

Non-negative matrices, digraphs and regular Markov chains

MATH 2740 – Mathematics of Data Science – Unit U08

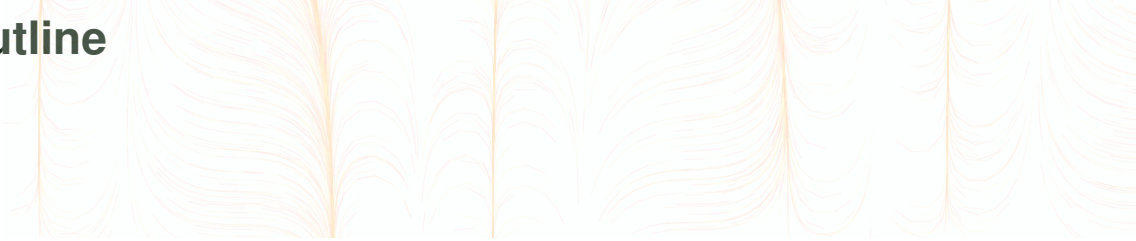
Julien Arino

`julien.arino@umanitoba.ca`

Department of Mathematics @ University of Manitoba

Fall 2026

Outline





Why study non-negative matrices and digraphs together?

Many matrices in data science have entries that represent quantities that cannot be negative:

- ▶ counts, weights, distances or similarity scores
- ▶ transition probabilities between states in a stochastic process
- ▶ hyperlinks between web pages or flows in a network

Such matrices have a rich geometric and spectral theory (Perron-Frobenius theory)

Their zero/non-zero sparsity pattern is naturally encoded by a **directed graph** (**digraph**), turning algebraic questions about matrix powers into visual questions about walks on a graph

Non-negative and positive vectors and matrices

Definition 107 (Non-negative and positive matrices)

Let $A = [a_{ij}] \in \mathcal{M}_{np}(\mathbb{R})$ be a real matrix

- ▶ A is **non-negative**, written $A \geq 0$ (or $A \in \mathcal{M}_{np}(\mathbb{R}_{\geq 0})$), if

$$a_{ij} \geq 0 \quad \text{for all } i = 1, \dots, n \text{ and } j = 1, \dots, p$$

- ▶ A is **positive** (or **strictly positive**), written $A > 0$, if

$$a_{ij} > 0 \quad \text{for all } i = 1, \dots, n \text{ and } j = 1, \dots, p$$

The same notation applies to vectors $\mathbf{x} = (x_1, \dots, x_n)^T \in \mathbb{R}^n$:

- ▶ $\mathbf{x} \geq \mathbf{0}$ means $x_i \geq 0$ for all $i = 1, \dots, n$
- ▶ $\mathbf{x} > \mathbf{0}$ means $x_i > 0$ for all $i = 1, \dots, n$

Componentwise inequalities and basic properties

For two matrices $A, B \in \mathcal{M}_{np}(\mathbb{R})$, we write $A \geq B$ if $A - B \geq 0$, and $A > B$ if $A - B > 0$

Proposition 108 (Arithmetic of non-negative matrices)

Let $A, B \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ and $\mathbf{x} \in \mathbb{R}_{\geq 0}^n$

1. $A + B \geq 0$ and $AB \geq 0$
2. For any integer $k \geq 1$, $A^k \geq 0$
3. If $A > 0$ and $\mathbf{x} \geq \mathbf{0}$ with $\mathbf{x} \neq \mathbf{0}$, then $A\mathbf{x} > \mathbf{0}$
4. If $A \geq 0$ and $\mathbf{x} > \mathbf{0}$, then $A\mathbf{x} > \mathbf{0}$ if and only if A has no zero row

Why does property 3 hold? Because $\mathbf{x} \neq \mathbf{0}$ has at least one entry $x_{j_0} > 0$, so for every row i ,

$$(A\mathbf{x})_i = \sum_{j=1}^n a_{ij}x_j \geq a_{i,j_0}x_{j_0} > 0$$

Directed graphs (digraphs)

Definition 109 (Directed graph)

A **directed graph** (or **digraph**) is a pair $G = (V, A)$ where:

- ▶ $V = \{1, 2, \dots, n\}$ is a finite set of elements called **vertices** (or **nodes**)
- ▶ $A \subseteq V \times V$ is a set of ordered pairs of vertices called **arcs** (or **directed edges**)

Definition 110 (Arc, tail, head and loop)

An **arc** $e = (u, v) \in A$ represents a directed connection from vertex u to vertex v

- ▶ u is the **initial endpoint** (or **tail**) and v is the **terminal endpoint** (or **head**)
- ▶ u is a **predecessor** of v and v is a **successor** of u
- ▶ An arc of the form $(u, u) \in A$ from a vertex to itself is called a **loop** (or **self-loop**)

Indegree, outdegree, sources and sinks

Let $G = (V, A)$ be a digraph and $v \in V$

Definition 111 (Outdegree and indegree)

- ▶ The **outdegree** $d_G^+(v)$ is the number of arcs directed away from v :

$$d_G^+(v) = |\{w \in V : (v, w) \in A\}|$$

- ▶ The **indegree** $d_G^-(v)$ is the number of arcs directed towards v :

$$d_G^-(v) = |\{u \in V : (u, v) \in A\}|$$

- ▶ A vertex with $d_G^-(v) = 0$ is a **source**; a vertex with $d_G^+(v) = 0$ is a **sink**

Since every arc has one tail and one head,

$$\sum_{v \in V} d_G^+(v) = \sum_{v \in V} d_G^-(v) = |A|$$

The digraph associated with a non-negative matrix

Definition 112 (Digraph of a matrix and adjacency matrix)

- ▶ Given a square non-negative matrix $M = [m_{ij}] \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$, its associated **digraph** is $\mathcal{G}(M) = (V, A)$ with vertex set $V = \{1, \dots, n\}$ and arc set

$$A = \{(i, j) \in V \times V : m_{ij} > 0\}$$

If we attach weight m_{ij} to each arc $(i, j) \in A$, we obtain a **weighted digraph** (or **directed network**)

- ▶ Conversely, given a digraph $G = (V, A)$ on $V = \{1, \dots, n\}$, its (binary) **adjacency matrix** is $A(G) = [a_{ij}] \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ defined by

$$a_{ij} = \begin{cases} 1 & \text{if } (i, j) \in A \\ 0 & \text{otherwise} \end{cases}$$

Note that diagonal entries $m_{ii} > 0$ correspond to self-loops (i, i) at vertex i

Orientation convention: general matrices vs Markov chains

There are two natural conventions for drawing arcs from a matrix $M = [m_{ij}]$:

1. **Row-to-column (adjacency) convention:** arc (i, j) goes from i to j when $m_{ij} > 0$
 - ▶ Row i lists outgoing arcs from i ; column j lists incoming arcs into j
2. **Column-to-row (transition) convention:** in column-stochastic Markov chains, $p_{ij} = \mathbb{P}(S_i(t+1) \mid S_j(t))$ is the probability of moving **from j to i** , so we draw an arc from j to i when $p_{ij} > 0$
 - ▶ Column j lists outgoing transitions from state j (summing to 1); row i lists incoming transitions into state i
 - ▶ This transition digraph is simply $\mathcal{G}(P^T)$, obtained by reversing all arrows of $\mathcal{G}(P)$

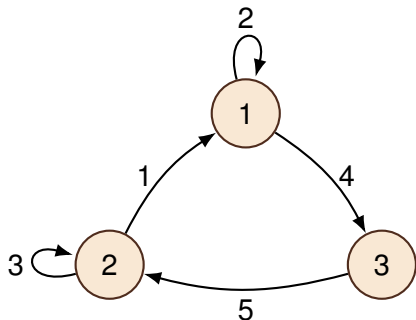
Crucial fact: Reversing all arrows preserves closed walks, cycle lengths and strong connectedness, so irreducibility and primitivity of M and M^T are identical

Example: 3×3 non-negative matrix and its digraph $\mathcal{G}(M)$

Consider the 3×3 non-negative matrix M and its binary sparsity pattern $A(\mathcal{G}(M))$:

$$M = \begin{pmatrix} 2 & 0 & 4 \\ 1 & 3 & 0 \\ 0 & 5 & 0 \end{pmatrix}, \quad A(\mathcal{G}(M)) = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Weighted digraph $\mathcal{G}(M)$



Reading $\mathcal{G}(M)$ directly from M :

- ▶ $m_{11} = 2 > 0 \implies$ self-loop $(1, 1)$
- ▶ $m_{13} = 4 > 0 \implies$ arc $(1, 3)$
- ▶ $m_{21} = 1 > 0 \implies$ arc $(2, 1)$
- ▶ $m_{22} = 3 > 0 \implies$ self-loop $(2, 2)$
- ▶ $m_{32} = 5 > 0 \implies$ arc $(3, 2)$
- ▶ All other entries are 0 (no other arcs)

Example: 4×4 sparse matrix and its digraph $\mathcal{G}(M)$

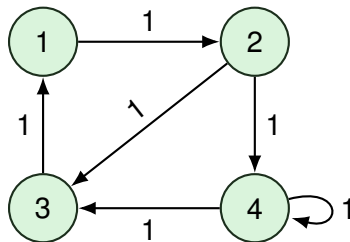
$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

- ▶ Row sums of $A(\mathcal{G}(M))$ give the outdegrees:

$$d^+ = (1, 2, 1, 2)^T$$

- ▶ Column sums of $A(\mathcal{G}(M))$ give the indegrees:

$$d^- = (1, 1, 2, 2)^T$$



Building a digraph from a matrix in igraph

```
M <- matrix(c(0, 1, 0, 0,
              0, 0, 1, 1,
              1, 0, 0, 0,
              0, 0, 1, 1), nrow = 4, byrow = TRUE)
G <- graph_from_adjacency_matrix(M, mode = "directed")
degree(G, mode = "out")

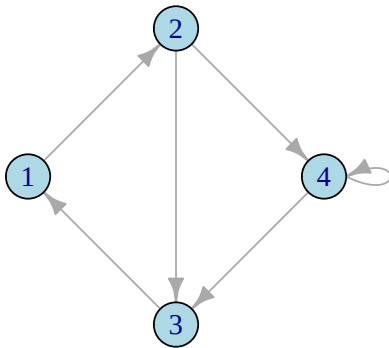
## [1] 1 2 1 2

degree(G, mode = "in")

## [1] 1 1 2 2

plot(G, edge.arrow.size = 0.7, vertex.color = "lightblue",
      vertex.size = 30, main = "Digraph G(M) from adjacency matrix M")
```

Digraph $G(M)$ from adjacency matrix M



Walks, trails, paths and cycles in a digraph

Let $G = (V, A)$ be a digraph

Definition 113 (Directed walk, length and closed walk)

- ▶ A **directed walk** from vertex v_0 to vertex v_k in G is a sequence of vertices $v_0, v_1, \dots, v_k$ such that $(v_{r-1}, v_r) \in A$ is an arc in G for all $r = 1, \dots, k$
- ▶ The **length** of the directed walk is k , the number of arcs traversed
- ▶ The walk is **closed** if its initial and terminal vertices coincide ($v_0 = v_k$)

Definition 114 (Directed trail, path and cycle)

- ▶ A **directed trail** is a directed walk in which all arcs are distinct
- ▶ A **directed path** is a directed walk in which all vertices $v_0, v_1, \dots, v_k$ are distinct
- ▶ A **directed cycle** (of length $k \geq 1$) is a closed walk $v_0, v_1, \dots, v_k = v_0$ in which $v_0, v_1, \dots, v_{k-1}$ are distinct (a cycle of length 1 is a self-loop)

Powers of a matrix and walks in its digraph

Theorem 115 (Matrix powers count walks)

Let $A = [a_{ij}] \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ and denote the (i, j) entry of A^k by $a_{ij}^{(k)}$ for $k \in \mathbb{N}_+$. Then

$$a_{ij}^{(k)} = \sum_{r_1=1}^n \sum_{r_2=1}^n \cdots \sum_{r_{k-1}=1}^n a_{i,r_1} a_{r_1,r_2} \cdots a_{r_{k-1},j}$$

In particular:

1. $a_{ij}^{(k)} > 0$ if and only if there is at least one directed walk of length k from i to j in $\mathcal{G}(A)$
2. If $A \in \{0, 1\}^{n \times n}$ is a binary adjacency matrix, then $a_{ij}^{(k)}$ equals the exact number of directed walks of length k from vertex i to vertex j in $\mathcal{G}(A)$

Proof of Theorem ??

We prove the formula by mathematical induction on the walk length $k \geq 1$

Base case ($k = 1$): $A^1 = A$, and $a_{ij}^{(1)} = a_{ij}$ is the weight of the single arc (i, j) (a walk of length 1 from i to j), which equals 1 if (i, j) is an arc in the binary case and 0 otherwise

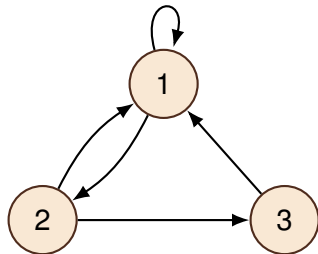
Inductive step: Suppose the result holds for $k \geq 1$. Since $A^{k+1} = A^k A$, the definition of matrix multiplication gives

$$a_{ij}^{(k+1)} = \sum_{r=1}^n a_{ir}^{(k)} a_{rj}$$

In the binary case $A \in \{0, 1\}^{n \times n}$, any walk of length $k + 1$ from i to j decomposes uniquely into a walk of length k from i to some penultimate vertex r , followed by the arc (r, j) . For a fixed r , there are $a_{ir}^{(k)} a_{rj}$ such walks. Summing over all $r \in \{1, \dots, n\}$ gives $a_{ij}^{(k+1)}$, completing the induction □

Example: counting walks of length 2 and 3

Consider the binary adjacency matrix A and its digraph $\mathcal{G}(A)$:



$$A = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}, \quad A^2 = \begin{pmatrix} 2 & 1 & 1 \\ 2 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

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

Let us verify individual entries directly from $\mathcal{G}(A)$:

- ▶ $(A^2)_{11} = 2$: the two walks of length 2 from 1 to 1 are $1 \rightarrow 1 \rightarrow 1$ and $1 \rightarrow 2 \rightarrow 1$
- ▶ $(A^2)_{23} = 0$: from 2, length-1 steps reach 1 or 3, neither of which has an arc to 3
- ▶ $(A^3)_{11} = 4$: the four walks of length 3 from 1 to 1 are

$$1 \rightarrow 1 \rightarrow 1 \rightarrow 1, \quad 1 \rightarrow 1 \rightarrow 2 \rightarrow 1, \quad 1 \rightarrow 2 \rightarrow 1 \rightarrow 1, \quad 1 \rightarrow 2 \rightarrow 3 \rightarrow 1$$



Strong connectedness and strongly connected components

Definition 116 (Strongly connected digraph)

A digraph $G = (V, A)$ is **strongly connected** if for every ordered pair of vertices $(i, j) \in V \times V$ with $i \neq j$, there exists a directed walk (and hence a directed path) in G from i to j and from j to i

Define the mutual reachability relation $i \equiv j$ on V by:

$$i \equiv j \iff i = j \text{ or there is a directed walk from } i \text{ to } j \text{ and from } j \text{ to } i$$

- ▶ $\equiv$ is reflexive, symmetric and transitive, so it is an **equivalence relation** on V
- ▶ Its equivalence classes partition V into maximal strongly connected subdigraphs called the **strongly connected components** (SCCs) of G
- ▶ G is strongly connected if and only if it has a single strongly connected component

Reducible and irreducible matrices

Definition 117 (Reducible and irreducible matrices)

A matrix $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ ($n \geq 2$) is **reducible** if there exists a permutation matrix $P_\pi \in \mathcal{M}_n$ such that

$$P_\pi^T A P_\pi = \begin{pmatrix} A_{11} & A_{12} \\ \mathbf{0} & A_{22} \end{pmatrix}$$

where $A_{11} \in \mathcal{M}_r$ and $A_{22} \in \mathcal{M}_{n-r}$ are square blocks with $1 \leq r < n$, and $\mathbf{0}$ is the $(n-r) \times r$ zero matrix

The matrix A is **irreducible** if no such permutation matrix exists (by convention, any 1×1 matrix is irreducible)

Note that $P_\pi^T A P_\pi$ simply relabels the rows and columns of A using the same permutation π of the vertices $\{1, \dots, n\}$

Why block upper-triangular form prevents strong connectedness

Suppose A has been relabelled into the reducible block form

$$\tilde{A} = P_{\pi}^T A P_{\pi} = \begin{pmatrix} A_{11} & A_{12} \\ \mathbf{0} & A_{22} \end{pmatrix} \quad \left. \begin{array}{l} \} V_1 = \{1, \dots, r\} \\ \} V_2 = \{r+1, \dots, n\} \end{array} \right\}$$

- ▶ The bottom-left block is $\mathbf{0}$, which means $\tilde{a}_{ij} = 0$ for all $i \in V_2$ and $j \in V_1$: there is **no arc** from any vertex in V_2 to any vertex in V_1
- ▶ Furthermore, for every power $k \geq 1$, block multiplication yields

$$\tilde{A}^k = \begin{pmatrix} A_{11}^k & * \\ \mathbf{0} & A_{22}^k \end{pmatrix}$$

so $(\tilde{A}^k)_{ij} = 0$ for all $i \in V_2$ and $j \in V_1$

- ▶ Thus no walk of any length $k \geq 1$ can ever go from V_2 to V_1 : the vertex set V_2 is a **closed trap** under the row-to-column walk dynamics

Characterisations of irreducibility

Theorem 118 (Characterisations of irreducible matrices)

For a non-negative matrix $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$, the following statements are equivalent:

- 1. A is irreducible*
- 2. The associated digraph $\mathcal{G}(A)$ is strongly connected*
- 3. $\mathbb{I}_n + A + A^2 + \cdots + A^{n-1} > 0$ (strictly positive entrywise)*
- 4. $(\mathbb{I}_n + A)^{n-1} > 0$ (strictly positive entrywise)*

Items 3 and 4 provide a purely algebraic test for irreducibility without searching over all $n!$ permutation matrices

Proof of Theorem ??

(1) $\iff$ (2): A is reducible $\iff$ vertices can be partitioned into non-empty sets V_1, V_2 with no arc from V_2 to V_1 $\iff$ no walk can go from V_2 to V_1 $\iff \mathcal{G}(A)$ is not strongly connected

(2) $\iff$ (3): By Theorem ??, $(A^k)_{ij} > 0 \iff$ there is a walk of length k from i to j . Any shortest walk between two distinct vertices $i \neq j$ is a simple path visiting at most n vertices, so its length is $k \in \{1, \dots, n-1\}$. With $(\mathbb{I}_n)_{ii} = 1 > 0$ covering $i = j$, $\mathcal{G}(A)$ is strongly connected $\iff$ for every pair (i, j) there is some $k \in \{0, 1, \dots, n-1\}$ with $(A^k)_{ij} > 0 \iff \sum_{k=0}^{n-1} A^k > 0$

(3) $\iff$ (4): By the binomial expansion with positive coefficients $\binom{n-1}{k} > 0$ and $A^k \geq 0$,

$$(\mathbb{I}_n + A)^{n-1} = \sum_{k=0}^{n-1} \binom{n-1}{k} A^k$$

A sum of non-negative terms with positive weights is strictly positive if and only if at least one term is strictly positive, so $(\mathbb{I}_n + A)^{n-1} > 0 \iff \sum_{k=0}^{n-1} A^k > 0$ □

Exam-style example 1: testing irreducibility via $\mathcal{G}(A)$

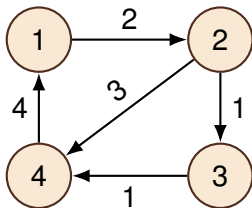
Problem: Determine whether the following 4×4 matrix is irreducible:

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

Solution:

- ▶ Draw the 4-node digraph $\mathcal{G}(A)$
- ▶ Observe the directed Hamiltonian cycle

$$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$$

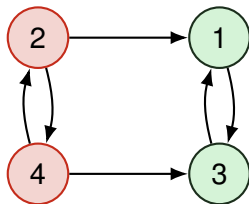


- ▶ Since this cycle visits all 4 vertices, every vertex can reach every other vertex along the cycle
- ▶ Therefore $\mathcal{G}(A)$ is **strongly connected**, so A is **irreducible**

Exam-style example 2: reducible matrix and block triangular form

Problem: Show that B is reducible, find its SCCs and write $P_{\pi}^T B P_{\pi}$ in block upper-triangular form:

$$B = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{pmatrix}$$



- ▶ SCCs are $C_1 = \{2, 4\}$ and $C_2 = \{1, 3\}$
- ▶ Arcs go from C_1 to C_2 , but **no arc goes from C_2 to C_1** , so $\mathcal{G}(B)$ is not strongly connected (B is **reducible**)
- ▶ Ordering vertices as $(2, 4, 1, 3)$ gives

$$P_{\pi}^T B P_{\pi} = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{array} \right) = \begin{pmatrix} B_{11} & B_{12} \\ \mathbf{0} & B_{22} \end{pmatrix}$$

Checking strong connectedness and $(\mathbb{I}_n + A)^{n-1}$ in R

```
A <- matrix(c(0, 2, 0, 0,
              0, 0, 1, 3,
              0, 0, 0, 1,
              4, 0, 0, 0), nrow = 4, byrow = TRUE)
is_connected(graph_from_adjacency_matrix(A > 0), mode = "strong")

## [1] TRUE

# Compute  $(I_4 + A)^3$ 
I4_plus_A <- diag(4) + A
I4_plus_A %*% I4_plus_A %*% I4_plus_A

##      [,1] [,2] [,3] [,4]
## [1,]  25   6   6  20
## [2,]  40  25   3  12
## [3,]  12   8   1   3
## [4,]  12  24   8  25
```



Primitive matrices

Definition 119 (Primitive matrix)

A non-negative matrix $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ is **primitive** if there exists a positive integer $m \in \mathbb{N}_+$ such that

$$A^m > 0 \quad (\text{every entry of } A^m \text{ is strictly positive})$$

The smallest positive integer m such that $A^m > 0$ is called the **index of primitivity** of A , denoted $\gamma(A)$

An irreducible non-negative matrix that is not primitive is called **imprimitive** (or **periodic**)

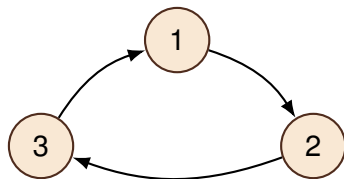
Proposition 120

1. Every primitive matrix is irreducible
2. If $A^m > 0$ for some $m \in \mathbb{N}_+$, then $A^k > 0$ for all $k \geq m$

Irreducible does not imply primitive!

Consider the 2×2 and 3×3 cyclic permutation matrices:

$$C_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad C_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$



- ▶ Both $\mathcal{G}(C_2)$ and $\mathcal{G}(C_3)$ are single directed cycles, so C_2 and C_3 are **irreducible**
- ▶ Yet for C_2 : $C_2^{2k} = \mathbb{I}_2$ and $C_2^{2k+1} = C_2$
- ▶ And for C_3 : $C_3^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$, $C_3^3 = \mathbb{I}_3$,
 $C_3^4 = C_3, \dots$
- ▶ Every power C_2^m and C_3^m contains zeros, so neither matrix is primitive!

Why? Because walks from i to j only occur at lengths m belonging to a fixed congruence class modulo 2 (for C_2) or modulo 3 (for C_3)

Period of a digraph and characterisation of primitivity

Definition 121 (Period and aperiodicity)

Let $G = (V, A)$ be a strongly connected digraph. The **period** (or **index of imprimitivity**) $d(G)$ of G is the greatest common divisor of the lengths of all directed cycles (or closed walks) in G :

$$d(G) = \gcd\{\text{length}(C) : C \text{ is a directed cycle in } G\}$$

- ▶ If $d(G) = 1$, we say that G (and its matrix A) is **aperiodic**
- ▶ If $d(G) > 1$, we say that G (and its matrix A) is **periodic** with period $d(G)$

Theorem 122 (Graphical characterisation of primitive matrices)

A non-negative matrix $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ is **primitive** if and only if $\mathcal{G}(A)$ is **strongly connected** and **aperiodic** ($d(\mathcal{G}(A)) = 1$)

Practical corollaries for exams

Checking gcd of cycle lengths is immediate in two common situations:

Corollary 123 (Self-loop test for primitivity)

*If $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ is irreducible ($\mathcal{G}(A)$ is strongly connected) and has **at least one positive diagonal entry** $a_{ii} > 0$ for some $i \in \{1, \dots, n\}$, then A is **primitive***

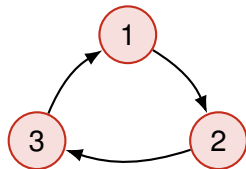
Reason: $a_{ii} > 0$ gives a self-loop $i \rightarrow i$ of length 1 in $\mathcal{G}(A)$, and $\gcd(1, \dots) = 1$

Corollary 124 (Coprime cycle lengths)

*If $\mathcal{G}(A)$ is strongly connected and contains two directed cycles of coprime lengths ℓ_1 and ℓ_2 (i.e. $\gcd(\ell_1, \ell_2) = 1$, such as a 2-cycle and a 3-cycle), then A is **primitive***

Visualising periodic vs primitive digraphs

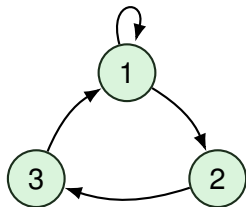
Case 1: Irreducible, periodic ($d = 3$)



$$A_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Cycle length: $\{3\}$
 $\gcd(3) = 3 \implies$ **not primitive**

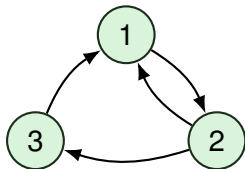
Case 2: Adding one self-loop ($d = 1$)



$$A_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Cycle lengths: $\{1, 3\}$
 $\gcd(1, 3) = 1 \implies$ **primitive** ($A^4 > 0$)

Case 3: Zero trace, coprime cycles ($d = 1$)



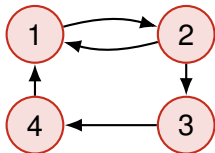
$$A_3 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Cycles: $1 \leftrightarrow 2$ (2),
 $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ (3)
 $\gcd(2, 3) = 1 \implies$ **primitive** ($A_3^5 > 0$)

Exam-style comparison: two 4×4 irreducible matrices

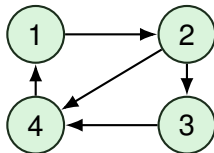
Both matrices below are irreducible with zero diagonal ($\text{tr} = 0$). Which one is primitive?

$$M_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$



- Directed cycles: $1 \rightarrow 2 \rightarrow 1$ (length 2) and $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ (length 4)

$$M_2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$



- Directed cycles: $1 \rightarrow 2 \rightarrow 4 \rightarrow 1$ (length 3) and $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ (length 4)
- $d = \gcd(3, 4) = 1$: **primitive**

Wielandt's bound on the index of primitivity

How large can the smallest exponent $\gamma(A)$ with $A^{\gamma(A)} > 0$ be for an $n \times n$ primitive matrix?

Theorem 125 (Wielandt's theorem)

Let $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ with $n \geq 2$. Then A is primitive if and only if

$$A^{n^2-2n+2} > 0$$

Moreover, if A is irreducible and has $d \geq 1$ positive diagonal entries, then

$$\gamma(A) \leq 2n - d - 1$$

- ▶ For $n = 3$, $n^2 - 2n + 2 = 5$: the matrix A_3 has $A_3^4 \not> 0$ and $A_3^5 > 0$
- ▶ For $n = 4$, $n^2 - 2n + 2 = 10$: the matrix M_2 on the previous slide has $M_2^9 \not> 0$ and $M_2^{10} > 0$, showing that Wielandt's bound $n^2 - 2n + 2$ is sharp!

Spectrum and spectral radius

Definition 126 (Spectrum and spectral radius)

Let $A \in \mathcal{M}_n(\mathbb{C})$ be a square matrix

- ▶ The **spectrum** of A , denoted $\sigma(A)$, is the set of all eigenvalues of A :

$$\sigma(A) = \{\lambda \in \mathbb{C} : \det(A - \lambda \mathbb{I}_n) = 0\}$$

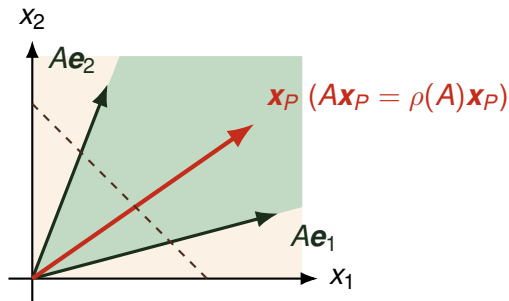
- ▶ The **spectral radius** of A , denoted $\rho(A)$, is the maximum modulus of its eigenvalues:

$$\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$$

Geometrically, in the complex plane $\mathbb{C}$, all eigenvalues of A lie inside or on the closed disc $\{z \in \mathbb{C} : |z| \leq \rho(A)\}$ of radius $\rho(A)$ centred at the origin

Geometric intuition in $\mathbb{R}^2$: trapping an invariant ray

Why must a positive matrix $A > 0$ have a positive eigenvalue with a positive eigenvector?



- ▶ Consider the non-negative quadrant $\mathbb{R}_{\geq 0}^2 = \{\mathbf{x} \geq \mathbf{0}\}$
- ▶ Since $A > 0$, both columns $A\mathbf{e}_1 > \mathbf{0}$ and $A\mathbf{e}_2 > \mathbf{0}$ lie **strictly inside** the open first quadrant
- ▶ Thus A maps the entire quadrant $\mathbb{R}_{\geq 0}^2 \setminus \{\mathbf{0}\}$ into a narrower cone strictly inside the first quadrant
- ▶ By Brouwer's fixed-point theorem (or the intermediate value theorem on the angle of $\mathbf{x}$), A must map a unique ray in the first quadrant onto itself!

Perron's theorem for positive matrices

Theorem 127 (Perron's theorem, 1907)

Let $A \in \mathcal{M}_n(\mathbb{R})$ be a strictly positive matrix ($A > 0$). Then:

1. $\rho(A) > 0$ is an eigenvalue of A (the **Perron root** or **dominant eigenvalue**)
2. $\rho(A)$ is algebraically and geometrically simple (multiplicity 1)
3. $\rho(A)$ strictly dominates every other eigenvalue: $|\lambda| < \rho(A)$ for all $\lambda \in \sigma(A) \setminus \{\rho(A)\}$
4. There exists a **unique** strictly positive right eigenvector $\mathbf{x} > \mathbf{0}$ such that

$$A\mathbf{x} = \rho(A)\mathbf{x} \quad \text{and} \quad \mathbf{1}^T \mathbf{x} = \sum_{i=1}^n x_i = 1$$

(the **right Perron vector**, and similarly a unique left Perron vector $\mathbf{y} > \mathbf{0}$)

5. No other eigenvalue of A has a non-negative eigenvector

The Perron-Frobenius theorem for primitive and irreducible matrices

Frobenius (1912) extended Perron's theorem to irreducible non-negative matrices:

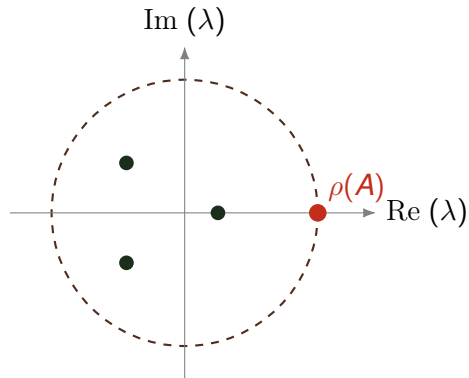
Theorem 128 (Perron-Frobenius theorem)

Let $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ be an **irreducible** non-negative matrix

1. $\rho(A) > 0$ is a simple eigenvalue of A (algebraic and geometric multiplicity 1)
2. There exists a unique strictly positive right eigenvector $\mathbf{x} > \mathbf{0}$ with $A\mathbf{x} = \rho(A)\mathbf{x}$ and $\mathbf{1}^T \mathbf{x} = 1$ (and no other eigenvalue has a non-negative eigenvector)
3. **If A is primitive** ($d = 1$): $|\lambda| < \rho(A)$ for all $\lambda \in \sigma(A)$ with $\lambda \neq \rho(A)$
4. **If A is imprimitive with period $d > 1$** : A has exactly d eigenvalues of modulus $\rho(A)$, given by $\lambda_k = \rho(A)e^{2\pi i k/d}$ ($k = 0, 1, \dots, d-1$), and $|\lambda| < \rho(A)$ for all others

Spectral circle: primitive ($d = 1$) vs periodic ($d = 3$)

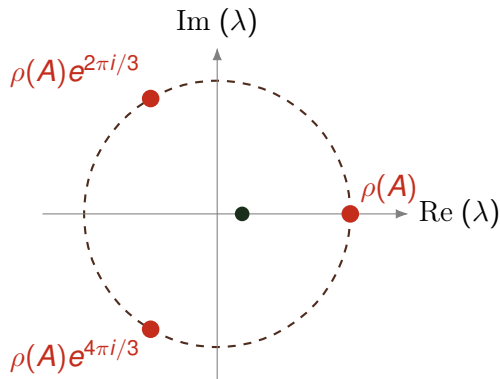
Primitive matrix ($d = 1$)



Strict spectral gap $|\lambda_j| < \rho(A)$ for $j \geq 2$

$\implies \left(\frac{1}{\rho(A)}A\right)^k$ **converges** as $k \rightarrow \infty$

Irreducible periodic matrix ($d = 3$)

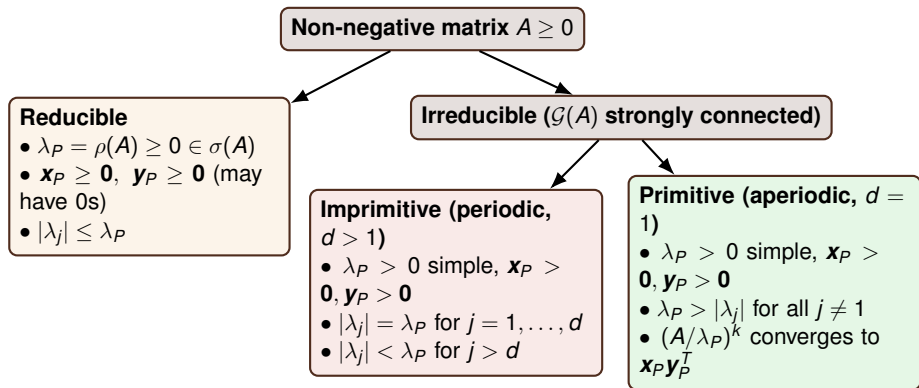


$d = 3$ eigenvalues on the circle

$|\lambda| = \rho(A)$
 $\implies \left(\frac{1}{\rho(A)}A\right)^k$ **oscillates** with period 3

Summary of Perron-Frobenius across all non-negative matrices

Let $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ have Perron root $\lambda_P = \rho(A)$ and right/left Perron vectors $\mathbf{x}_P, \mathbf{y}_P$:



Asymptotic behaviour of powers of a primitive matrix

Theorem 129 (Asymptotics of primitive matrices)

Let $A \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ be a primitive matrix with Perron root $\rho(A) > 0$. Let $\mathbf{x} > \mathbf{0}$ and $\mathbf{y} > \mathbf{0}$ be the unique positive right and left eigenvectors satisfying $A\mathbf{x} = \rho(A)\mathbf{x}$, $\mathbf{y}^T A = \rho(A)\mathbf{y}^T$, $\mathbf{1}^T \mathbf{x} = 1$ and $\mathbf{y}^T \mathbf{x} = 1$. Then

$$\lim_{k \rightarrow \infty} \left(\frac{1}{\rho(A)} A \right)^k = \mathbf{x} \mathbf{y}^T > 0$$

Why this holds: Write $\mathbf{z} = (\mathbf{y}^T \mathbf{z}) \mathbf{x} + \sum_{j=2}^n c_j \mathbf{v}_j$. Since $|\lambda_j / \rho(A)| < 1$ for all $j \geq 2$,

$$\left(\frac{1}{\rho(A)} A \right)^k \mathbf{z} = (\mathbf{y}^T \mathbf{z}) \mathbf{x} + \sum_{j=2}^n c_j \left(\frac{\lambda_j}{\rho(A)} \right)^k \mathbf{v}_j \xrightarrow{k \rightarrow \infty} (\mathbf{y}^T \mathbf{z}) \mathbf{x} = (\mathbf{x} \mathbf{y}^T) \mathbf{z}$$

Inspecting the Perron-Frobenius spectrum in R

Compare the eigenvalues of the periodic matrix A_1 ($d = 3$) and primitive matrix A_2 ($d = 1$):

```
A1 <- matrix(c(0, 1, 0, 0, 0, 1, 1, 0, 0), nrow = 3, byrow = TRUE)
A2 <- matrix(c(1, 1, 0, 0, 0, 1, 1, 0, 0), nrow = 3, byrow = TRUE)
# Moduli of eigenvalues of A1 (all 3 have modulus 1)
round(Mod(eigen(A1)$values), 4)

## [1] 1 1 1

# Eigenvalues and moduli for primitive A2 (unique dominant Perron root)
round(eigen(A2)$values, 4)

## [1] 1.4656+0.0000i -0.2328+0.7926i -0.2328-0.7926i

round(Mod(eigen(A2)$values), 4)

## [1] 1.4656 0.8260 0.8260
```





Discrete-time Markov chains

A **Markov chain** is a discrete-time stochastic process whose evolution depends **only on the current state** of the system (the **memoryless** or **Markov property**):

- ▶ Consider a system that can be in one of n finite states $S = \{S_1, \dots, S_n\}$ at discrete times $t = 0, 1, 2, \dots$
- ▶ The probability of moving to state S_i at time $t + 1$ given that the system is in state S_j at time t is

$$p_{ij} = \mathbb{P}(S_i(t+1) \mid S_j(t))$$

- ▶ Note the index order carefully: p_{ij} is the transition probability **from state S_j (column j) to state S_i (row i)**

Markov chains are a cornerstone of data science: they underpin PageRank, Hidden Markov Models (HMMs), Markov Chain Monte Carlo (MCMC) sampling and reinforcement learning

The transition matrix P

Definition 130 (Discrete-time Markov chain and transition matrix)

A sequence of random states from $S = \{S_1, \dots, S_n\}$ is a (time-homogeneous) **discrete-time Markov chain** (DTMC) if for all $t = 0, 1, 2, \dots$ and all $1 \leq i, j \leq n$, the transition probability

$$p_{ij} = \mathbb{P}(S_i(t+1) \mid S_j(t))$$

depends only on i and j . The matrix $P = [p_{ij}] \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ is called the **transition matrix** of the Markov chain

If the system is in state S_j at time t , it must transition to one of $S_1, \dots, S_n$ at time $t+1$, so the entries in **column** j of P must sum to 1:

$$p_{1j} + p_{2j} + \dots + p_{nj} = \sum_{i=1}^n p_{ij} = 1 \quad \text{for all } j = 1, \dots, n$$

How do we get to state S_i at time $t + 1$?

For $1 \leq i \leq n$, let $p_i(t) = \mathbb{P}(S_i(t))$ be the probability that the system is in state S_i at time t

List all n mutually exclusive ways to arrive in state S_i at time $t + 1$:

1. We were in S_1 at time t (probability $p_1(t)$) and transitioned from S_1 to S_i (probability p_{i1}), contributing $p_{i1}p_1(t)$
2. We were in S_2 at time t (probability $p_2(t)$) and transitioned from S_2 to S_i (probability p_{i2}), contributing $p_{i2}p_2(t)$
- $\vdots$
- n . We were in S_n at time t (probability $p_n(t)$) and transitioned from S_n to S_i (probability p_{in}), contributing $p_{in}p_n(t)$

Summing over all $j = 1, \dots, n$ gives the law of total probability:

$$p_i(t + 1) = \sum_{j=1}^n p_{ij}p_j(t) = p_{i1}p_1(t) + p_{i2}p_2(t) + \dots + p_{in}p_n(t)$$

State distribution update in matrix form

Collecting the state probabilities into the column **probability vector**

$$\mathbf{p}(t) = \begin{pmatrix} p_1(t) \\ \vdots \\ p_n(t) \end{pmatrix} \geq \mathbf{0}, \quad \mathbf{1}^T \mathbf{p}(t) = \sum_{i=1}^n p_i(t) = 1$$

the update equations for $i = 1, \dots, n$ take the compact matrix form

$$\begin{pmatrix} p_1(t+1) \\ \vdots \\ p_n(t+1) \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{pmatrix} \begin{pmatrix} p_1(t) \\ \vdots \\ p_n(t) \end{pmatrix} \iff \mathbf{p}(t+1) = P\mathbf{p}(t)$$

Iterating from the initial distribution $\mathbf{p}(0)$ gives $\mathbf{p}(1) = P\mathbf{p}(0)$,
 $\mathbf{p}(2) = P\mathbf{p}(1) = P^2\mathbf{p}(0)$ and for any $t \in \mathbb{N}$:

$$\mathbf{p}(t) = P^t \mathbf{p}(0)$$

Transition matrix and transition digraph

For a column-stochastic transition matrix $P = [p_{ij}]$, recall that $p_{ij} > 0$ represents a transition **from state S_j to state S_i** :

Column-stochastic matrix P

$$P = \begin{pmatrix} 0.5 & 0.8 & 0.1 \\ 0.5 & 0 & 0.2 \\ 0 & 0.2 & 0.7 \end{pmatrix}$$

► Column 1 (from S_1):

$$0.5 + 0.5 + 0 = 1$$

► Column 2 (from S_2):

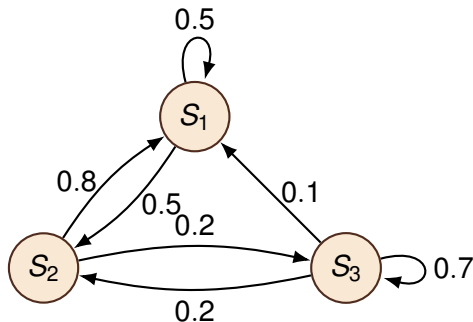
$$0.8 + 0 + 0.2 = 1$$

► Column 3 (from S_3):

$$0.1 + 0.2 + 0.7 = 1$$

Out-weights from each vertex S_j in the transition digraph correspond to **column j** of P and always sum to 1

Transition digraph



Stochastic matrices

Definition 131 (Column-stochastic, row-stochastic and doubly stochastic matrices)

Let $M = [m_{ij}] \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ be a non-negative matrix and $\mathbf{1} = (1, \dots, 1)^T \in \mathbb{R}^n$

- ▶ M is **column-stochastic** if every column sums to 1:

$$\sum_{i=1}^n m_{ij} = 1 \quad \text{for all } j = 1, \dots, n \quad \Longleftrightarrow \quad \mathbf{1}^T M = \mathbf{1}^T$$

- ▶ M is **row-stochastic** if every row sums to 1:

$$\sum_{j=1}^n m_{ij} = 1 \quad \text{for all } i = 1, \dots, n \quad \Longleftrightarrow \quad M \mathbf{1} = \mathbf{1}$$

- ▶ M is **doubly stochastic** if it is both column-stochastic and row-stochastic ($\mathbf{1}^T M = \mathbf{1}^T$ and $M \mathbf{1} = \mathbf{1}$)

Products and powers of column-stochastic matrices

Theorem 132

1. If $M, N \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ are column-stochastic matrices, then their product MN is also column-stochastic
2. For any column-stochastic matrix $P \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ and any integer $t \geq 1$, P^t is column-stochastic, and if $\mathbf{p}(0)$ is a probability vector, then $\mathbf{p}(t) = P^t \mathbf{p}(0)$ is a probability vector

Proof:

1. Since $M \geq 0$ and $N \geq 0$, their product satisfies $MN \geq 0$. Furthermore, using the associativity of matrix multiplication and $\mathbb{1}^T M = \mathbb{1}^T$, $\mathbb{1}^T N = \mathbb{1}^T$:

$$\mathbb{1}^T (MN) = (\mathbb{1}^T M)N = \mathbb{1}^T N = \mathbb{1}^T$$

Thus MN is non-negative and its columns sum to 1, so MN is column-stochastic

2. By induction on $t \geq 1$, P^t is column-stochastic, and $\mathbf{p}(t) = P^t \mathbf{p}(0) \geq \mathbf{0}$ with $\mathbb{1}^T \mathbf{p}(t) = (\mathbb{1}^T P^t) \mathbf{p}(0) = \mathbb{1}^T \mathbf{p}(0) = 1$



Spectral radius of a column-stochastic matrix

Theorem 133 (Spectrum of a stochastic matrix)

Let $P \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ be a column-stochastic matrix. Then:

1. $\lambda_1 = 1$ is an eigenvalue of P
2. Every eigenvalue $\lambda \in \sigma(P)$ satisfies $|\lambda| \leq 1$, so the spectral radius is $\rho(P) = 1$

Proof:

1. Since P is column-stochastic, $\mathbb{1}^T P = \mathbb{1}^T$. Taking the transpose gives $P^T \mathbb{1} = \mathbb{1}$ with $\mathbb{1} \neq \mathbf{0}$, so $\lambda = 1$ is an eigenvalue of P^T . Since $\det(P^T - \lambda \mathbb{I}_n) = \det(P - \lambda \mathbb{I}_n)$, P and P^T have the same eigenvalues ($\sigma(P) = \sigma(P^T)$), so $1 \in \sigma(P)$
2. Let $\lambda \in \sigma(P) = \sigma(P^T)$ and let $\mathbf{x} \neq \mathbf{0}$ be a right eigenvector of P^T : $P^T \mathbf{x} = \lambda \mathbf{x}$. Pick an index k maximising $|x_k| = \max_{1 \leq j \leq n} |x_j| > 0$. Then from row k of $P^T \mathbf{x} = \lambda \mathbf{x}$:

$$|\lambda| |x_k| = |\lambda x_k| = \left| \sum_{i=1}^n p_{ik} x_i \right| \leq \sum_{i=1}^n p_{ik} |x_i| \leq \left(\sum_{i=1}^n p_{ik} \right) |x_k| = 1 \cdot |x_k|$$

Regular Markov chains

Definition 134 (Regular Markov chain)

A discrete-time Markov chain with column-stochastic transition matrix $P \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ is **regular** if P is **primitive**, i.e. there exists a positive integer $k \geq 1$ such that

$$P^k > 0 \quad (\text{every entry of } P^k \text{ is strictly positive})$$

Interpretation:

- ▶ The entry $(P^k)_{ij}$ is the k -step transition probability of moving from state S_j to state S_i in exactly k steps
- ▶ Thus a Markov chain is regular if and only if there is some time horizon $k \geq 1$ such that from **any** starting state S_j , there is a strictly positive probability of being in **any** target state S_i after k steps
- ▶ By Theorem ??, a Markov chain is regular $\iff$ its transition digraph is **strongly connected** (irreducible) and **aperiodic** ($d = 1$, e.g. at least one state has a self-loop $p_{ii} > 0$)

The fundamental limit theorem for regular Markov chains

Applying Theorem ?? with $\rho(P) = 1$ and left Perron vector $\mathbf{y} = \mathbf{1}$ yields:

Theorem 135 (Fundamental limit theorem for regular Markov chains)

Let $P \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ be the column-stochastic transition matrix of a **regular** Markov chain. Then:

1. There exists a **unique** strictly positive probability vector $\mathbf{v} = (v_1, \dots, v_n)^T > \mathbf{0}$ satisfying $P\mathbf{v} = \mathbf{v}$ and $\mathbf{1}^T \mathbf{v} = \sum_{i=1}^n v_i = 1$, called the **stationary distribution** (or **equilibrium** / **steady-state distribution**) of the Markov chain
2. As $t \rightarrow \infty$, the matrix powers P^t converge to the rank-one column-stochastic matrix

$$\lim_{t \rightarrow \infty} P^t = W = (\mathbf{v} \quad \mathbf{v} \quad \cdots \quad \mathbf{v}) = \mathbf{v} \mathbf{1}^T > \mathbf{0}$$

3. For **any** initial probability distribution $\mathbf{p}(0)$, the state vector $\mathbf{p}(t) = P^t \mathbf{p}(0)$ converges to $\mathbf{v}$:

$$\lim_{t \rightarrow \infty} \mathbf{p}(t) = W \mathbf{p}(0) = \mathbf{v} (\mathbf{1}^T \mathbf{p}(0)) = \mathbf{v}$$

Finding the stationary distribution $\mathbf{v}$ in practice

How do we compute the unique stationary distribution $\mathbf{v}$ of a regular transition matrix $P \in \mathcal{M}_n(\mathbb{R}_{\geq 0})$ on an exam or by hand?

1. **Write the fixed-point equation $P\mathbf{v} = \mathbf{v}$ as a homogeneous linear system:**

$$(P - \mathbb{I}_n)\mathbf{v} = \mathbf{0} \quad (\text{so } \mathbf{v} \in \text{null}(P - \mathbb{I}_n))$$

- ▶ Since $\lambda_1 = 1$ is a simple eigenvalue of P , $\dim \text{null}(P - \mathbb{I}_n) = 1$
- ▶ Because $\mathbf{1}^T(P - \mathbb{I}_n) = \mathbf{0}^T$, the rows of $P - \mathbb{I}_n$ always sum to zero (one row is redundant)

2. **Row-reduce $P - \mathbb{I}_n$** (or solve the first $n - 1$ equations) to express $v_1, \dots, v_n$ in terms of a single free parameter, giving a positive basis vector $\mathbf{w} \in \text{null}(P - \mathbb{I}_n)$

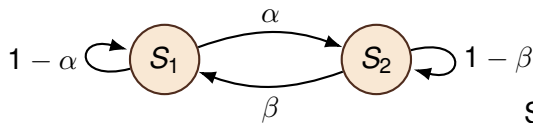
3. **Normalise to a probability vector** using $\mathbf{1}^T \mathbf{v} = v_1 + \dots + v_n = 1$:

$$\mathbf{v} = \frac{\mathbf{w}}{\|\mathbf{w}\|_1} = \frac{\mathbf{w}}{w_1 + \dots + w_n}$$



Exam-style example 1: general 2×2 regular Markov chain

Consider a two-state Markov chain with transition probabilities $\alpha = \mathbb{P}(S_2 \mid S_1)$ and $\beta = \mathbb{P}(S_1 \mid S_2)$, where $0 < \alpha, \beta < 1$:



$$P = \begin{pmatrix} 1 - \alpha & \beta \\ \alpha & 1 - \beta \end{pmatrix} > 0$$

Since $P > 0$, the chain is regular ($P^1 > 0$)

Finding the stationary distribution $\mathbf{v} = (v_1, v_2)^T$:

$$(P - \mathbb{I}_2)\mathbf{v} = \begin{pmatrix} -\alpha & \beta \\ \alpha & -\beta \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \iff -\alpha v_1 + \beta v_2 = 0 \iff v_2 = \frac{\alpha}{\beta} v_1$$

Imposing $v_1 + v_2 = 1$ gives $v_1 \left(1 + \frac{\alpha}{\beta}\right) = 1$, hence:

$$\mathbf{v} = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta \\ \alpha \end{pmatrix}, \quad \lim_{t \rightarrow \infty} P^t = W = \frac{1}{\alpha + \beta} \begin{pmatrix} \beta & \beta \\ \alpha & \alpha \end{pmatrix}$$

Exam-style example 2: 3×3 regular chain via row reduction

Problem: Show that the Markov chain with transition matrix P is regular and find its stationary distribution $\mathbf{v}$ and $\lim_{t \rightarrow \infty} P^t$:

$$P = \begin{pmatrix} 0.5 & 0.4 & 0 \\ 0.5 & 0.2 & 0.5 \\ 0 & 0.4 & 0.5 \end{pmatrix}$$

Step 1 (Regularity):

- ▶ From P , we have transitions $S_1 \rightarrow S_2 \rightarrow S_3 \rightarrow S_2 \rightarrow S_1$, so all 3 states communicate ($\mathcal{G}(P^T)$ is strongly connected, P is irreducible)
- ▶ Diagonal entries $p_{11} = 0.5 > 0$, $p_{22} = 0.2 > 0$, $p_{33} = 0.5 > 0$ give self-loops, so by Corollary ??, P is **primitive** (in fact $P^2 > 0$) and the chain is **regular**

Exam-style example 2 (continued): solving $(P - \mathbb{I}_3)\mathbf{v} = \mathbf{0}$

Step 2 (Nullspace of $P - \mathbb{I}_3$): Form $P - \mathbb{I}_3$ (notice the columns sum to 0):

$$P - \mathbb{I}_3 = \begin{pmatrix} -0.5 & 0.4 & 0 \\ 0.5 & -0.8 & 0.5 \\ 0 & 0.4 & -0.5 \end{pmatrix} \xrightarrow{R_2+R_1} \begin{pmatrix} -0.5 & 0.4 & 0 \\ 0 & -0.4 & 0.5 \\ 0 & 0.4 & -0.5 \end{pmatrix} \xrightarrow{R_3+R_2} \begin{pmatrix} -0.5 & 0.4 & 0 \\ 0 & -0.4 & 0.5 \\ 0 & 0 & 0 \end{pmatrix}$$

From the two pivot rows with v_3 as free variable:

► Row 2: $-0.4v_2 + 0.5v_3 = 0 \implies v_2 = \frac{5}{4}v_3$

► Row 1: $-0.5v_1 + 0.4v_2 = 0 \implies v_1 = \frac{4}{5}v_2 = \frac{4}{5} \cdot \frac{5}{4}v_3 = v_3$

Step 3 (Normalisation): Impose $v_1 + v_2 + v_3 = 1$, i.e.

$(1 + \frac{5}{4} + 1)v_3 = \frac{13}{4}v_3 = 1 \implies v_3 = \frac{4}{13}$, giving:

$$\mathbf{v} = \begin{pmatrix} 4/13 \\ 5/13 \\ 4/13 \end{pmatrix}, \quad \lim_{t \rightarrow \infty} P^t = W = \begin{pmatrix} 4/13 & 4/13 & 4/13 \\ 5/13 & 5/13 & 5/13 \\ 4/13 & 4/13 & 4/13 \end{pmatrix}$$

Running example: the “orange” gene in cats

A cat's coat colour is determined by many genes; the “orange” trait comes from a specific gene called the **orange locus**

It has two *alleles* (versions):

- ▶ **O** $\implies$ produces *phaeomelanin* (red/orange pigment)
- ▶ **o** $\implies$ produces *eumelanin* (black/brown pigment)

This gene is *sex-linked*: it is located on the **X chromosome**

- ▶ Females have two X chromosomes (**XX**), receiving one allele from each parent: genotypes $X^O X^O$, $X^o X^o$ or $X^O X^o$
- ▶ Males have one X and one Y chromosome (**XY**), receiving only one allele (always from the mother): genotypes $X^O Y$ or $X^o Y$

Genotype vs phenotype in cats

Males (simple):

- ▶ $X^O Y \implies$ **orange cat**
- ▶ $X^o Y \implies$ **non-orange cat** (e.g. black)

Females (the special case):

- ▶ $X^O X^O \implies$ **orange cat** (S_1)
- ▶ $X^o X^o \implies$ **non-orange cat** (S_2 , e.g. black)
- ▶ $X^O X^o \implies$ **tortoiseshell cat** (S_3)

A “tortie” is not a simple blend: because of random X-inactivation in female cells, both alleles (O and o) are active in different patches of skin, creating the mottled orange and black pattern (male tortoiseshells are extremely rare genetic anomalies with **XXY** genotype $X^O X^o Y$)

Punnett square warm-up: crossing two parents

Cross 1: orange dad ($X^O Y$) + black mum ($X^o X^o$)

		Father	
		X^O	Y
Mother	X^o	$X^O X^o$	$X^O Y$
	X^o	$X^o X^o$	$X^o Y$

- ▶ All daughters ($X^O X^o$) are **tortoiseshell**
- ▶ All sons ($X^o Y$) are **black**

Cross 2: black dad ($X^o Y$) + tortie mum ($X^O X^o$)

		Father	
		X^o	Y
Mother	X^O	$X^O X^o$	$X^O Y$
	X^o	$X^o X^o$	$X^o Y$

- ▶ Daughters: 50% tortie ($X^O X^o$), 50% black ($X^o X^o$)
- ▶ Sons: 50% orange ($X^O Y$), 50% black ($X^o Y$)

Setting up the regular Markov chain for female genotypes

Track the 3 female genotypes across generations:

- ▶ $S_1: X^O X^O$ (orange female)
- ▶ $S_2: X^O X^o$ (black female)
- ▶ $S_3: X^O X^o$ (tortoiseshell female)

In each generation, we mate the current female with a male chosen at random from a **fixed population** consisting of:

- ▶ 50% orange males ($X^O Y$)
- ▶ 50% black males ($X^o Y$)

and select one of her daughters at random as the next generation's female

State 1: orange female ($X^O X^O$)

The mother is $X^O X^O$. We pick a father with 50/50 probability:

Case 1: father is $X^O Y$

Mother		X^O	Y
	X^O	$X^O X^O (S_1)$	Male
	X^O	$X^O X^O (S_1)$	Male

Daughters: 100% S_1

Case 2: father is $X^O Y$

Mother		X^O	Y
	X^O	$X^O X^O (S_3)$	Male
	X^O	$X^O X^O (S_3)$	Male

Daughters: 100% S_3

Transitions from S_1 (column 1 of P):

- ▶ $p_{11} = \mathbb{P}(S_1 \mid S_1) = 0.5 \times 1.0 = \mathbf{0.5}$
- ▶ $p_{21} = \mathbb{P}(S_2 \mid S_1) = \mathbf{0}$
- ▶ $p_{31} = \mathbb{P}(S_3 \mid S_1) = 0.5 \times 1.0 = \mathbf{0.5}$

State 2: black female (X^oX^o)

The mother is X^oX^o . We pick a father with 50/50 probability:

Case 1: father is X^oY

		X^o	Y
Mother	X^o	$X^oX^o (S_3)$	Male
	X^o	$X^oX^o (S_3)$	Male

Daughters: 100% S_3

Case 2: father is X^oY

		X^o	Y
Mother	X^o	$X^oX^o (S_2)$	Male
	X^o	$X^oX^o (S_2)$	Male

Daughters: 100% S_2

Transitions from S_2 (column 2 of P):

- ▶ $p_{12} = \mathbb{P}(S_1 \mid S_2) = 0$
- ▶ $p_{22} = \mathbb{P}(S_2 \mid S_2) = 0.5 \times 1.0 = \mathbf{0.5}$
- ▶ $p_{32} = \mathbb{P}(S_3 \mid S_2) = 0.5 \times 1.0 = \mathbf{0.5}$

State 3: tortoiseshell female ($X^O X^o$)

The mother is $X^O X^o$. We pick a father with 50/50 probability:

Case 1: father is $X^O Y$

Mother		X^O	Y
	X^O	$X^O X^O (S_1)$	Male
	X^o	$X^O X^o (S_3)$	Male

Daughters: 50% S_1 , 50% S_3

Case 2: father is $X^o Y$

Mother		X^o	Y
	X^O	$X^O X^o (S_3)$	Male
	X^o	$X^o X^o (S_2)$	Male

Daughters: 50% S_2 , 50% S_3

Transitions from S_3 (column 3 of P):

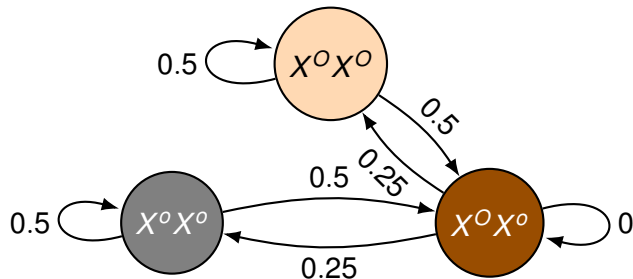
- ▶ $p_{13} = \mathbb{P}(S_1 \mid S_3) = 0.5 \times 0.5 = \mathbf{0.25}$
- ▶ $p_{23} = \mathbb{P}(S_2 \mid S_3) = 0.5 \times 0.5 = \mathbf{0.25}$
- ▶ $p_{33} = \mathbb{P}(S_3 \mid S_3) = 0.5 \times 0.5 + 0.5 \times 0.5 = \mathbf{0.5}$

Transition matrix and digraph of the orange cat chain

$$P = \begin{pmatrix} 0.5 & 0 & 0.25 \\ 0 & 0.5 & 0.25 \\ 0.5 & 0.5 & 0.5 \end{pmatrix}$$

Is this Markov chain regular?

- ▶ **Strongly connected:**
 $S_1 \rightarrow S_3 \rightarrow S_2$ and
 $S_2 \rightarrow S_3 \rightarrow S_1$, so all states communicate (P is irreducible)
- ▶ **Aperiodic:** $p_{11}, p_{22}, p_{33} > 0$ (self-loops), so P is **primitive** ($P^2 > 0$) and the chain is **regular**



Stationary distribution of the orange cat chain

We solve $(P - \mathbb{I}_3)\mathbf{v} = \mathbf{0}$ with $v_1 + v_2 + v_3 = 1$:

$$(P - \mathbb{I}_3)\mathbf{v} = \begin{pmatrix} -0.5 & 0 & 0.25 \\ 0 & -0.5 & 0.25 \\ 0.5 & 0.5 & -0.5 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

From the first two equations:

- ▶ $-0.5v_1 + 0.25v_3 = 0 \implies v_1 = 0.5v_3$
- ▶ $-0.5v_2 + 0.25v_3 = 0 \implies v_2 = 0.5v_3$

Substituting into the normalisation condition $v_1 + v_2 + v_3 = 1$:

$$0.5v_3 + 0.5v_3 + v_3 = 2v_3 = 1 \implies v_3 = 0.5, \quad v_1 = 0.25, \quad v_2 = 0.25$$

Thus the unique stationary distribution is $\mathbf{v} = (0.25, 0.25, 0.5)^T$: in the long run, regardless of the initial female's genotype, 25% of females are orange, 25% are black and 50% are tortoiseshell

Computing matrix powers and stationary distributions in R

Let us verify the convergence of P^t and the eigenpair $P\mathbf{v} = \mathbf{v}$ in R:

```
P_reg <- matrix(c(0.5, 0, 0.25,
                  0, 0.5, 0.25,
                  0.5, 0.5, 0.5),
               nrow = 3, byrow = TRUE)
# Dominant right eigenvector corresponding to eigenvalue 1
ev <- eigen(P_reg)
v_stat <- Re(ev$vectors[, 1])
v_stat <- v_stat / sum(v_stat)
v_stat

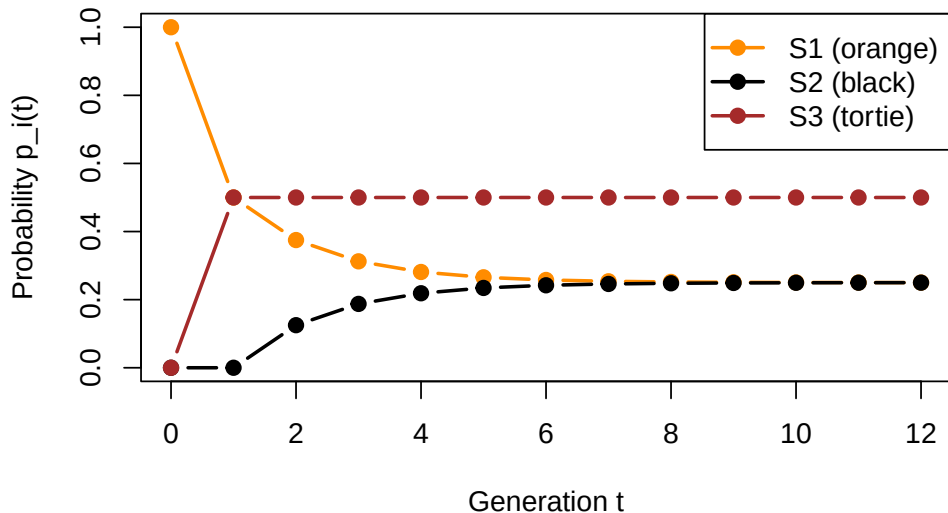
## [1] 0.25 0.25 0.50

# Compare with  $P^{20}$  computed by repeated multiplication
P_20 <- Reduce(`%*%`, replicate(20, P_reg, simplify = FALSE))
round(P_20, 6)
```

Convergence of $\mathbf{p}(t) = \mathbf{P}^t \mathbf{p}(0)$ from an initial state

```
t_max <- 12
p_hist <- matrix(0, nrow = 3, ncol = t_max + 1)
p_hist[, 1] <- c(1, 0, 0) # Start with 100% orange female (S_1)
for (t in 1:t_max) {
  p_hist[, t + 1] <- P_reg %*% p_hist[, t]
}
matplot(0:t_max, t(p_hist), type = "b", pch = 19, lty = 1, lwd = 2,
        col = c("darkorange", "black", "brown"),
        xlab = "Generation t", ylab = "Probability p_i(t)",
        ylim = c(0, 1), main = "Convergence to stationary distribution v = (",
        legend("topright", legend = c("S1 (orange)", "S2 (black)", "S3 (tortie)"),
        col = c("darkorange", "black", "brown"), lty = 1, pch = 19, lwd = 2)
```

Convergence to stationary distribution $v = (0.25, 0.25, 0.5)$

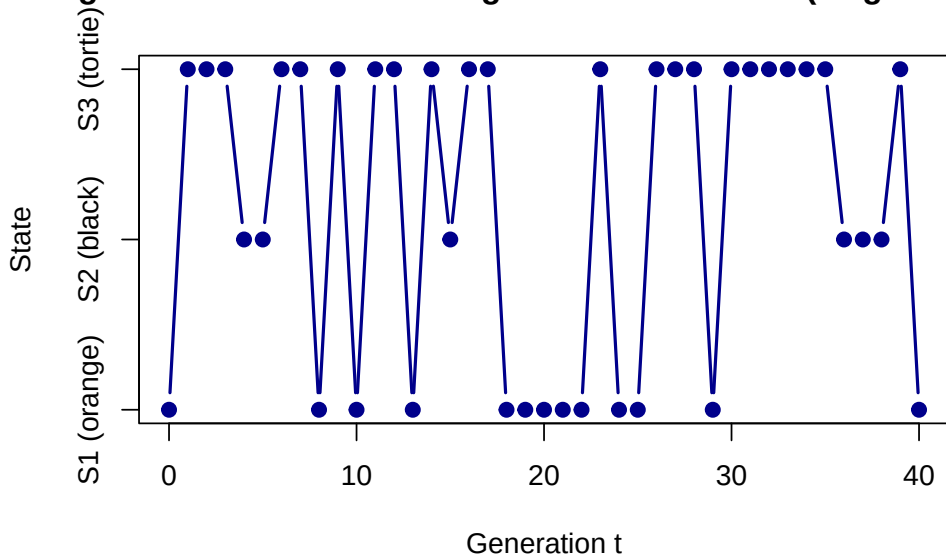


Simulating a single realisation of the regular Markov chain

In a column-stochastic chain, when in state j at step t , the next state at step $t + 1$ is drawn with probabilities given by **column** j of P :

```
set.seed(2740)
n_steps <- 40
states <- numeric(n_steps + 1)
states[1] <- 1 # Start in state S_1 (orange female)
for (t in 1:n_steps) {
  states[t + 1] <- sample(1:3, size = 1, prob = P_reg[, states[t]])
}
plot(0:n_steps, states, type = "b", pch = 19, lwd = 1.8, col = "darkblue",
     yaxt = "n", xlab = "Generation t", ylab = "State",
     main = "Single realisation of the orange cat Markov chain (40 generations)",
     axis(2, at = 1:3, labels = c("S1 (orange)", "S2 (black)", "S3 (tortie)"))
```

Single realisation of the orange cat Markov chain (40 generations)



Empirical state frequencies vs stationary distribution $\mathbf{v}$

By the ergodic theorem for regular Markov chains, the long-run fraction of time spent in each state along a single trajectory converges to $\mathbf{v} = (0.25, 0.25, 0.5)^T$:

```
set.seed(2740)
n_long <- 10000
traj <- numeric(n_long)
traj[1] <- 1
for (t in 1:(n_long - 1)) {
  traj[t + 1] <- sample(1:3, size = 1, prob = P_reg[, traj[t]])
}
# Empirical frequencies over 10,000 generations vs exact v
rbind(empirical = prop.table(table(factor(traj, levels = 1:3))),
      stationary = v_stat)
```

##	1	2	3
## empirical	0.2414	0.2505	0.5081
## stationary	0.2500	0.2500	0.5000