathbf{P}^2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad ??$$



(7.10)

Equilibrium distribution(s) of a Markov chain

Markov chains are a rich subject with numerous applications and theoretical considerations.

One particular aspect of them that we can examine using matrix methods is the existence of an **equilibrium distribution** or **stationary distribution** or **steady-state distribution**.

Define. Let π denote an N -dimensional vector $\pi = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_N \end{bmatrix}$. We call π a **stationary distribution** or **equilibrium distribution** or **steady-state distribution** of a Markov chain having **transition matrix P** iff

$$\underbrace{\pi_n \geq 0, \quad n = 1, \dots, N,}_{\text{nonnegative elements}} \quad \underbrace{\sum_{n=1}^N \pi_n = \mathbf{1}'_N \pi = 1,}_{\text{sum to unity}} \quad \underbrace{P\pi = \pi.}_{\text{eigenvector}}$$

In words, π has nonnegative elements that sum to unity and is an eigenvector of P with eigenvalue 1. Such a π is sometimes called a **stochastic eigenvector**. See preceding example for two such π .

- Because a transition matrix P is nonnegative and has $\rho(P) = 1$, the **Perron-Frobenius theorem for nonnegative matrices** ensures that there always **exists** such a (nonzero) eigenvector having nonnegative elements with eigenvalue 1, so we can normalize that eigenvector to sum to unity, establishing **existence** of an **equilibrium distribution**. || π ||₁ = 1 ✓
- However, that eigenvalue of 1 need not be unique and there can be multiple stochastic eigenvectors in general. See preceding example.

(strongly connected graph)

 If P is **irreducible**, then the equilibrium distribution is unique. (Follows from P-F theorem.)

(Read)

The importance of an **equilibrium distribution** is the following.

In general, by the **law of total probability**:

$$\boxed{p(X_{k+1} = i) = \sum_{j=1}^N p(X_{k+1} = i | X_k = j) p(X_k = j).}$$

The chain is in **equilibrium** or **steady state** when $p(X_{k+1} = n) = p(X_k = n) = \pi_n$, i.e., when:

$$\pi_i = \sum_{j=1}^N p(X_{k+1} = i | X_k = j) \pi_j = \sum_{j=1}^N P_{ij} \pi_j.$$

In matrix vector form:

$$\boldsymbol{\pi} = \mathbf{P}\boldsymbol{\pi}.$$

In other words, an equilibrium distribution $\boldsymbol{\pi}$ is a stochastic eigenvector of the transition matrix \mathbf{P} associated with eigenvalue 1.

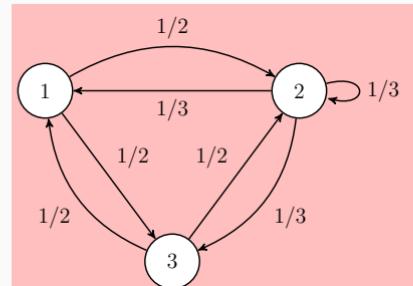
Caution: “**steady state**” does not mean “stuck in one state” in general, although one of the two equilibrium distributions for the preceding 3 state example happens to have that property.



Example. Returning to an earlier example, the eigenvalues of P =

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

↑



This is a **primitive** matrix.

A (unit norm) eigenvector associated with eigenvalue 1 is

$$\mathbf{v}_1 = \begin{bmatrix} 0.485071 \\ 0.727607 \\ 0.485071 \end{bmatrix}.$$

One can verify that $\|\mathbf{v}_1\|_2 = 1$. But this is not the normalization we want for a probability distribution. We want $\sum_n \pi_n = 1$ i.e., $\mathbf{1}' \mathbf{\pi} = 1$.

So define $\mathbf{\pi} = \mathbf{v}_1 / (\mathbf{1}' \mathbf{v}_1)$, which will also take care of any sign flip if needed, for which $\mathbf{\pi} = \begin{bmatrix} 0.285714 \\ 0.428571 \\ 0.285714 \end{bmatrix}$.

unique

This **stochastic eigenvector** is the **equilibrium distribution** and its elements show the (equilibrium) probability of each state. Note that state 2 is more likely, and this is intuitive in light of the graph.

20. Which state in the demo has the lowest probability in equilibrium?

A: 1

B: 2

C: 3

D: 4

E: None: equally likely

??

Limiting distribution(s) of a Markov chain

We can also use matrix methods to examine the **limiting behavior** or **asymptotics** of Markov chains.

If we start the chain in some state (perhaps chosen at random, or perhaps not) and run the system for a long time, what is the probability of being in the n th state? Formally we want to examine:

$$\lim_{k \rightarrow \infty} p(X_k = n), \text{ for } n = 1, \dots, N.$$

Fact. If the **transition matrix P** is **primitive**, then it follows from the Perron-Frobenius theorem for primitive matrices [18] that (see proof sketch on later page):

- the **equilibrium distribution π** is **unique** (as stated previously), (because P is irreducible)
- the limit of powers of P is a special rank-1 matrix:

$$\lim_{k \rightarrow \infty} P^k = \pi \mathbf{1}_N' = [\pi \ \pi \ \dots \ \pi] \quad (7.12)$$

rank (outer product)

Thus, when P is **primitive**, then regardless of the initial distribution,

$$\lim_{k \rightarrow \infty} p(X_k = n) = \pi_n,$$

i.e., the **limiting distribution** of a Markov chain having a primitive transition matrix is the **unique equilibrium distribution**.

$$\pi_0 \geq 0 \quad \sum_i \pi_0 = 1 \quad P^k \pi_0 \rightarrow \pi \quad \text{as } k \rightarrow \infty \quad \|P^k \pi_0 - \pi\| \rightarrow 0$$

if π_0 including e_i

Example. For our running example on p. 7.62 with 3 states, the transition matrix *is* primitive.

```
julia> P = [0 1/3 1/2; 1/2 1/3 1/2; 1/2 1/3 0]
0.0 0.333333 0.5
0.5 0.333333 0.5
0.5 0.333333 0.0
```

```
julia> P^5
0.270062 0.285751 0.301312
0.428627 0.428498 0.428627
0.301312 0.285751 0.270062
```

$$\approx [\pi \quad \pi \quad \pi] = \pi \mathbf{1}'_3$$

$$= \pi \mathbf{1}'_N = P_\pi$$

21. If transition matrix P is a **primitive matrix**, then $\lim_{k \rightarrow \infty} P^k$ is a **projection matrix**. (?)
- A: True B: False

??

$$P_\pi P_\pi = (\pi \mathbf{1}'_N)(\pi \mathbf{1}'_N) = \pi \underbrace{(\mathbf{1}'_N \pi)}_I \mathbf{1}'_N = P_\pi$$

(Read)

Proof of (7.12) in the case where \mathbf{P} is **diagonalizable**, with magnitude-sorted eigenvalues:

$$\mathbf{P} = \mathbf{V}\Lambda\mathbf{V}^{-1} \implies \mathbf{P}^k = \mathbf{V}\Lambda^k\mathbf{V}^{-1} \rightarrow \mathbf{P}_* \triangleq \mathbf{V} \operatorname{Diag}\{1, 0, \dots, 0\} \mathbf{V}^{-1} = \mathbf{V}\mathbf{e}_1\mathbf{e}'_1\mathbf{V}^{-1} = \mathbf{v}_1\mathbf{e}'_1\mathbf{V}^{-1}$$

as $k \rightarrow \infty$, because the Perron Frobenius theorem says that there is only one eigenvalue $\lambda_{(1)} = 1$ and all others have $|\lambda_i| < 1$ so $\lambda_i^k \rightarrow 0$ as $k \rightarrow \infty$. So we see that \mathbf{P}_* is a rank 1 outer product.

Now we know that $\mathbf{1}'\mathbf{P} = \mathbf{1}'$ and it follows from the law of total probability that $\mathbf{1}'\mathbf{P}^k = \mathbf{1}'$ for all $k \in \mathbb{N}$. Thus in the limit we also have $\mathbf{1}'\mathbf{P}_* = \mathbf{1}'$ so:

$$\mathbf{1}' = \mathbf{1}'\mathbf{P}_* = \mathbf{1}'\mathbf{v}_1\mathbf{e}'_1\mathbf{V}^{-1} = (\mathbf{1}'\mathbf{v}_1)\mathbf{e}'_1\mathbf{V}^{-1} \implies \mathbf{e}'_1\mathbf{V}^{-1} = \mathbf{1}'/(\mathbf{1}'\mathbf{v}_1) \implies \mathbf{P}_* = \frac{\mathbf{v}_1}{\mathbf{1}'\mathbf{v}_1}\mathbf{1}' = \boldsymbol{\pi}\mathbf{1}'. \quad \square$$

The proof for non-diagonalizable matrices uses the Jordan form in a similar way.

Having \mathbf{P} be **primitive** is sufficient, but not necessary, for $\{\mathbf{P}^k\}$ to converge.

Example. Consider $\mathbf{P} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$ for which $\mathbf{P}^k = \mathbf{P}$ so $\{\mathbf{P}^k\}$ converges (immediately) to

$\begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \boldsymbol{\pi}_* \mathbf{1}'_2$ and $\{\mathbf{P}^k \boldsymbol{\pi}\}$ converges (immediately) to $\boldsymbol{\pi}_* = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ for any initial distribution $\boldsymbol{\pi}$.

Markov chains with strongly connected graphs

(P irreducible)

Example. The transition matrix corresponding to a loop of 3 states is $P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$,

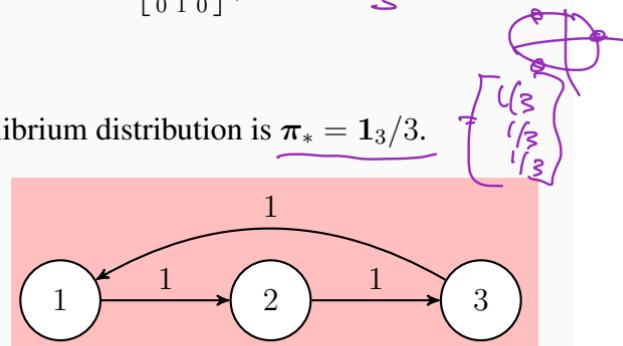
$$= G_3$$

for which $P^2 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$, $P^3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $P^4 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$,

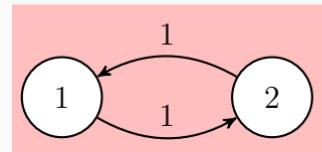
This matrix is nonnegative and irreducible and it has period 3.

It has three eigenvalues whose magnitudes are 1. Its (only) equilibrium distribution is $\pi_* = \underline{1_3/3}$.

However, the matrix power sequence $\{P^m\}$ does not converge.



Example. This two-state Markov chain is **strongly connected**.



22. The corresponding transition matrix $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ is primitive. (?)

A: True

B: False

??

$$\overset{?}{P} = I \quad P^3 = P \quad \dots$$

- The corresponding eigenvalues are $\{+1, -1\}$.
- This 2×2 matrix is irreducible, and its unit-norm Perron vector is $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$.
- So the (unique) equilibrium distribution of this chain is $\pi_* = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$.

23. Let $\pi^{(n)} \triangleq P^{n-1}\pi^{(1)}$ where $\pi^{(1)} \triangleq \begin{bmatrix} P\{X_1 = 1\} \\ P\{X_1 = 2\} \end{bmatrix}$ denotes an initial probability distribution.

Does $\pi^{(n)}$ approach the equilibrium distribution π_* as $n \rightarrow \infty$?

A: Yes, for any $\pi^{(1)} \geq 0$ s.t. $1_2' \pi^{(1)} = 1$.

B: Yes, but only for certain $\pi^{(1)}$ choices ~~X~~.

C: No, never, because P^n oscillates between $\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

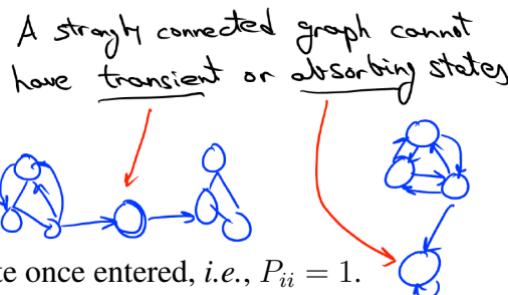
D: None of these.

??

$$\pi^{(1)} = \pi_* = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \quad \pi^{(1)} = P \quad \pi^{(1)} = \pi^{(1)} = \pi_*$$

(Read)

A Markov chain with a strongly connected graph cannot have **transient** or **absorbing** states:



- Absorbing state: impossible to leave state once entered, i.e., $P_{ii} = 1$.
- Transient state: non-zero probability of never returning.

Explore 7.0.9: Express existence of a transient state in terms of transition matrix P .

If a Markov chain has a **strongly connected graph**, then the following properties hold.

- Its transition matrix is **irreducible**.
- An **equilibrium distribution** exists and is *unique* and has all *positive* entries.
- The associated eigenvalue is 1 and is *unique*.

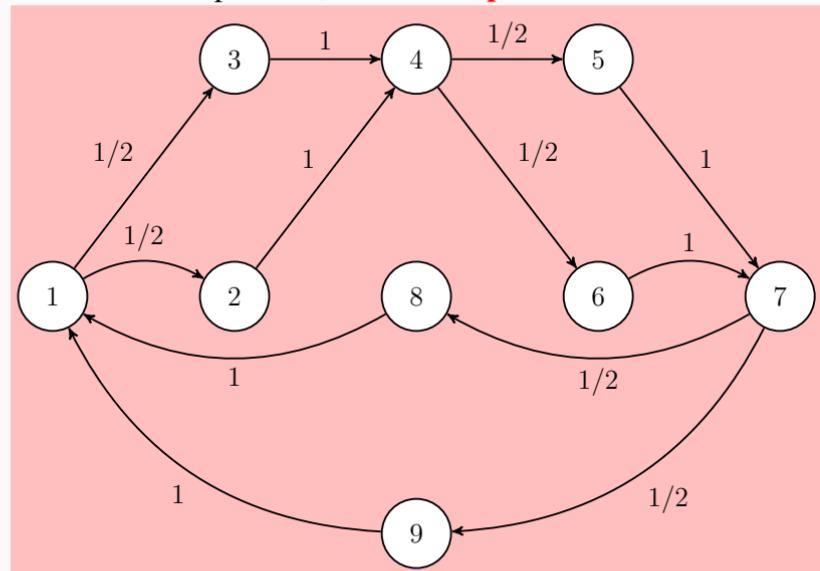
(However, there can be other eigenvalues whose *magnitude* is 1.)

We *cannot* conclude that the Markov chain approaches in the limit that equilibrium distribution.

Making the stronger assumption that the transition matrix is **primitive** ensures that limiting behavior.

(Read)

Example. This Markov chain has a **strongly connected graph** and thus has an **irreducible** transition matrix. From any state it takes exactly 6 steps to return to that state, no matter which path one takes. In other words $[P^k]_{ii} = 1 > 0$ when $k = 6, 12, 18, \dots$ but is zero otherwise. Thus the transition matrix has period = 6, and hence P is not aperiodic, so P is not **primitive**.



Google's PageRank method

This section uses the preceding ideas to summarize **Google's PageRank** method [18]. [US Patent 6,285,999]

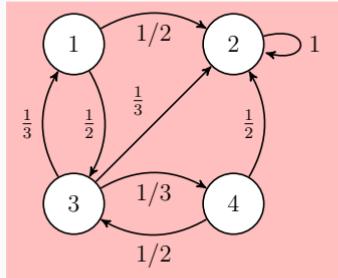
To rank web pages presented after a search, two natural criteria are

- The relevance of each page (number of related terms) etc.
- The “importance” (link popularity) of each web page.

One way to quantify importance is to imagine making a **random walk** between pages, choosing the next page *at random* from all the out links on the current page. This is a Markov chain having a transition matrix:

$$p_{ij} = \begin{cases} 1, & i = j \text{ and page } j \text{ has no out links} \\ 1/N_j, & i \neq j \text{ and page } j \text{ has } N_j \text{ out links, one of which is page } i \\ 0, & \text{otherwise.} \end{cases}$$

Example.



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

The idea is that someone who hypothetically walked pages at random this way would tend to visit more important pages more often because important pages have many incoming links.

So to rank pages by importance we would like to find “the” equilibrium distribution of this Markov chain and then rank importance based on sorting the probabilities from largest to smallest.

In practice, the transition matrix \mathbf{P} has columns that sum to one by construction, but the graph of web page outlinks is *not strongly connected*, so \mathbf{P} is not **irreducible**.

- Web has **absorbing** states (“dangling nodes”), e.g., PDF documents with no links and many business sites
- Web is not strongly connected (e.g., probably cannot reach canvas.umich.edu from whitehouse.gov)

Thus there is not a unique equilibrium distribution.

One solution is to instead use a modified transition matrix for some $0 < \alpha < 1$:

$$\tilde{\mathbf{P}}_\alpha \triangleq \alpha \mathbf{P} + (1 - \alpha) \underbrace{\frac{1}{N} \mathbf{1}_N \mathbf{1}'_N}_{\substack{85\% \\ 15\%}} = \alpha \mathbf{P} + (1 - \alpha) \underbrace{\begin{bmatrix} 1/N & \dots & 1/N \\ \vdots & & \vdots \\ 1/N & \dots & 1/N \end{bmatrix}}_{\text{jump at random to any of the } N \text{ known web pages}}.$$

This matrix $\tilde{\mathbf{P}}_\alpha$ is clearly **positive** by construction, so it has a unique equilibrium distribution π having all positive elements. This vector π has been called the \$25 billion eigenvector [25].

$$x_{n+1} = \frac{Ax_n}{\|Ax_n\|_2}$$

7.72
usual
power iteration
 $\|x_n\|_2 = 1$

PageRank algorithm

- Generate adjacency matrix by web crawling. It is (extremely) sparse, so store accordingly
- Form the transition matrix P . Again, this is (extremely) sparse.
- Select value for **damping factor** α . The original paper by Sergei Brin and (UM alum) Larry Page used $\alpha = 0.85$, apparently based on how often users return to their bookmarks.
- Pick an initial distribution vector π_0 at random from unit simplex (nonnegative entries and sums to 1).
- Perform a modified version of the power iteration where we keep the sum equal to one instead of the Euclidean norm equal to one:

$$\pi_{k+1} = \tilde{P}_\alpha \pi_k = \alpha P \pi_k + \frac{1-\alpha}{N} \mathbf{1}_N.$$

- Perform 40-50 iterations (according to original paper)
- Use the final π vector (sorting?) to quantify the “importance” of web pages.

Example. Applying this method to the preceding 4-state example with $\alpha = 0.85$ yields $\pi = \begin{bmatrix} 0.0634 \\ 0.7818 \\ 0.0914 \\ 0.0634 \end{bmatrix}$.

State 2 is the most important and state 3 is the next, and states 1 and 4 are tied for last. In this particular example, we could have ranked states based on the number of incoming links. However, Brin and Page argue in their **patent** that link counting is suboptimal in general; it is more informative if your web page is linked by other important web pages, not just by many (possibly spam) web pages.

PageRank computation

(Read)

A key step in the algorithm is multiplying the $N \times N$ sparse matrix \mathbf{P} by a (non-sparse) vector: $\mathbf{P}\pi_k$. Using ordinary matrix multiplication this would require N^2 operations.

In 2016, N was over 130 trillion says [this article](#). Google's page says [hundreds of billions](#).

Fortunately, \mathbf{P} is extremely sparse; most web pages link to just a few other pages, and the average number of outgoing links is probably roughly a constant that does not really scale with N . Using the fact that \mathbf{P} is sparse, with an average of, say, $L \ll N$ links per page, the product $\mathbf{P}\pi_k$ requires "just" $O(NL)$ operations. This is still enormous, requiring massive computation and considerable energy.

If N is about 100 trillion, then $1/N$ is about 10^{-14} which is smaller than machine precision (about 10^{-7}) of 32-bit floating point numbers, so apparently π_k must be stored as a 64-bit (double precision) vector. The number of bytes just to stored π is then $8N$ which is over 700 petabytes. Surely Google must have some tricks to deal with this.

PageRank is independent of a user's query so it can be done relatively infrequently (hourly? daily? weekly?) instead of for every individual search.

Numerous practical issues arise to deal with web sites trying to boost scores artificially.

PageRank summary

One of the most influential algorithms of modern times has its roots in Markov chains, directed graphs, positive matrices, and eigenvector computation using the power method.

For more reading, see the [article](#) by Mathworks founder **Cleve Moler** and the [related MATLAB demo](#).

7.8 Summary

This chapter has introduced just a few of the many important special matrices used in practice, culminating with a summary of the PageRank method.

Another application of the Perron-Frobenius theorem is graph matching [26].

The following table summarizes the relationship between the properties of transition matrix of a Markov chain and its corresponding graph.

Transition matrix	Markov Chain graph	Key property
Nonnegative	Directed	Equilibrium distribution(s) exist with eigenvalue 1.
Irreducible	Strongly connected	" Unique, positive equilibrium distribution π
Primitive	Strongly connected (and more?)	" $P^k \rightarrow \pi 1'$
Positive	"Fully" connected	"

eigen
w

Exercise 7-9-3. When reviewing this chapter, consider these additional properties:

- $\rho \geq 0$ vs $\rho > 0$
- $\lambda_{(1)} = \rho$ vs $|\lambda_{(1)}| = \rho$
- uniqueness of $\lambda_{(1)}$ (can any other $\lambda_i = \lambda_{(1)}$?)
- uniqueness of $|\lambda_{(1)}|$ (can any other $|\lambda_i| = |\lambda_{(1)}|$?)
- existence of $v_1 \geq 0$ vs existence of $v_1 > 0$

Summary of properties for **Markov chain transition matrix P** (square and nonnegative), where $\mathbf{1}'P = \mathbf{1}'$.

	any	not aperiodic	irreducible	primitive	positive
$\rho(\mathbf{P}) = 1$	y	y	y	y	y
\mathbf{P} has an eigenvalue equal to 1	y	y	y	y	y
\mathbf{P} has exactly one eigenvalue equal to 1	.	n	y	y	y
\mathbf{P} has exactly one eigenvalue whose magnitude is 1	.	n	n	y	y
\mathbf{P} has all nonnegative eigenvalues	.	n	n	n	(i) <i>v</i>
\mathbf{P} has a nonnegative eigenvector with eigenvalue $\rho(\mathbf{P})$	y	y	y	y	y
\mathbf{P} has a positive eigenvector with eigenvalue $\rho(\mathbf{P})$	n	?	y	y	y
\mathbf{P} has a unique equilibrium distribution	.	n	y	y	y
power iter. $\{\mathbf{P}^k \pi\}$ converges to the same limit $\forall \pi$ in N -simplex	.	n	n	y	y
$\{\mathbf{P}^k\}$ converges to a rank-1 matrix $\rightarrow \text{rank } 1$.	n	n	y	y
can reach any state from any state	.	n	y	y	y
graph is strongly connected	.	n	y	y	y
$\exists m \in \mathbb{N}$: can reach any state from any state in m steps	.	n	n	y	y
can reach any state from any state in 1 step	.	n	n	n	(ii) <i>y</i>
counter-example for "n" entries	$\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$		

24.

The entries (i),(ii) are:

A: (n,n)

B: (n,y)

C: (y,n)

D: (y,y)

??

$$\frac{1}{4} \begin{bmatrix} 1 & 3 \\ 3 & 1 \end{bmatrix} \quad \text{eig: } 1, -\frac{1}{2}$$

$$\mathbf{1}' \mathbf{P} = \mathbf{1}'$$

$\mathbf{1}' \mathbf{1} = 1$

25. For any Markov chain with a $N \times N$ circulant transition matrix, the vector $\pi = \mathbf{1}_N/N$ is an equilibrium distribution of the chain. (?)

A: True

$$\mathbf{P} \mathbf{1} = c \mathbf{1}$$

B: False

$$\mathbf{P} \pi = \pi$$

??

$$\Rightarrow \frac{1}{N} \mathbf{1}' \mathbf{P} \mathbf{1} = c \mathbf{1}' \mathbf{1} = cN \Rightarrow c = 1$$

26. For any Markov chain with a $N \times N$ circulant transition matrix, the vector $\pi = \mathbf{1}_N/N$ is the unique equilibrium distribution of the chain. (?)

A: True

$$\mathbf{B} \quad \mathbf{B}' \quad \mathbf{B}'$$

B: False

$$\mathbf{P} = \mathbf{I}_N$$

$$\mathbf{I} \pi = \pi \quad \forall \pi$$

??

27. For any Markov chain with a $N \times N$ circulant transition matrix, exactly one eigenvalue of that transition matrix has magnitude equal to 1. (?)

A: True

B: False

$$\text{eig}(\mathbf{I}) = 1, 1, \dots, 1$$

??

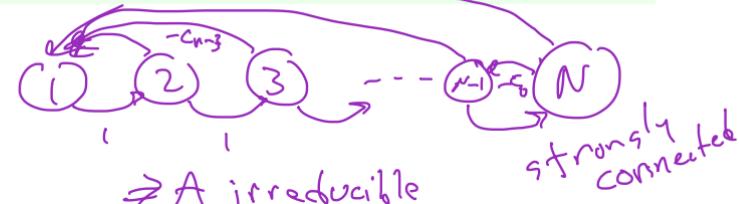
28. If a monic polynomial has all negative coefficients (except for the leading coefficient), then the corresponding companion matrix has an eigenvector with all positive elements. (?)

A: True

B: False

$$-c_{m-1} > 0$$

$$A = \begin{pmatrix} -c_{m-1} & -c_{m-2} & \cdots & -c_1 & -c_0 \\ 1 & & & & \\ \vdots & & & & \end{pmatrix}$$


 $\Rightarrow A \text{ irreducible}$

$$(x-3)(x+1)$$

$$\text{cr}_q(A) = 3, -1$$

29.

The companion matrix for the polynomial $p(x) = x^2 - 2x - 3$ has spectral radius:

A: 1

B: 2

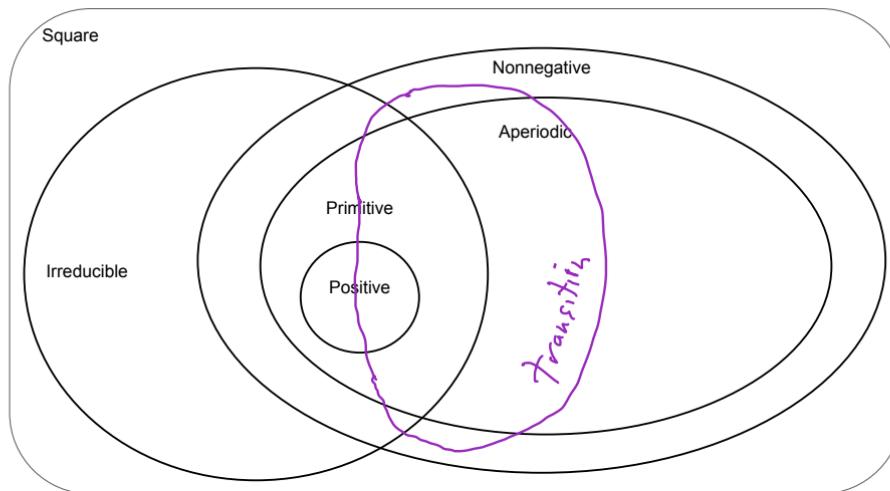
C: 3

D: 4

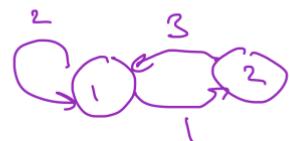
E: 5

??

This diagram summarizes the relationships between many of the matrix types in this chapter.



$$A = \begin{bmatrix} 2 & 3 \\ 1 & 0 \end{bmatrix}$$



An important category that is missing from this diagram is transition matrices.

Exercise 7-9-4. Complete the Venn diagram by adding **transition** matrices.

(in class) ✓

??

nonzero

Exercise 7-9-5. Give an example of a matrix that is nonnegative but is in no other (square) category.

??

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

↑ not positive

not a transition matrix



graph
not strongly connected
⇒ A not irreducible
 $I^T A \neq I^T$ ⇒ not primitive

$$A^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\left[A^m \right]_{2 \times 2}$$

always zero, so no period

⇒ not aperiodic

Bibliography

- [1] M. Akcakaya and V. Tarokh. “A frame construction and a universal distortion bound for sparse representations”. In: *IEEE Trans. Sig. Proc.* 56.6 (June 2008), 2443–50 (cit. on p. [7.11](#)).
- [2] A. J. Laub. *Matrix analysis for scientists and engineers*. Soc. Indust. Appl. Math., 2005 (cit. on pp. [7.14](#), [7.15](#)).
- [3] J. P. Boyd. “Finding the zeros of a univariate equation: proxy rootfinders, Chebyshev interpolation, and the companion matrix”. In: *SIAM Review* 55.2 (2013), 375–396 (cit. on p. [7.15](#)).
- [4] J. W. Cooley and J. W. Tukey. “An algorithm for machine calculation of complex Fourier series”. In: *Mathematics of Computation* 19.90 (Apr. 1965), 297–301 (cit. on p. [7.21](#)).
- [5] H. Akaike. “Block Toeplitz matrix inversion”. In: *SIAM J. Appl. Math.* 24.2 (Mar. 1973), 234–41 (cit. on p. [7.24](#)).
- [6] M. Wax and T. Kailath. “Efficient inversion of Toeplitz-block Toeplitz matrix”. In: *IEEE Trans. Acoust. Sp. Sig. Proc.* 31.5 (Oct. 1983), 1218–21 (cit. on p. [7.24](#)).
- [7] L. Rodman and T. Shalom. “On inversion of symmetric Toeplitz matrices”. In: *SIAM J. Matrix. Anal. Appl.* 13.2 (1992), 530–49 (cit. on p. [7.24](#)).
- [8] M. K. Ng, K. Rost, and Y-W. Wen. “On inversion of Toeplitz matrices”. In: *Linear Algebra and its Applications* 348.1 (June 2002), 145–51 (cit. on p. [7.24](#)).
- [9] I. Gohberg and G. Heinig. *Inversion of finite Toeplitz matrices*. 2010 (cit. on p. [7.24](#)).
- [10] V. Sukhoy and A. Stoytchev. “Generalizing the inverse FFT off the unit circle”. In: *Nature: Scientific Reports* 9.1 (2019) (cit. on p. [7.24](#)).
- [11] H. G. Feichtinger, K. Gröchenig, and T. Strohmer. “Efficient numerical methods in non-uniform sampling theory”. In: *Numerische Mathematik* 69.4 (Feb. 1995), 423–40 (cit. on p. [7.24](#)).
- [12] J. A. Fessler, S. Lee, V. T. Olafsson, H. R. Shi, and D. C. Noll. “Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity”. In: *IEEE Trans. Sig. Proc.* 53.9 (Sept. 2005), 3393–402 (cit. on p. [7.24](#)).
- [13] Z. Yang, L. Xie, and P. Stoica. “Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution”. In: *IEEE Trans. Info. Theory* 62.6 (June 2016), 3685–701 (cit. on p. [7.24](#)).
- [14] E. K. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin. “Plug-and-play methods provably converge with properly trained denoisers”. In: *Proc. Intl. Conf. Mach. Learn.* 2019 (cit. on p. [7.25](#)).

- [15] B. G. Horne. "Lower bounds for the spectral radius of a matrix". In: *Linear Algebra and its Applications* 263 (Sept. 1997), 261–73 (cit. on p. 7.33).
- [16] N. J. Higham and Francoise Tisseur. "Bounds for eigenvalues of matrix polynomials". In: *Linear Algebra and its Applications* 358.1 (2003), 5–22 (cit. on p. 7.33).
- [17] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge: Cambridge Univ. Press, 1985 (cit. on p. 7.33).
- [18] S. U. Pillai, T. Suel, and S. Cha. "The Perron-Frobenius theorem: some of its applications". In: *IEEE Sig. Proc. Mag.* 22.2 (Mar. 2005), 62–75 (cit. on pp. 7.34, 7.50, 7.63, 7.70).
- [19] H. Schneider. "Wielandt's proof of the exponent inequality for primitive nonnegative matrices". In: *Linear Algebra and its Applications* 353.1 (Sept. 2002), 5–10 (cit. on p. 7.36).
- [20] R. S. Varga. *Geršgorin and his circles*. Springer, 2004 (cit. on p. 7.44).
- [21] C. D. Meyer. *Matrix analysis and applied linear algebra*. Soc. Indust. Appl. Math., 2000 (cit. on pp. 7.44, 7.47).
- [22] H. Minc. *Nonnegative matrices*. New York: Wiley, 1988 (cit. on p. 7.50).
- [23] F. G. Frobenius. "Über Matrizen aus nicht negativen Elementen". In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* 26 (May 1912), 456–77 (cit. on p. 7.50).
- [24] A. Berman and R. J. Plemmons. *Nonnegative matrices in the mathematical sciences*. Soc. Indust. Appl. Math., 1994 (cit. on p. 7.50).
- [25] K. Bryan and T. Leise. "The \$25,000,000,000 eigenvector: The linear algebra behind Google". In: *SIAM Review* 48.3 (Sept. 2006), 569–81 (cit. on p. 7.71).
- [26] S. Feizi, G. Quon, M. Recamonde-Mendoza, M. Medard, M. Kellis, and A. Jadbabaie. *Spectral alignment of graphs*. 2017 (cit. on p. 7.74).

Solutions to explorations

Explore 7.0.1: For $n = 1$ we have $\mathbf{A} = \begin{bmatrix} -c_1 & -c_0 \\ 1 & 0 \end{bmatrix}$ which has characteristic polynomial $\det\{\mathbf{zI} - \mathbf{A}\} =$

n=2 actually

$$\det \left\{ \begin{bmatrix} z + c_1 & c_0 \\ -1 & z \end{bmatrix} \right\} = z^2 + zc_1 + c_0.$$

Explore 7.0.2: The numerical computation of the eigenvalues of \mathbf{A} is imperfect, returning $3 \pm i\epsilon$ instead of $[3, 3]$.

Explore 7.0.3: Consider $\mathbf{A} = \mathbf{I}_2$, $\mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

Explore 7.0.4: $\langle \mathbf{v}_k, \mathbf{v}_l \rangle = \sum_{n=0}^{N-1} w_k^n (w_l^n)^* = \sum_{n=0}^{N-1} e^{i2\pi(k-l)n/N} = 0$, $k \neq l$.

Explore 7.0.5: $Q_{nk} = w_k^n = e^{i2\pi kn/N} = Q_{kn} = w_n^k$.

Explore 7.0.6: If one can reach any node from any node in m steps, then one can also do so in any $n \geq m$ steps.

Explore 7.0.7: The graph is “fully” connected.

Explore 7.0.8: $\mathbf{P} = \begin{bmatrix} 1/2 & 0 \\ 1/2 & 1 \end{bmatrix}$.

Explore 7.0.9: $\exists N \in \mathbb{N}$ s.t. $[\mathbf{P}^n]_{ii} = 0$, $\forall n \geq N$.