Chapter 7

Special matrices

Contents (class version)

7.0 Introduction

This chapter contains topics related to matrices with special structures that arise in many applications.

7.1 Companion matrices

L§10.4

Previously we defined the **eigenvalues** of a square matrix $\boldsymbol{A}$ to be the roots of the **characteristic polynomial**:

$$\det\{\boldsymbol{A} - z\boldsymbol{I}\} = 0.$$

Here we work somewhat in the reverse direction by starting with a polynomial and then defining a matrix from it. Consider the **monic polynomial**

$$p(z) = z^n + c_{n-1}z^{n-1} + \cdots + c_1 z + c_0, \quad z \in \mathbb{C},$$

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

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

Which matrix class is this?
A: Lower triangular B: Upper triangular C: Lower Hessenberg D: Upper Hessenberg E: None of these

??

Example. For $n = 3$ we have

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

To examine the eigenvalues of $\boldsymbol{A}$, evaluate the **determinant** using the usual minors:

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

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

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

$$\det\{z\boldsymbol{I} - \boldsymbol{A}\} = \det\left\{\begin{bmatrix} z + c_{n-1} & c_{n-2} & \dots & c_1 & c_0 \\ -1 & z & 0 & \dots & 0 \\ 0 & -1 & z & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & z \end{bmatrix}\right\} = (z + c_{n-1}) \underbrace{\det\left\{\begin{bmatrix} z & 0 & \dots & 0 \\ -1 & z & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & -1 & z \end{bmatrix}\right\}}_{(n-1)\times(n-1)}$$

$$- c_{n-2} \underbrace{\det\left\{\left[\begin{array}{c|ccc} -1 & 0 & 0 & \dots & 0 \\ \hline 0 & z & 0 & \dots & 0 \\ 0 & -1 & z & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -1 & z \end{array}\right]\right\}}_{(n-1)\times(n-1)} + c_{n-3} \underbrace{\det\left\{\left[\begin{array}{cc|ccc} -1 & z & 0 & 0 & \dots & 0 \\ 0 & -1 & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & z & 0 & \dots & 0 \\ 0 & 0 & -1 & z & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & -1 & z \end{array}\right]\right\}}_{(n-1)\times(n-1)} + c_{n-4} \cdots$$

$$= (z + c_{n-1}) z^{n-1} + c_{n-2} z^{n-2} + c_{n-3} z^{n-3} + \cdots .$$

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 $\det\left\{\begin{bmatrix} \boldsymbol{B} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{C} \end{bmatrix}\right\} = \det\{\boldsymbol{B}\}\det\{\boldsymbol{C}\}$. Thus

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

In words: the polynomial $p(z)$ is the **characteristic polynomial** of $\boldsymbol{A}$.
It is also its **minimal polynomial**. (See next page.)
Due to this property, the matrix $\boldsymbol{A}$ so defined is called a **companion matrix** of the monic polynomial $p(z)$.

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

There are other rearrangements of $\boldsymbol{A}$ (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 -> [-reverse(c,dims=1)'; [I zeros(length(c)-1)]]
A = compan(c)
```

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

(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

The **characteristic polynomial** $p(z)$ of a matrix $N \times N$ matrix $\boldsymbol{A}$ has the property that

$$p(\boldsymbol{A}) = \sum_{k=0}^{N} c_k \boldsymbol{A}^k = c_0 \boldsymbol{I}_N + \sum_{k=1}^{N} c_k \boldsymbol{A}^k = \boldsymbol{0}_{N \times N},$$

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

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

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

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

$$\mu(\boldsymbol{A}) = 1 \cdot \boldsymbol{A}^1 + (-\beta)\boldsymbol{A}^0 = \boldsymbol{A} - \beta \boldsymbol{I} = \boldsymbol{0}.$$

The degree of the minimal polynomial of a $N \times N$ matrix is always in the set $\{1, \ldots, N\}$.
A: True B: False **??**

Eigenvectors of companion matrices

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

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

$$= \begin{bmatrix} -c_{n-1}t^{n-1} - \cdots - c_1 t - c_0 \\ t^{n-1} \\ \vdots \\ t \end{bmatrix} \stackrel{?}{=} \begin{bmatrix} t^n \\ t^{n-1} \\ \vdots \\ t \end{bmatrix} = t \begin{bmatrix} t^{n-1} \\ \vdots \\ t \\ 1 \end{bmatrix} = t\boldsymbol{v},$$

where the inner equality holds iff

$$-c_{n-1}t^{n-1} - \cdots - c_1 t - c_0 = t^n, \; \textit{i.e.}, \; 0 = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0 = p(t).$$

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

Diagonalizing a companion matrix

Each root of $p(z)$ has a corresponding eigenvector (that we can normalize to have unit norm).
If $p(z)$ has any repeated roots, then the matrix $\boldsymbol{A}$ is not diagonalizable.
(Recall fact above: Matrix $\boldsymbol{A}$ is **diagonalizable** iff its **minimal polynomial** is a product of distinct factors.)

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

$$\boldsymbol{V}^{-1} \boldsymbol{A} \boldsymbol{V} = \mathsf{diag}\{z_1, \ldots, z_n\},$$

where $\{z_1, \ldots, z_n\}$ denote the roots of the polynomial (*i.e.*, the eigenvalues of $\boldsymbol{A}$), and where the eigenvector matrix $\boldsymbol{V}$ is an $n \times n$ **Vandermonde matrix** of the form on the right:

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

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 $z_k = \mathrm{e}^{-\imath 2\pi k/N}$.

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

The companion matrix for the polynomial $p(z) = z^2 - 9z + 6$ is diagonalizable.
A: True B: False **??**

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 $\boldsymbol{A}$ with a $N \times N$ matrix $\boldsymbol{B}$ is the following $MN \times MN$ matrix [2, p.143]:

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

L§13.2

Wikipedia uses this definition:

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

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

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

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

Fact. [2, Thm. 13.16]
If $\boldsymbol{A}$ has eigenvalues $\{\lambda_m,\ m = 1, \dots, M\}$ and $\boldsymbol{B}$ has eigenvalues $\{\mu_n,\ n = 1, \dots, N\}$,
then the MN eigenvalues of $\boldsymbol{A} \oplus \boldsymbol{B}$ are

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

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

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

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 $\boldsymbol{A}$ and $\boldsymbol{B}$, then $p_1(z)$ and $p_2(z)$ share a common root iff $\boldsymbol{A}$ and $\boldsymbol{B}$ 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 $\boldsymbol{A} \oplus (-\boldsymbol{B})$ 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].

7.2 Circulant matrices

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

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

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

$$\boldsymbol{G}_N \triangleq \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

This special case is a **circulant matrix**. (Indeed it is the parent of all circulant matrices, as shown below.)

What are the eigenvectors of this matrix $\boldsymbol{G}_N$? The roots of $p(z)$ here are solutions to $z^N - 1 = 0$, *i.e.*, the Nth roots of unity: $z_k = \mathrm{e}^{-\imath 2\pi k/N}$, $k = 0, \dots, N-1$. An eigenvector corresponding to the kth root z_k is

$$\boldsymbol{v}_k \triangleq \frac{1}{\sqrt{N}} z_k^{1-N} \begin{bmatrix} z_k^{N-1} \\ z_k^{N-2} \\ \vdots \\ z_k \\ 1 \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} z_k^{0} \\ z_k^{-1} \\ \vdots \\ z_k^{2-N} \\ z_k^{1-N} \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 \\ w_k \\ \vdots \\ w_k^{N-2} \\ w_k^{N-1} \end{bmatrix}, \quad w_k \triangleq \mathrm{e}^{\imath 2\pi k/N},$$

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 another Vandermonde matrix:

$$\boldsymbol{Q} \triangleq \begin{bmatrix} \boldsymbol{v}_0 & \dots \boldsymbol{v}_{N-1} \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{bmatrix} & \dots & \begin{bmatrix} 1 \\ w_k \\ \vdots \\ w_k^{N-2} \\ w_k^{N-1} \end{bmatrix} & \dots & \begin{bmatrix} 1 \\ w_{N-1} \\ \vdots \\ w_{N-1}^{N-2} \\ w_{N-1}^{N-1} \end{bmatrix} \end{bmatrix}. \tag{7.1}$$

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

$$\boldsymbol{G}_N = \boldsymbol{Q} \, \mathsf{diag}\{z_0, \dots, z_{N-1}\} \, \boldsymbol{Q}'.$$

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

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

Exercise. Verify that $\boldsymbol{Q}$ is a symmetric (not Hermitian!) matrix.

Now consider a general **circulant matrix**:

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

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

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

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

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

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

Thus

$$\boldsymbol{Q}'\boldsymbol{C}\boldsymbol{Q} = \boldsymbol{Q}' \left(\sum_{n=0}^{N-1} c_n \boldsymbol{G}_N^n \right) \boldsymbol{Q} = \sum_{n=0}^{N-1} c_n \left(\boldsymbol{Q}' \boldsymbol{G}_N^n \boldsymbol{Q} \right) = \sum_{n=0}^{N-1} c_n \, \mathsf{diag}\left\{ z_0^n, \ldots, z_{N-1}^n \right\}$$
$$= \mathsf{diag}\left\{ \sum_{n=0}^{N-1} c_n z_0^n, \ldots, \sum_{n=0}^{N-1} c_n z_{N-1}^n \right\} \boldsymbol{\Lambda} = \mathsf{diag}\{\lambda_k\}, \quad \lambda_k \triangleq \sum_{n=0}^{N-1} c_n z_k^n,$$

so the eigendecomposition of $\boldsymbol{C}$ is

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

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

The kth eigenvalue of $\boldsymbol{C}$ is

$$\lambda_k = \sum_{n=0}^{N-1} c_n z_k^n = \sum_{n=0}^{N-1} c_n \, \mathrm{e}^{-\imath 2\pi nk/N}, \; k = 0, \ldots, N-1,$$

which is the (ordinary, *not* orthonormal) DFT of the *first column* of $\boldsymbol{C}$.

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:

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

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

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

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

The number of distinct eigenvalues of the matrix $\boldsymbol{G}_6^3$ is:

A: 1 B: 2 C: 3 D: 4 E: 6 **??**

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 $\boldsymbol{C}$ by a length-N vector, *i.e.*, $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}$ via `y = C * x`, would require $O(N^2)$ operations. When $\boldsymbol{C}$ 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 = ifft(fft(c) .* fft(x))
```

where `c` is the first column of $\boldsymbol{C}$.

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 $\boldsymbol{C}$ and $\boldsymbol{x}$ are both real then of course $\boldsymbol{y} = \boldsymbol{C}\boldsymbol{x}$ 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

Every circulant matrix is a **normal matrix** (?)
A: True B: False **??**

If $\boldsymbol{C}$ is a circulant matrix, then its **spectral norm** is $\sigma_1 =$ `maximum(abs.(fft(C[:,1])))` (?)
A: True B: False **??**

Inverting a circulant matrix

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

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

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

Inverting a Toeplitz matrix

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

7.3 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, but it was used recently as part of training a CNN for image denoising [11], so it continues to be important.

We showed previously that if square matrix $\boldsymbol{A} \in \mathbb{F}^{N \times N}$ is **diagonalizable**, then

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

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

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

Now make some assumptions:

- WLOG we order the eigenvalues with decreasing magnitudes: $|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_N|$.
- Assume $\lambda_1 \neq 0$ to avoid the trivial case of $\boldsymbol{A} = \boldsymbol{0}$.
- Assume $z_1 = [\boldsymbol{V}^{-1} \boldsymbol{x}_0]_1 \neq 0$, *i.e.*, the initial vector $\boldsymbol{x}_0$ has a nonzero coordinate in the direction associated with the dominant eigenvalue λ_1. Choosing $\boldsymbol{x}_0$ at random ensures $z_1 \neq 0$ with probability 1.

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

$$\boldsymbol{x}_k \triangleq \frac{\boldsymbol{A}^k \boldsymbol{x}_0}{\|\boldsymbol{A}^k \boldsymbol{x}_0\|_2}, \quad k = 1, 2, \dots. \tag{7.4}$$

This definition leads to the following simple recursion:

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

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

$$\boldsymbol{x}_{k+1} = \frac{\boldsymbol{A}\boldsymbol{x}_k}{\|\boldsymbol{A}\boldsymbol{x}_k\|_2}, \quad k = 0, 1, \dots. \tag{7.5}$$

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

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

where $\boldsymbol{v}_n$ denotes a (unit-norm) eigenvector of $\boldsymbol{A}$ corresponding to λ_n, *i.e.*, the nth column of $\boldsymbol{V}$.

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

Convergence of the power iteration

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

$$|\lambda_n| < |\lambda_1|, \; n = 2, \ldots, N,$$

then one can show that $\left\| \mathrm{e}^{-\imath \angle z_1} \left(\mathrm{e}^{-\imath \angle \lambda_1} \right)^k \boldsymbol{x}_k - \boldsymbol{v}_1 \right\| \to 0$, because $\mathrm{e}^{-\imath \angle z_1} \, z_1 = |z_1|$ so

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

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

Another way of saying this is $\left\| \boldsymbol{P}_{\boldsymbol{v}_1}^{\perp} \boldsymbol{x}_k \right\| \to 0$ as $k \to \infty$.

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

$$\boldsymbol{x}_k' \boldsymbol{A} \boldsymbol{x}_k \to \lambda_1.$$

Here we do not need to worry about the extra phase term $\mathrm{e}^{\imath \angle z_1}\, \mathrm{e}^{\imath k \angle \lambda_1}$ because it cancels out due to $\boldsymbol{x}_k'$ and $\boldsymbol{x}_k$:

$$(\mathrm{e}^{\imath\phi}\, \boldsymbol{v}_1)' \boldsymbol{A} (\mathrm{e}^{\imath\phi}\, \boldsymbol{v}_1) = \mathrm{e}^{-\imath\phi}\, \mathrm{e}^{\imath\phi}\, \boldsymbol{v}_1' (\boldsymbol{A}\boldsymbol{v}_1) = \boldsymbol{v}_1' (\lambda_1 \boldsymbol{v}_1) = \lambda_1 .$$

- 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 $\boldsymbol{x}_k$ 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 $\mathrm{e}^{\imath\phi}$ 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 $\boldsymbol{A}$ is diagonalizable here, but the **power iteration convergence analysis** generalizes to non-diagonalizable matrices using the **Jordan form**. So sufficient conditions are:
  - $\boldsymbol{A}$ is square
  - $z_1 = [\boldsymbol{A}\boldsymbol{x}_0]_1 \neq 0$
  - $|\lambda_1|$ dominates the other eigenvalue magnitudes
- For analyzing Markov chains (next), we will see that the dominant eigenvalue of the transition matrix $\boldsymbol{P}$ is real and positive. For any matrix where λ_1 is real and positive, we have that $\mathrm{e}^{\imath \angle \lambda_1} = 1$ so there is no problematic oscillation term. In this case, under the sufficient conditions above, we can say that the power

iteration converges:

$$\boldsymbol{x}_k \to \mathrm{e}^{\imath \angle z_1} \, \boldsymbol{v}_1.$$

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

To determine the principal left singular vector $\boldsymbol{u}_1$, we apply the power iteration to

A: $\boldsymbol{A}'$ B: $\boldsymbol{A}$ C: $\boldsymbol{A}\boldsymbol{A}'$ D: $\boldsymbol{A}'\boldsymbol{A}$ E: None of these **??**

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*: $\rho(\boldsymbol{A}) = \max_i |\lambda_i(\boldsymbol{A})|$.

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

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

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

$$\lambda_i \in$$

More generally, if the union of J disks is disjoint from the union of the other $N - J$ discs then the former union contains exactly J and the latter $N - J$ eigenvalues of $\boldsymbol{A}$.

Example. If $\boldsymbol{P}$ 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 $\boldsymbol{P}$ are 1 and $\boldsymbol{P}$ is normal so its singular values are the absolute value of its eigenvalues.)

General spectral radius bounds

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

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

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

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

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

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

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

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

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

from which we could obtain a (often loose) lower bound on ρ.
Challenge: find a useful lower bound for general square matrices [12] [13].

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

7.4 Nonnegative matrices and Perron-Frobenius theorem

Define. A matrix $\boldsymbol{A}$ is called a **positive matrix** iff all of its elements are real and positive.

Define. A matrix $\boldsymbol{A}$ is called a **nonnegative matrix** iff all of its elements are real and nonnegative.

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

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

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

These definitions are unrelated to **positive definite** and **positive semi-definite**, except in special cases like **diagonal** matrices.

There are many applications of such matrices.

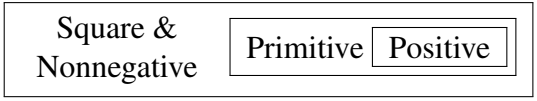
- Such matrices arise when describing transition probabilities in certain **Markov chains**.
- Google's **PageRank** algorithm for ranking web sites is built on a certain nonnnegative matrix.
- Other applications include power control, commodity pricing, and population growth [14].

Any positive semi-definite diagonal matrix is a nonnegative matrix. (?)
A: True B: False **??**

Any positive definite diagonal matrix is a positive matrix. (?)
A: True B: False **??**

The circulant generator matrix $\boldsymbol{G}_N$ from the previous class is (choose most specific answer):
A: Square B: Nonnegative C: Primitive D: Positive E: None of these **??**

Venn diagram (for square matrices only):



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

Square positive, primitive, and nonnegative matrices have various special properties in terms of their eigenvalues and eigenvectors that are important in applications. We focus on those properties next.

Properties of nonnegative matrices

We start with the broadest category: **nonnegative matrices**.

Fact. If $\boldsymbol{A}$ is **square** and **nonnegative**, then we have the following bounds on its **spectral radius** [14]:
- Row sums provide bounds on the spectral radius: $\min_i \sum_j a_{ij} \le \rho(\boldsymbol{A}) \le \max_i \sum_j a_{ij} = \|\boldsymbol{A}\|_\infty$
- Column sums provide bounds on the spectral radius: $\min_j \sum_i a_{ij} \le \rho(\boldsymbol{A}) \le \max_j \sum_i a_{ij} = \|\boldsymbol{A}\|_1$

Combining yields the following lower bound on the spectral radius of a nonnegative matrix:

$$\max\left(\min_j \sum\nolimits_i a_{ij}, \min_i \sum\nolimits_j a_{ij} \right) \le \rho(\boldsymbol{A}).$$

Challenge: [14] claims the bounds are easy to show. Find a simple proof of the lower bound.

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

The **Perron-Frobenius theorem** for any square, **nonnegative** matrix $\boldsymbol{A}$ states the following:
- $r \triangleq \rho(\boldsymbol{A}) \ge 0$ is an eigenvalue of $\boldsymbol{A}$.
  In other words, the largest magnitude eigenvalue is a real, nonnegative eigenvalue. Think $\lambda_1 = r \ge 0$.
  We call r the **Perron root** or **Perron-Frobenius eigenvalue**.
- $\boldsymbol{A}$ has a (right) eigenvector $\boldsymbol{v}$ *with nonnegative elements* corresponding to that eigenvalue, *i.e.*, $\boldsymbol{A}\boldsymbol{v} = r\boldsymbol{v}$.
- There likewise exists a left eigenvector $\boldsymbol{w}$ having all nonnegative elements for which $\boldsymbol{w}'\boldsymbol{A} = r\boldsymbol{w}'$.
  We refer to such $\boldsymbol{v}$ and $\boldsymbol{w}$ as right and left **Perron vectors**, respectively.

Remarks.

- Because $\boldsymbol{A}$ and $\boldsymbol{A}'$ have the same eigenvalues, the existence of the nonnegative left eigenvector follows directly from the existence of the nonnegative right eigenvector, because $\boldsymbol{A}'$ 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** (5.22), followed by noting that any nonnegative matrix is a limit of positive matrices.
- By definition of spectra radius: $|\lambda_i(\boldsymbol{A})| \leq \rho(\boldsymbol{A})$, so for a square nonnegative matrix $|\lambda_i(\boldsymbol{A})| \leq r = \lambda_1$.
- This inequality is not strict and in general there can be multiple eigenvalues with the same magnitude. 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-collinear right nonnegative eigenvectors having an eigenvalue where $|\lambda_i| = r$.
- In other words, we cannot 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 $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ the eigenvalues are $1, 1$. Any nonzero vector of the form $\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]$ with $x, y \geq 0$ is a (right and left) Perron vector.

Example. For $\boldsymbol{A} = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ the eigenvalues are $+1, -1$. The vector $\left[\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right]$ is a (right and left) Perron vector; the vector $\left[\begin{smallmatrix} -1 \\ 1 \end{smallmatrix}\right]$ is also an eigenvector whose eigenvalue, -1, also has magnitude 1. Powers $\boldsymbol{A}^k$ do not converge.

Properties of primitive matrices

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

The **Perron-Frobenius theorem** for any (square) **primitive matrix** $\boldsymbol{A}$ states that:

- All the conclusions for (square) nonnegative matrices hold.
- $r \triangleq \rho(\boldsymbol{A}) > 0$ is a (real) eigenvalue of $\boldsymbol{A}$.
  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: For any other eigenvalue λ, $|\lambda| < r = \rho(\boldsymbol{A})$, *i.e.*, there is no other λ such that $|\lambda| = r$.
- There exists a unit-norm eigenvector $\boldsymbol{v}$ with eigenvalue r, *i.e.*, $\boldsymbol{A}\boldsymbol{v} = r\boldsymbol{v}$, where $\boldsymbol{v}$ has all *positive* elements.
- There likewise exists a unit-norm left eigenvector $\boldsymbol{w}$ having all positive elements for which $\boldsymbol{w}'\boldsymbol{A} = r\boldsymbol{w}'$.
  We call this $\boldsymbol{v}$ and $\boldsymbol{w}$ the right and left **Perron vectors**, respectively
- Uniqueness: All other (left and right) unit-norm eigenvectors have negative and/or non-real elements.
  So $\boldsymbol{v}$ and $\boldsymbol{w}$ are unique in that sense.

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

Because a primitive matrix $\boldsymbol{A}$ has one eigenvalue whose magnitude dominates all others, the **power iteration converges** to the **Perron vector** (to within a constant $\mathrm{e}^{\imath\phi}$).

Some, but not all, of these properties generalize to the broader class of **irreducible matrices** [14].

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

Markov chains

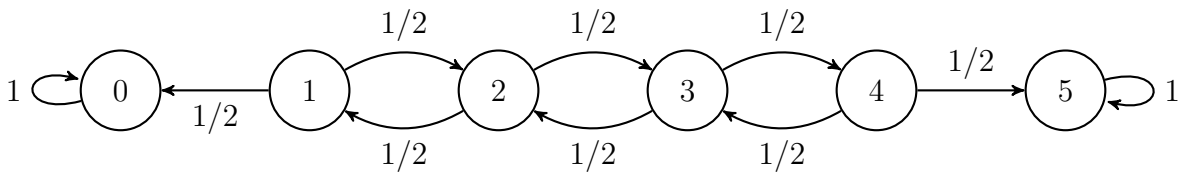
Define. A (first-order) **Markov chain** is a random process (or random sequence, *i.e.*, a sequence of random variables) $X_1, X_2, \ldots$ whose joint distribution is such that the conditional distribution of X_{k+1} given the values (states) of all preceding variables depends only on the preceding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

$$\mathsf{p}(X_{k+1} = x \,|\, X_k = x_k, X_{k-1} = x_{k-1}, \ldots, X_1 = x_1) = \mathsf{p}(X_{k+1} = x \,|\, X_k = x_k)\,.$$

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

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

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 that 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).

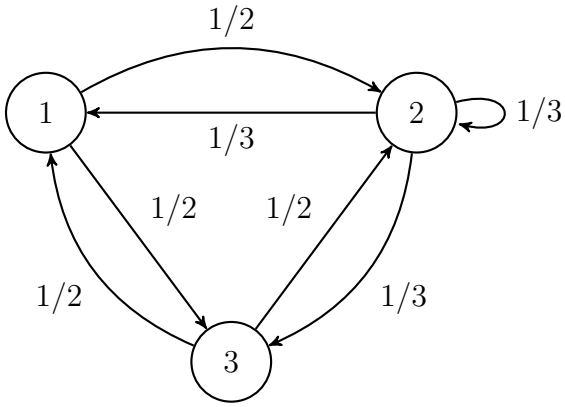
In this typical case, the statistical properties are governed by the **transition probabilities**:

$$\mathsf{p}(X_{k+1} = i \,|\, X_k = j), \quad i, j = 1, \dots, N.$$

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 **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.

Example.



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

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

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

$$= \begin{bmatrix} \\ \mathsf{p}(X_{k+1} = i \,|\, X_k = j) \\ \\ \end{bmatrix}$$

Sometimes 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.*, $\mathbf{1}_3' \boldsymbol{P} = \mathbf{1}_3'$.

Here is a demo that also shows 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

Properties of transition matrices

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

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

$$\mathbf{1}_N' \boldsymbol{P} = \mathbf{1}_N' \text{ because } \sum_{n=1}^{N} \mathsf{p}(X_{k+1} = n \,|\, X_k = j) = 1. \tag{7.6}$$

  In words, the transition probabilities for leaving the jth state sum to one, for every j.
- Due to (7.6), a $N \times N$ transition matrix $\boldsymbol{P}$ always has $\mathbf{1}_N$ as a left eigenvector with eigenvalue 1.
- Because all columns sum to unity: $\rho(\boldsymbol{P}) \leq 1$ so $|\lambda_i(\boldsymbol{P})| \leq 1$. (No eigenvalues are outside the unit disk.)
- Thus $\rho(\boldsymbol{P}) = 1$.

In general, there can be multiple nonnegative eigenvectors having eigenvalue equal to unity.
Example. The following transition matrix has two distinct eigenvectors with eigenvalue 1:

$$\boldsymbol{P} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \implies \boldsymbol{v}_1 = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix} \text{ and } \boldsymbol{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{7.7}$$

Is this $\boldsymbol{P}$ **primitive**?
A: Yes B: No **??**

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

$$\pi_n \geq 0, \quad n = 1, \ldots, N, \qquad \sum_{n=1}^{N} \pi_n = \mathbf{1}_N' \boldsymbol{\pi} = 1, \qquad \boldsymbol{P}\boldsymbol{\pi} = \boldsymbol{\pi}.$$

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

- Because a transition matrix $\boldsymbol{P}$ is nonnegative and has $\rho(\boldsymbol{P}) = 1$, 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.
- However, that eigenvalue of 1 need not be unique and there can be multiple stochastic eigenvectors. Again, see preceding example.

The importance of an equilibrium distribution is the following.

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

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

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

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

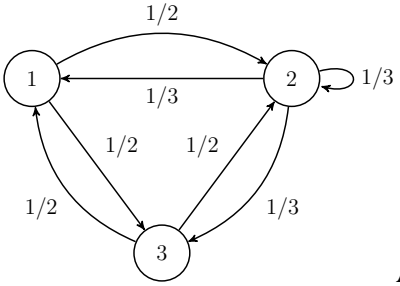
In matrix vector form:

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

In other words, an equilibrium distribution $\boldsymbol{\pi}$ is a stochastic eigenvector of the transition matrix $\boldsymbol{P}$ 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 the earlier example, the eigenvalues of $\boldsymbol{P} = \begin{bmatrix} 0 & 1/3 & 1/2 \\ 1/2 & 1/3 & 1/2 \\ 1/2 & 1/3 & 0 \end{bmatrix}$ are $\{1, -1/2, -1/6\}$.



A (unit norm) eigenvector associated with eigenvalue 1 is $\boldsymbol{v}_1 = \begin{bmatrix} 0.485071 \\ 0.727607 \\ 0.485071 \end{bmatrix}$.

One can verify that $\|\boldsymbol{v}_1\|_2 = 1$. But this is not the normalization we want for a probability distribution.

We want $\sum_n \pi_n = 1$ *i.e.*, $\mathbf{1}_N' \boldsymbol{\pi} = 1$.

So define $\boldsymbol{\pi} = \boldsymbol{v}_1 / (\mathbf{1}_N' \boldsymbol{v}_1)$, possibly with sign flip if needed, for which $\boldsymbol{\pi} = \begin{bmatrix} 0.285714 \\ 0.428571 \\ 0.285714 \end{bmatrix}$.

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.

Limiting distribution

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 nth state? Formally we want to examine:

$$\lim_{k \to \infty} \mathsf{p}(X_k = n) \,.$$

Fact. If the **transition matrix** is **primitive**, then it follows from a version of the Perron-Frobenius theorem [14] that:

- the equilibrium distribution $\boldsymbol{\pi}$ is **unique**
- $$\lim_{k \to \infty} \boldsymbol{P}^k = \boldsymbol{\pi} \mathbf{1}_N'$$

Thus, regardless of the initial distribution,

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

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

Example.

For our running example 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

julia> P^10
 0.286203  0.285714  0.285226
 0.428571  0.428571  0.428571
 0.285226  0.285714  0.286203
```

If transition matrix $\boldsymbol{P}$ is a **primitive matrix**, then $\lim_{k\to\infty} \boldsymbol{P}^k$ is a **projection matrix**. (?)

A: True B: False **??**

Irreducible matrix

Define. A square matrix $\boldsymbol{A}$ is called **irreducible** iff

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

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

There are other equivalent definitions of irreducible matrices.

Any **primitive** matrix is an **irreducible** matrix. (?)
A: True B: False **??**

The **circulant** generator matrix $\boldsymbol{G}_N$ is **irreducible**. (?)
A: True B: False **??**

Fact. If an nonnegative **irreducible** matrix $\boldsymbol{A}$ has at least one non-zero diagonal element, then $\boldsymbol{A}$ is **primitive** [15, p. 678].

Some of the Perron-Frobenius theorem statements hold for a nonnegative irreducible matrix, but not all. In particular, we lose uniqueness of the maximum eigenvalue magnitude.

Example. For the circulant generator matrix $\boldsymbol{G}_N$, the eigenvalue magnitudes are all $\left|\mathrm{e}^{-\imath 2\pi k/N}\right| = 1$.

That is why we focused on **primitive** matrices previously.

Matrix period

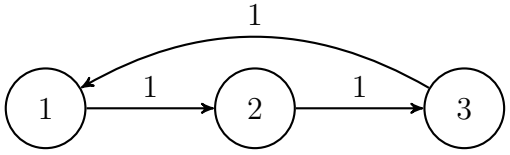
Define. For a square nonnegative matrix, the **period of the ith index** is the greatest common divisor of all $m \in \mathbb{N}$ such that $[\boldsymbol{A}^m]_{i,i} > 0$, if any such m exists.

Define. If the period of the ith index exists and is 1 for all i, then $\boldsymbol{A}$ is called **aperiodic**.

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

Any square **positive** matrix is **aperiodic**. (?)
A: True B: False **??**

Example. The transition matrix corresponding to a loop of 3 states is $\boldsymbol{P} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$.
This matrix is nonnegative and irreducible and it has period 3.
It has three eigenvalues whose magnitude is 1. Its (only) equilibrium distribution is $\begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}$.
However, $\boldsymbol{P}^m$ does not converge.



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

Properties of nonnegative irreducible matrices

The **Perron-Frobenius theorem** for any (square) nonnegative **irreducible matrix** $\boldsymbol{A}$ states that:

- All the conclusions for (square) nonnegative matrices hold.
- $r \triangleq \rho(\boldsymbol{A}) > 0$ is a (real) eigenvalue of $\boldsymbol{A}$.
  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**, *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 m denotes the period of $\boldsymbol{A}$, then there are exactly m eigenvalues for which $|\lambda| = r = \rho(\boldsymbol{A})$.
- There exists a unit-norm eigenvector $\boldsymbol{v}$ with eigenvalue r, *i.e.*, $\boldsymbol{A}\boldsymbol{v} = r\boldsymbol{v}$, where $\boldsymbol{v}$ has all *positive* elements.
- There likewise exists a unit-norm left eigenvector $\boldsymbol{w}$ having all positive elements for which $\boldsymbol{w}'\boldsymbol{A} = r\boldsymbol{w}'$.
  We call this $\boldsymbol{v}$ and $\boldsymbol{w}$ the right and left **Perron vectors**, respectively
- Uniqueness: All other (left and right) unit-norm eigenvectors have negative and/or non-real elements.
  So $\boldsymbol{v}$ and $\boldsymbol{w}$ are unique in that sense.

So in short, the property we "lose" in going from **primitive** matrices to the broader family of **irreducible matrices** is uniqueness of the magnitude of the Perron root. This prevents us from drawing conclusions using the power method.

Strongly connected graphs

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

Fact. A graph is **strongly connected** iff the corresponding (transition) matrix $\boldsymbol{A}$ is **irreducible**.

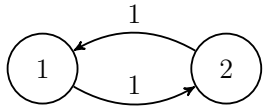
All (square) primitive matrices correspond to strongly connected graphs. (?)
A: True B: False **??**

Example. The running example with 3 states on p. 7.37 *is* strongly connected.

Example. The three-state loop on p. 7.41 *is* strongly connected.

Example. The graph corresponding to the 3×3 matrix (7.7) is *not* strongly connected.

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



The corresponding transition matrix $\boldsymbol{P} = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ is primitive. (?)

A: True B: False **??**

- 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 $\begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$.

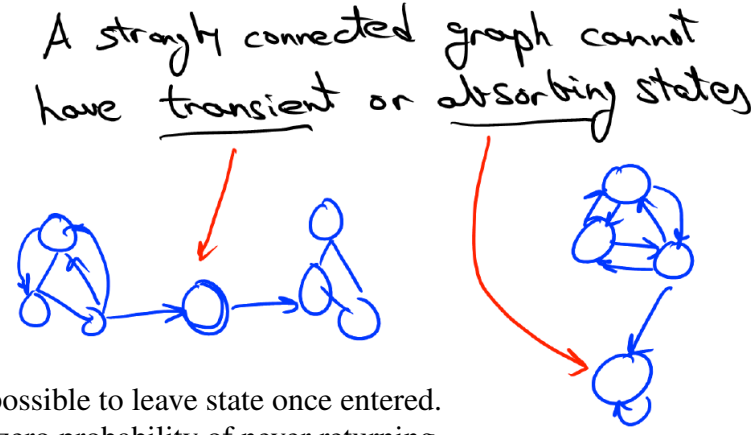
Let $\boldsymbol{\pi}^{(n)} \triangleq \boldsymbol{P}^{n-1}\boldsymbol{\pi}^{(1)}$ where $\boldsymbol{\pi}^{(1)} \triangleq \begin{bmatrix} \mathsf{P}\{X_1 = 1\} \\ \mathsf{P}\{X_1 = 2\} \end{bmatrix}$ denotes the starting probability distribution.

Does $\boldsymbol{\pi}^{(n)}$ approach the equilibrium distribution as $n \to \infty$?

A: Yes, for any $\boldsymbol{\pi}^{(1)} \geq \mathbf{0}$ s.t. $\mathbf{1}_2'\boldsymbol{\pi}^{(1)} = 1$.
B: Yes, but only for certain $\boldsymbol{\pi}^{(1)}$ choices.
C: No, never, because $\boldsymbol{P}^n$ oscillates between $\left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ and $\left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$.
D: None of these.

??

A strongly connected graph cannot have **transient** or **absorbing** states:



- Absorbing state: impossible to leave state once entered.
- Transient state: non-zero probability of never returning.

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.

Google's PageRank method

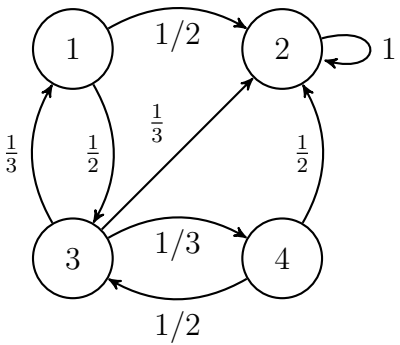
This section uses the preceding ideas to summarize **Google's PageRank** method [14]. [US Patent 6,285,999]
To rank web pages presented after a search, two natural criteria are

- 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.



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

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 $\boldsymbol{P}$ has columns that sum to one by construction, but the graph of web page outlinks is *not* **strongly connected**, so $\boldsymbol{P}$ 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 ctools.umich.edu from whitehouse.gov)

Thus there is not a unique equilibrium distribution.

One solution is to instead use a modified transition matrix:

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

This matrix $\tilde{\boldsymbol{P}}_\alpha$ is clearly **positive** by construction, so it has a unique equilibrium distribution $\boldsymbol{\pi}$ having all positive elements. This vector $\boldsymbol{\pi}$ has been called the $25 billion eigenvector [16].

PageRank algorithm

- Generate adjacency matrix by web crawling. It is (extremely) sparse, so store accordingly
- Form the transition matrix $\boldsymbol{P}$. 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 $\boldsymbol{\pi}_0$ 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:

$$\boldsymbol{\pi}_{k+1} = \tilde{\boldsymbol{P}}_\alpha \boldsymbol{\pi}_k = \alpha \boldsymbol{P} \boldsymbol{\pi}_k + \frac{1-\alpha}{N} \mathbf{1}_N.$$

- Perform 40-50 iterations (according to original paper)
- Use the final $\boldsymbol{\pi}$ vector (sorting?) to quantify the "importance" of web pages.

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

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 is in the algorithm is multiplying the $N \times N$ sparse matrix $\boldsymbol{P}$ by a (non-sparse) vector: $\boldsymbol{P}\boldsymbol{\pi}_k$.
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, $\boldsymbol{P}$ 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 $\boldsymbol{P}$ is sparse, with an average of, say, $L \ll N$ links per page, the product $\boldsymbol{P}\boldsymbol{\pi}_k$ requires "just" $O(NL)$ operations. This is 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 $\boldsymbol{\pi}_k$ must be stored as a 64-bit (double precision) vector. The number of bytes just to stored $\boldsymbol{\pi}$ 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.

7.5 Summary

This chapter has introduced just a few of the many important special matrices used in practice.

Another application of the Perron-Frobenius theorem is graph matching [17].

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	All	Equilibrium distribution(s) exist with eigenvalue 1.
Irreducible	Strongly connected	" Unique, positive equilibrium distribution $\boldsymbol{\pi}$
Primitive	Strongly connected (and more?)	" $\boldsymbol{P}^k \to \boldsymbol{\pi}\mathbf{1}'$
Positive	"Fully" connected	"

Other properties to consider when reviewing:

- $\rho \geq 0$ vs $\rho > 0$
- $\lambda_1 = \rho$ vs $|\lambda_1| = \rho$
- uniqueness of λ_1 (can any other $\lambda_i = \lambda_1$?)
- uniqueness of $|\lambda_1|$ (can any other $|\lambda_i| = |\lambda_1|$?)
- existence of $\boldsymbol{v}_1 \geq 0$ vs existence of $\boldsymbol{v}_1 > 0$

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

A: True B: False **??**

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

A: True B: False **??**

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 **??**

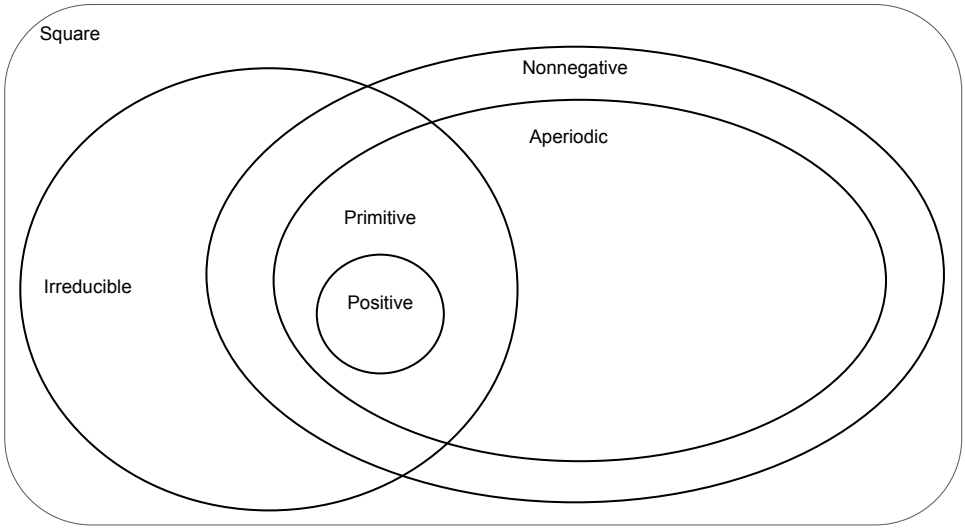
If a monic polynomial has all negative coefficients (except for the leading coefficient), then the corresponding companion matrix has an eigenvector with all positive elements. (?)

A: True B: False **??**

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.



An important category that is missing from this diagram is **transition** matrices.
Exercise. Complete the Venn diagram by adding **transition** matrices. (in class) **??**

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. 7.9).

[2] A. J. Laub. *Matrix analysis for scientists and engineers*. Soc. Indust. Appl. Math., 2005 (cit. on pp. 7.10, 7.11).

[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. 7.11).

[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. 7.17).

[5] H. Akaike. "Block Toeplitz matrix inversion". In: *SIAM J. Appl. Math.* 24.2 (Mar. 1973), 234–41 (cit. on p. 7.18).

[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. 7.18).

[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. 7.18).

[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. 7.18).

[9] I. Gohberg and G. Heinig. *Inversion of finite Toeplitz matrices*. 2010 (cit. on p. 7.18).

[10] V. Sukhoy and A. Stoytchev. "Generalizing the inverse FFT off the unit circle". In: *Nature: Scientific Reports* 9.1 (2019) (cit. on p. 7.18).

[11] 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. 7.19).

[12] 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. 7.25).

[13] 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. 7.25).

[14] 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. 7.26, 7.28, 7.30, 7.38, 7.46).

[15] C. D. Meyer. *Matrix analysis and applied linear algebra*. Soc. Indust. Appl. Math., 2000 (cit. on p. 7.40).

[16] 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. 7.47).

[17] S. Feizi, G. Quon, M. Recamonde-Mendoza, M. Medard, M. Kellis, and A. Jadbabaie. *Spectral alignment of graphs*. 2017 (cit. on p. 7.50).