POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Periodicity in Large-Scale Markov Chains: Algorithmic Detection and Spectral Characterization

Tesi di Laurea Triennale in
Mathematical Engineering - Ingegneria Matematica

Authors: **Cosimo Raugei, Leonardo Torti**

Student IDs: 209973, 211882
Advisors: Prof. Federico Bassetti, Prof. Giuseppina Guatteri
Academic Year: 2024-25

Contents

Abstract

Identifying the period - especially aperiodicity - of discrete-time finite-state Markov Chains is essential for predicting stable states and limit behavior within the state space. However, as the number of states increases, the complexity of the state space and its transition structure makes this task computationally demanding.

Our thesis is organized as follows: in the first chapter we find and prove the smallest number of steps (as function of the size of the matrix and some of its properties) such that, beyond this integer, the period becomes definitive. Leveraging our theoretical result, in the second chapter we propose two novel algorithms for period determination in large-scale Markov Chains: Systemic Period Evaluator (SysPe), which determines periodicity based on communication class structure, and Single-State Period Evaluator (SSPE), which exclusively analyzes a single state's return probabilities.

The third chapter focuses on the spectral structure of stochastic matrices and building upon known results, we derive, to the best of our knowledge, an original factorization of the return probability for a single state. This formulation separates two fundamental components of the chain's behavior: the limit properties, which govern long-term recurrence, and the transient phase. By expressing these two components as independent yet equally significant factors, our result provides a deeper structural understanding of periodicity.

Abstract in lingua italiana

Identificare il periodo - e, in particolare, l'eventuale aperiodicità - di una catena di Markov a tempo discreto e spazio degli stati finito è fondamentale per prevederne il comportamento a lungo termine. All'aumentare della dimensione dello spazio in cui le catene evolvono, ne accresce drasticamente la complessità del quadro evolutivo, rendendo complicato il problema di determinarne il periodo tramite la definizione.
La nostra tesi è articolata come segue: nel primo capitolo troviamo e dimostriamo il numero minimo di passi (in funzione della dimensione della matrice e altre sue proprietà) tale per cui il periodo diventa definitivo. Sfruttando questo risultato, nel secondo capitolo proponiamo due algoritmi originali per determinarlo: il "Systemic Period Evaluator" (SysPe), che valuta la periodicità basandosi sulla struttura delle classi di comunicazione, e il "Single-State Period Evaluator" (SSPE), che analizza esclusivamente le probabilità di ritorno di un singolo stato.
Il terzo capitolo è dedicato alla struttura spettrale delle matrici stocastiche. Sfruttando risultati già noti, deriviamo una nostra formulazione fattorizzata della probabilità di ritorno per un singolo stato di una catena. Tramite essa, separiamo le due componenti fondamentali del comportamento della catena: la dinamica persistente, che governa il comportamento a lungo termine e quello transitorio. Trattando questi due elementi come fattori indipendenti ma ugualmente rilevanti, nella dissertazione offriamo una prospettiva sui loro effetti che si discosta dalla trattazione accademica classica.

Preliminary Results

In this section, we introduce the fundamental concepts necessary for understanding the subsequent chapters. This serves as a focused introduction rather than a comprehensive overview of Markov chains. For a more in-depth treatment, we refer the reader to *Markov Chains* by P. Brémaud [1], from which the definitions in this section are drawn.

Definition 0.0.1. *A stochastic process is a family of random variables* $\{X_t\}_{t\in T}$ *defined on the same probability space and taking values in a common measurable space (called state space):* $X_t : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathscr{E})$ *where* t *is a parameter* $\in T \subseteq \mathbb{R}^+$, $|E| < +\infty$

Definition 0.0.2. *A square matrix* $P = \{p_{i,j}\}_{i,j\in E}$ *of size* $|E| \times |E|$ *is a stochastic transition matrix if:*

(i) $p_{i,j} \geq 0 \quad \forall\, i, j \in E$

(ii) $\sum_{j\in E} p_{i,j} = 1 \quad \forall\, i \in E$

From now on, we will focus on **Markov Chains**.

Definition 0.0.3. *A discrete-time Markov Chain with finite state space* E*, transition matrix* P *and initial discrete distribution* v *on* $(E, \mathscr{E})$ *is a sequence of random variables* $\{X_n\}_{n\geq 0} : (\Omega, \mathcal{A}, \mathbb{P}) \to (E, \mathscr{E})$ *which satisfies the following:*

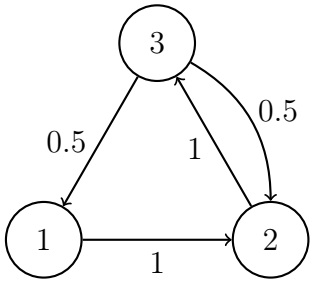
(i) $X_0 \sim v$*, that is* $\mathbb{P}(X_0 = i) = v(i) \quad \forall i \in E$

(ii) $\mathbb{P}(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = \mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{ij} \quad \forall i, j \in E, \quad \forall n \in \mathbb{N}$

Definition 0.0.4. *Given a Markov Chain with transition matrix* P *and state space* E*, the* ***Transition Graph*** *associated to the chain is the directed graph* $\Gamma = (\mathcal{N}(\Gamma) = E, \mathcal{E}(\Gamma) = \{(i, j) : p_{i,j} > 0\})$.

Thus, it is possible to obtain a graphical representation of the Chain where the states are represented as nodes of the graph and directed edges between nodes represent possible transitions between states.

For example, consider a system with three states:



$$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.5 & 0.5 & 0 \end{bmatrix}$$

Figure 0.1: *graph representation of a simple Markov Chain and its corresponding transition matrix*

Theorem 0.0.5. $(P^n)_{i,j} = \mathbb{P}(X_n = j \mid X_0 = i)\ \forall i, j \in E$ *and* $\forall n \in \mathbb{N}$

A proof of this classical result can be found in [1]. From this point onward, we will denote $(P^n)_{i,j}$ as $p_{i,j}^{(n)}$, with $p_{i,j}^{(0)} := \delta_{i,j}$. For Theorem 0.0.5, $p_{i,j}^{(n)}$ can be interpreted as the probability of transitioning from state i to state j in n steps.

Definition 0.0.6. *State j is said to be* ***accessible*** *from state i if there exists $n \geq 0$ such that $p_{ij}^{(n)} > 0$. States i and j are said to* ***communicate*** *if i is accessible from j and j is accessible from i, denoted by $i \leftrightarrow j$.*

For $n \geq 1$, we have $p_{ij}^{(n)} = \sum_{i_1,\ldots,i_{n-1}} p_{ii_1} p_{i_1 i_2} \ldots p_{i_{n-1} j}$, and therefore $p_{ij}^{(n)} > 0$ if and only if there exists at least one path $i, i_1, \ldots, i_{n-1}, j$ from i to j such that:

$$p_{ii_1} p_{i_1 i_2} \ldots p_{i_{n-1} j} > 0,$$

or, equivalently, if there is a directed path from i to j in the transition graph Γ.

Clearly:

- $i \leftrightarrow i$ (reflexivity),
- $i \leftrightarrow j \implies j \leftrightarrow i$ (symmetry),
- $i \leftrightarrow j$ and $j \leftrightarrow k \implies i \leftrightarrow k$ (transitivity).

The communication relation $\leftrightarrow$ is an equivalence relation and generates a partition of the nodes $\mathcal{N}(\Gamma)$ into disjoint equivalence classes called **communication classes**. We will call $\mathcal{H}_i$ the communication class containing a state $i \in \mathcal{N}(\Gamma)$

Definition 0.0.7. *The **period** $d(i)$ of a state $i \in E$ is defined as:*

$$d(i) = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\},$$

where $p_{ii}^{(n)}$ represents the probability of returning to state i after n steps. If there is no $n \geq 1$ such that $p_{ii}^{(n)} > 0$, then we set $d(i) = +\infty$ by convention.

If $d(i) = 1$, the state i is called **aperiodic**. We are now prepared to explore the concept of the period of a Markov chain, which will be the central topic of the next two chapters.

1 | Period Analysis on the Underlying Oriented Graph

1.1. Relations between Communication and Period

Lemma 1.1.1. *Chicken McNugget Theorem or Frobenius Coin Problem [3].*
Let a and b be two positive integers such that $\gcd(a, b) = 1$ (i.e., a and b are coprime). The largest number N that cannot be expressed as a non-negative integer linear combination of a and b (i.e., of the form $n = ax + by$ for non-negative integers x and y) is given by:

$$N = ab - a - b.$$

Thus, any integer greater than N, which we will call McNugget's Number, can be expressed as a non-negative integer linear combination of a and b. More can be found here [3]

Theorem 1.1.2. *If i and j communicate, then $d(i) = d(j)$.*

This is a classical result, also presented in [1]. Below, we provide an alternative proof of our own.

Proof. We aim to show that $d(i) \geq d(j)$ and, by symmetry, conclude that $d(i) = d(j)$.

By hypothesis, there exist $k, l \in \mathbb{N}$ such that $p_{ij}^{(k)} > 0$ and $p_{ji}^{(l)} > 0$. This implies that

$$p_{ii}^{(k+l)} \geq p_{ij}^{(k)} \cdot p_{ji}^{(l)} > 0,$$

and similarly,

$$p_{jj}^{(k+l)} \geq p_{ji}^{(l)} \cdot p_{ij}^{(k)} > 0.$$

Thus, both $d(i)$ and $d(j)$ divide $k + l$.

Case 1: $d(i) > 1$.

Let $n > 0$ be such that $d(i)$ does not divide n. Such an n exists since $d(i) > 1$. Then

$p_{ii}^{(n)} = 0$. Also, since $d(i)$ divides $k+l$ but not n, it does not divide $n+k+l$, so $p_{ii}^{(n+k+l)} = 0$.

However,

$$p_{ii}^{(n+k+l)} \geq p_{ij}^{(k)} \cdot p_{jj}^{(n)} \cdot p_{ji}^{(l)}.$$

Since $p_{ij}^{(k)} > 0$ and $p_{ji}^{(l)} > 0$, we have

$$0 = p_{ii}^{(n+k+l)} \geq (\text{positive constant}) \cdot p_{jj}^{(n)}.$$

Therefore, $p_{jj}^{(n)} = 0$, meaning $d(j)$ does not divide n.

This shows that every n not divisible by $d(i)$ is also not divisible by $d(j)$. Thus, the multiples of $d(j)$ are contained within the multiples of $d(i)$, so $d(j) \geq d(i)$.

By symmetry (since $i \leftrightarrow j$), we also have $d(i) \geq d(j)$. Therefore, $d(i) = d(j)$.

Case 2: $d(i) = 1$.

Since $d(i)$ is 1, state i is aperiodic. In addition, by the properties of gcd, there exist integers r and s such that:

$$\gcd(r, s) = 1 \quad \text{and} \quad p_{ii}^{(r)} > 0,\ p_{ii}^{(s)} > 0.$$

Because r and s are co-prime and associated with positive probabilities, we can generate any integer $n > N$ McNugget's number as a linear combination:

$$n = a \cdot r + b \cdot s,$$

where a and b are non-negative integers. Thus, $p_{ii}^{(n)} > 0$ for all $n \geq N$ (see Lemma 1.1.1).

For any $n \geq N + k + l$, consider $n' = n - k - l \geq N$. Then $p_{ii}^{(n')} > 0$. We have:

$$p_{jj}^{(n)} \geq p_{ji}^{(l)} \cdot p_{ii}^{(n')} \cdot p_{ij}^{(k)} > 0,$$

since $p_{ji}^{(l)} > 0$, $p_{ii}^{(n')} > 0$, and $p_{ij}^{(k)} > 0$.

Therefore, $p_{jj}^{(n)} > 0$ for all sufficiently large n, implying $d(j) = 1$.

Thus, $d(i) = d(j) = 1$. □

Remark: every state in a communication class has the same period.

1.2. Upper bounds for Period Stabilization

The period of a state is, by definition, the g.c.d. of the lengths of all closed paths starting and ending in the mentioned state, which could be arbitrarily big. Without additional clarifications, this would mean that checking paths with length less than or equal to a finite $l \in \mathbb{N}$, even if very large, would not be enough to determine the period, as it could exist a longer path changing its value (obviously, this can only happen if the chain is not already found to be aperiodic when checking paths with length less than or equal to l). For example, several theorems concerning useful properties of asymptotic behavior of Markov Chains require aperiodicity, hence, it is particularly useful to establish a bound that guarantees the definitiveness of the period (in this case, the certainty that we will never achieve aperiodicity). In this section, we address this problem and provide a concrete solution.

Proposition 1.2.1. *Let Γ be a graph such that $|\mathcal{N}(\Gamma)| = n$, and let $i \neq j \in \mathcal{N}(\Gamma)$. If $i \to j$, then there exists a path of length l, with $l < n$, that leads from i to j.*

Proof. Suppose, for contradiction, that the shortest path from i to j has length $l^* \geq n$. According to the Dirichlet Pigeonhole Principle (more can be found here [2]), since there are only n distinct nodes in Γ, but the path has $l^* \geq n$ steps, there must be at least one node that is visited more than once. Let this repeated node be k.

Thus, the path from i to j can be written as:

$$i \dots k \dots k \dots j$$

where the dots represent other states visited along the way. Since the node k appears twice, we can remove the subpath between the two occurrences of k (i.e., the subpath from the first occurrence of k to the second occurrence of k) to "compress" the path. The new path, after removing this cycle, is shorter but still leads from i to j. This contradicts our assumption that the shortest path had length $l^* \geq n$. Therefore, the shortest path from i to j must have length $l < n$, as required.

□

Remark: to construct the shortest path from i to j, we can repeat this process each time we find a repeated node in the path. By continually removing cycles, we eventually arrive at a path with no repeated nodes. Since there are only n distinct nodes in $\mathcal{N}(\Gamma)$,

the final compressed path must have length at most $n-1$.

Remark: in the remaining case where $i = j$, on the one hand by definition i leads to i in 0 steps, on the other hand if there exists a non-trivial path leading from i to i, by the same reasoning as before, we conclude that it must have length l s.t. $0 < l \leq n$.

Corollary 1.2.2. *The set of the entries of the powers of the transition matrix P up to n (specifically, the entries of $P^1, P^2, \ldots, P^{n-1}$) constitutes the minimal set of numbers required to characterize the partition of the state space $\mathcal{N}(\Gamma)$ into communication classes.*

Proof. For each state i, if j belongs to the communication class of i, then by definition $i \leftrightarrow j$. By Proposition 1.2.1, there must exist two paths of lengths l_1 and l_2 with $l_1, l_2 < n$ such that:

$$i \to j \text{ in } l_1 \text{ steps } (p_{ij}^{(l_1)} > 0) \quad \text{and} \quad j \to i \text{ in } l_2 \text{ steps } (p_{ji}^{(l_2)} > 0).$$

To conclude the proof, it is sufficient to note that these values correspond to some entries of powers of the transition matrix P lower than n, specifically $(P^{l_1})_{ij}$ and $(P^{l_2})_{ji}$. □

Remark: as interpretation of the Corollary, by Proposition 1.2.1, any path of length $l \geq n$ from one state i to another j can be "compressed" to a strictly shorter path $l^* < n$. Consequently, to check whether i leads to j, it suffices to look for paths of length at most n. In matrix terms, the existence of a path of length k from i to j is equivalent to $(P^k)_{ij} \neq 0$. Hence, it is enough to examine the entries of $P, P^2, \ldots, P^{n-1}$.

Definition 1.2.3. *Let Γ be a graph, P its transition matrix, $s \in \mathbb{N}$, $i \in \mathcal{N}(\Gamma)$. We define partial period $d_s(i)$ of a state i till s:*

$$d_s(i) = \begin{cases} \gcd\left\{1 \leq j \leq s : p_{ii}^{(j)} > 0\right\}, & \text{if } p_{ii}^{(j)} > 0 \text{ for at least one } j \in \{1, \ldots, s\} \\ +\infty, & \text{if } p_{ii}^{(j)} = 0 \ \forall\ j \in \{1, \ldots, s\} \end{cases}$$

Remark: The partial period $d_s(i)$ of a state i till s shows how the period would look like if only considering paths with length $l \leq s$. By increasing s, we want to see when $d_s(i)$ stabilizes (i.e. assumes the value of $d(i)$). Notice that the definitions of period and partial period are consistent with what we want to observe, namely $d_s(i) = d(i)$ for $s \to +\infty$.

Definition 1.2.4. *Let $\mathcal{H}$ be a communication class of a graph. We define the partial period $d_s(\mathcal{H})$ of $\mathcal{H}$ until s as:*

$$d_s(\mathcal{H}) = \begin{cases} \gcd\{d_s(i) \mid i \in \mathcal{H},\ d_s(i) < +\infty\}, & \text{if } \exists\, i \in \mathcal{H} \text{ such that } d_s(i) < +\infty, \\ +\infty, & \text{if } d_s(i) = +\infty\ \forall i \in \mathcal{H}. \end{cases}$$

Let Γ be a graph and $i \in \mathcal{N}(\Gamma)$. As said at the beginning of this section (but utilizing the definition of partial period), for every $l \in \mathbb{N}$ there could exist a path of length $l_1 > l, l_1 \in \mathbb{N}$ such that $d_l(i)$ does not divide l_1 (if $d_l(i) \neq 1$, otherwise it is safe to stop checking paths as said before). The following theorem gives us an upper bound for period stabilization.

Note: two paths $\gamma_1 = i \cdots j_1$ and $\gamma_2 = j_2 \cdots k$ can be concatenated if, and only if, $j_1 = j_2 = j$, and we write $\gamma = \gamma_1 \cup \gamma_2 = i \cdots j \cdots k$. The length of a path γ is referred to as $|\gamma|$ and is equal to the number of steps taken in the path. Observe that the starting state does not count as a step.

Theorem 1.2.5. *Let Γ be a graph. Then, for any $i \in \mathcal{N}(\Gamma)$, the following holds:*

$$\begin{aligned} &I)\ d_s(\mathcal{H}_i) = d(i),\ \forall s \geq |\mathcal{H}_i| \\ &II)\ d_s(i) = d(i),\ \forall s \geq 2|\mathcal{H}_i| - 1 \\ &III)\ \textit{The bounds established in (I,II) are optimal.} \end{aligned}$$

Remark: The partial period of a communication class (I) stabilizes earlier (with earlier referring to the length of paths to check) than the partial period of a state (II), but it requires the observation of a smaller number of paths. Thus, (II) maintains a practical utility (see the two different period-finding algorithms in chapter 2) for a complete overview of the computational aspects concerning these results.

Proof. Let $i \in \mathcal{N}(\Gamma)$ be a state and $\mathcal{H}_i$ its communication class. First, observe that we are only interested in paths entirely contained within $\mathcal{H}_i$. In fact, if a path has the form $i \ldots j \ldots k$, where $j \notin \mathcal{H}_i$, then $k \notin \mathcal{H}_i$. This implies that every closed path starting and ending at a state in $\mathcal{H}_i$ is completely contained in $\mathcal{H}_i$.

I) Now, consider a closed path $\gamma = i...i$ of length $|\gamma| > |\mathcal{H}_i|$. For simplicity, we assume:

$$\gamma \neq \underbrace{i \quad \ldots \quad i}_{\gamma_1} \quad \ldots \quad \underbrace{i}_{\gamma_2}$$

which means i is only reached at the beginning and end of the path, otherwise we could just separately consider γ_1 and γ_2 and, if $|\gamma_1| > |\mathcal{H}_i|$ or $|\gamma_2| > |\mathcal{H}_i|$, proceed as in the following. Instead, if $|\gamma_1| \leq |\mathcal{H}_i|$ and $|\gamma_2| \leq |\mathcal{H}_i|$ there would be nothing to prove.
Now, by the Dirichlet Pigeonhole Principle [2], there exists a state $j \in \mathcal{H}_i$ reached at least two times, meaning that we can rewrite the path γ as:

$$\gamma = \underbrace{i \quad \dots \quad j}_{\gamma_1} \underbrace{\quad \dots \quad j}_{\gamma_2} \underbrace{\quad \dots \quad i}_{\gamma_3}$$

Due to Proposition 1.2.1, we can choose j such that the path

$$\gamma_1 \cup \gamma_3$$

has length

$$|\gamma_1 \cup \gamma_3| \leq |\mathcal{H}_i|.$$

W.L.O.G., we assume that

$$|\gamma_2| \leq |\mathcal{H}_i|.$$

Otherwise, if $|\gamma_2| > |\mathcal{H}_i|$, we could reiterate the process by applying Proposition 1.2.1 to γ_2.

Thus, $\gamma_1 \cup \gamma_3$ and γ_2 are closed paths of lengths

$$|\gamma_1 \cup \gamma_3|, |\gamma_2| \leq |\mathcal{H}_i|.$$

Since, by definition, $d_{|\mathcal{H}_i|}(\mathcal{H}_i)$ divides both

$$|\gamma_1 \cup \gamma_3| = |\gamma_1| + |\gamma_3|$$

and

$$|\gamma_2|,$$

it also divides

$$|\gamma_1| + |\gamma_2| + |\gamma_3| = |\gamma|.$$

This holds for any closed path γ of length

$$|\gamma| > |\mathcal{H}_i|,$$

thus proving that

$$d_s(\mathcal{H}_i) = d(i), \quad \forall s \geq |\mathcal{H}_i|.$$

II) Let $i \in \mathcal{N}(\Gamma)$ be a state and $\mathcal{H}_i$ its communication class. Consider a closed path $\gamma = i \cdots i$ of length $|\gamma| > 2|\mathcal{H}_i| - 1$.

As made in (I), we assume:

$$\gamma \neq \underbrace{i \quad \dots \quad i}_{\gamma_1} \quad \dots \quad \underbrace{i}_{\gamma_2}$$

which means i is only reached at the beginning and end of the path, otherwise we could just separately consider γ_1 and γ_2 and, if $|\gamma_1| > 2|\mathcal{H}_i| - 1$ or $|\gamma_2| > 2|\mathcal{H}| - 1$, proceed as in the following. Otherwise, as above, the thesis follows.
By the Dirichlet Pigeonhole Principle, there exists a state $j \neq i$ reached at least three times, meaning that:

$$\gamma = \underbrace{i \quad \dots \quad j}_{\gamma_1} \quad \dots \quad \underbrace{j}_{\gamma_2} \quad \dots \quad \underbrace{j}_{\gamma_3} \quad \dots \quad \underbrace{i}_{\gamma_4}$$

Thanks to Proposition 1.2.1, without loss of generality, we can assume

$$|\gamma_1 \cup \gamma_4| \leq |\mathcal{H}_i|.$$

Thus, by definition, $d_{2|\mathcal{H}_i|-1}(i)$ divides

$$|\gamma_1 \cup \gamma_4| = |\gamma_1| + |\gamma_4|,$$

as $\gamma_1 \cup \gamma_4$ is a closed path starting and closing on i with length

$$|\gamma_1 \cup \gamma_4| \leq |\mathcal{H}_i| \leq 2|\mathcal{H}_i| - 1.$$

Furthermore, γ_2, γ_3 can only reach states in $\mathcal{H}_i \setminus \{i\}$, which has cardinality

$$|\mathcal{H}_i \setminus \{i\}| = |\mathcal{H}_i| - 1.$$

So, we can assume (by Proposition 1.2.1 again) that

$$|\gamma_2|, |\gamma_3| \leq |\mathcal{H}_i| - 1.$$

Observe that $\gamma_1 \cup \gamma_2 \cup \gamma_4$ is a closed path starting and closing on i with length

$$|\gamma_1 \cup \gamma_2 \cup \gamma_4| = |\gamma_1| + |\gamma_2| + |\gamma_4| \leq |\mathcal{H}_i| + |\mathcal{H}_i| - 1 = 2|\mathcal{H}_i| - 1,$$

and the same reasoning applies for $\gamma_1 \cup \gamma_3 \cup \gamma_4$.

By its definition,

$$d_{2|\mathcal{H}_i|-1}(i) \text{ divides } |\gamma_1 \cup \gamma_2 \cup \gamma_4| = |\gamma_1| + |\gamma_2| + |\gamma_4|$$

and

$$|\gamma_1 \cup \gamma_3 \cup \gamma_4| = |\gamma_1| + |\gamma_3| + |\gamma_4|.$$

Hence, by simple properties of divisibility,

$$d_{2|\mathcal{H}_i|-1}(i) \text{ divides } |\gamma_1| + |\gamma_2| + |\gamma_3| + |\gamma_4| = |\gamma|.$$

This holds for any closed path of length $l > 2|\mathcal{H}_i| - 1$, proving that

$$d_s(i) = d(i), \quad \forall s \geq 2|\mathcal{H}_i| - 1.$$

III) In this final section of the proof, we demonstrate that - in the worst case scenario for a fixed size of the state space - the period of a state or class does not stabilize before the previously established bounds, hence they are optimal. To support this claim, we construct a counterexample. Before addressing the general case, we first present a specific instance to guide the reader, then proceed with a rigorous argument.

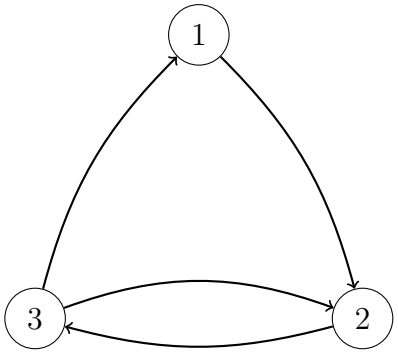


Figure 1.1: *a first specific counterexample with three states.*

Consider state 1 of the graph given in fig 1.2. To analyze the period of state 1, we observe the following paths:

- A path of length 3: $(1 \to 2 \to 3 \to 1)$.
- A path of length 5: $(1 \to 2 \to 3 \to 2 \to 3 \to 1)$, which corresponds to the bound $2|\mathcal{H}_1| - 1 = 5$, where $|\mathcal{H}_1| = 3$ (the size of the communication class).

These paths demonstrate that it is necessary to consider paths of length 5 (i.e., $2|\mathcal{H}_1| - 1$) for the period of state 1 to stabilize, as no shorter path captures the periodic behavior.

Moreover, by considering the period of the entire communication class $\mathcal{H}_1 = \{1, 2, 3\}$, we observe that:

- State 1 has paths of length 3: $(1 \to 2 \to 3 \to 1)$.
- States 2,3 have paths of length 2 : $(2 \to 3 \to 2), (3 \to 2 \to 3)$ respectively.

Thus, the periodic behavior of the class as a whole is fully captured by paths of length $3 = |\mathcal{H}_1|$.

This example can be **generalized to any** n by considering the graph reported below.

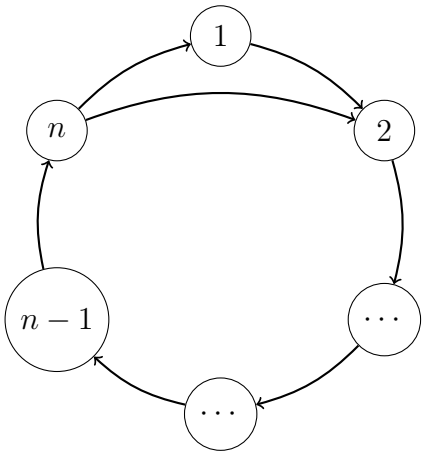


Figure 1.2: *general counterexample with n states.*

We begin by noting that this graph represents an **aperiodic communication class** of size n. We analyze the **optimality** of the bounds in two cases:

- **Optimality of I:** This example demonstrates that when considering paths within the class, the period stabilizes no previously then paths of lengths $n-1$ and n. Consequently, it is necessary to check paths up to length n.
- **Optimality of II:** If we instead fix a generic state i, the aperiodic behavior is fully captured by paths of the following lengths:
  - For $i = 1$, the relevant path lengths are n and $2n-1$.
  - For $i \neq 1$, the necessary path lengths reduce to n and $n-1$.

  Consequently it is necessary to check paths up to length $2n-1$.

Given the generality of n , Theorem 1.2.5 provides an upper bound, while this counterexample establishes a matching lower bound. Consequently, we have proven that these bounds are optimal and cannot be further refined without imposing additional hypotheses.

Thus, the proof is complete.

□

2 | Period Stabilization Algorithms

2.1. Systemic Period Evaluator (SysPe)

In this section, we present an efficient implementation of the method for finding the period and communication class of a state i by analyzing all the paths that start and end from all the possible states of length up to n. The algorithm, as explained below, follows from a direct application of result 1 of Theorem **1.2.5**.

Complexity:
Time complexity = $O(n^4)$
Memory space complexity = $O(n^3)$ (memory required to run the program)
where n as before corresponds to $|\mathcal{N}(\Gamma)|$

Input:

- A stochastic matrix P of size $n \times n$, representing the transition probabilities of a Markov chain.
- An integer i, representing the index of the state for which the period is to be found (1-based indexing).

Output:

- The period d of state i.
- The set of states $\mathcal{H}_i$ that communicate with state i, which forms its communication class.

Overview of the Algorithm:

- **Step 1: Input Validation**
  - The algorithm first checks if the input matrix P is stochastic (i.e., it is square, has non-negative elements, and the rows sum to 1).
  - It also verifies that the state i is a valid index within the matrix.

- **Step 2: Powers of the Transition Matrix**
  - The algorithm computes the first n powers of the matrix P, i.e., $P, P^2, \ldots, P^n$.
  - These powers are stored in a list R, where $R[k]$ represents P^{k+1}.
  - These powers help check if states communicate through any sequence of transitions.
- **Step 3: Dividing the State Space into Communication Classes**
  - Using the powers of the transition matrix, the algorithm identifies which states communicate with each other, forming communication classes.
  - Two states i and j are in the same communication class if there exists some $k, h < n$ such that both $P^k(i,j) > 0$ and $P^h(j,i) > 0$.
  - The communication classes are stored in a matrix C, where each row represents a class of states that communicate with each other.
- **Step 4: Identifying the Communication Class of State i**
  - The algorithm determines the communication class $\mathcal{H}_i$ for state i by finding the row in the matrix C that contains state i.
  - This set $\mathcal{H}_i$ represents all the states that can communicate with state i.
- **Step 5: Calculating the Period of State i**
  - The algorithm then calculates the period d of the communicant class of state i (which we know from Theorem **1.2.5** stabilizes in less then n powers).
  - For each power of P (from 1 to n), it updates the partial period of the class if there is an element > 0 on the principal diagonal of the sub-matrix corresponding to $\mathcal{H}_i$
- **Step 6: Returning the Results**
  - The algorithm returns the period d and the communication class $\mathcal{H}_i$ for state i.

Here is a link to the full algorithm code repository on GitHub available in C++,MATLAB and Python. The code can be accessed either directly in the folder or through the README under the section titled "1. n Algorithm Period Stabilization," where the preferred programming language can be chosen.

2.2. Single-State Period Evaluator (SSPE)

In this section, we present an efficient implementation of the method for finding the period of a state i by analyzing only the paths that start and end at i. The algorithm, as explained below, follows from a direct application of result II of Theorem **1.2.5**.

Complexity:
Time complexity = $O(n^4)$
Memory space complexity = $O(n^2)$
where n as before corresponds to $|N(\Gamma)|$

Input:

- A stochastic matrix P of size $n \times n$, representing the transition probabilities of a Markov chain.
- An integer i, representing the index of the state for which the period is to be found (1-based indexing).

Output:

- The period d of state i.

Overview of the Algorithm:

- **Step 1: Input Validation**
  - The algorithm first checks if the input matrix P is stochastic (i.e., it is square, has non-negative elements, and the rows sum to 1).
  - It also verifies that the state i is a valid index within the matrix.
- **Step 2: Initialize Matrix M and Period d**
  - The algorithm initializes the identity matrix M of size $n \times n$, which will be used to compute powers of the matrix P.
  - It also initializes the period d of state i to 0.
- **Step 3: Calculate Powers of the Transition Matrix**
  - The algorithm computes powers of the matrix P by iteratively multiplying M by P for up to $2n - 1$ steps (i.e., $M = P^k$ for $k = 1, 2, \ldots, 2n - 1$).
  - At each step k, it updates matrix M to P^k, representing the matrix after k steps of transitions.

- **Step 4: Check for Return to State i and Update Period**
  - During each iteration, the algorithm checks whether the entry $M[i, i]$ is greater than 0, indicating that state i can return to itself after k steps.
  - If a return to state i is detected (i.e., $M[i, i] > 0$), the period d is updated using the greatest common divisor (gcd) of the current value of d and k.
  - If this is the first time state i returns to itself, the period d is set to k.
  - The loop breaks early if the period d is found to be 1, as this is the smallest possible period.
- **Step 5: Return the Results**
  - After checking up to $2n - 1$ steps, the algorithm returns the period d of state i.

Here is a link to the full algorithm code repository on GitHub available in C++,MATLAB and Python. The code can be accessed either directly in the folder or through the README under the section titled "2. 2n-1 Algorithm Period Stabilization," where the preferred programming language can be chosen.

2.3. A comparison of the two Algorithms

The two methods are based on the results presented in Theorem1.2.5. It is important to highlight the different approaches these algorithms take, along with their respective pros and cons, to provide guidance on which to use depending on the situation.

Approach

- **Algorithm 1:** This algorithm uses the first result of Theorem1.2.5. It calculates the period by considering all paths within the communication class of state i, checking paths of length up to $|\mathcal{H}_i|$ for all states in the class. The output includes both the communication class and the period of the class.
- **Algorithm 2:** This algorithm uses the second result of Theorem1.2.5, focusing only on paths that start and end at state i, with path lengths up to $2n-1$. The output is solely the period of state i.

Complexity

- **Algorithm 1:** Time complexity $O(n^4)$, space complexity $O(n^3)$.
- **Algorithm 2:** Time complexity $O(n^4)$, space complexity $O(n^2)$. Although the time complexity is asymptotically double that of Algorithm 1 (due to the need to check powers of the transition matrix up to $2n-1$ vs n), the space complexity is lower, being polynomially one order smaller.

Conclusion and Guidelines

- **When to use Algorithm 1:** Algorithm 1 should be used when it is necessary to determine the full communication class along with the period. It is also the better choice when time constraints are more critical, as it runs twice as fast as Algorithm 2.
- **When to use Algorithm 2:** Algorithm 2 is more suitable when memory (space) limitations are significant. If only the period of a specific state is needed and communication class information is not required.

3 | Eigenvalues & Period Analysis

In this chapter, we aim to demonstrate the relationship between a Markov chain and the spectral structure of its transition matrix (i. e. the distribution of its eigenvalues on the complex plane). At the end, in Theorem 3.2.4 we provide our main contribution by showcasing and proving an algebraic representation of the n-steps return probability for each state.

3.1. Spectral Structure of Periodic Markov Chains

Definition 3.1.1. *A Markov chain is called* ***irreducible*** *if for every* $i, j \in \mathcal{N}(\Gamma)$*, then* $i \leftrightarrow j$

From now on, irreducibility will always be assumed. Furthermore, we will refer to irreducible transition matrices if the associated Markov Chain is irreducible. Similarly, we will say that a matrix has period d if it is associated to a Markov Chain of period d.

Theorem 3.1.2. *Let* P *be an irreducible transition matrix of period* d*, then the followings hold:*
i) P *has precisely* d *distinct eigenvalues of modulus 1 taking the form:*

$$\hat{\lambda}_k = e^{i\cdot\left(\frac{2\pi k}{d}\right)}, \quad k = 0, \dots, d-1,$$

ii) if $\lambda_0 = \rho \cdot e^{i\cdot\theta}$ *is an eigenvalue of* P*, then*

$$\lambda_k = \rho \cdot e^{i\cdot\left(\frac{2\pi k}{d}+\theta\right)}, \quad k = 1, \dots, d-1$$

are also eigenvalues of P*.*

A proof can be found on [5] at pages 23-26.

Remark: This means that the eigenvalues of a transition matrix are located in the

complex plane on concentric circumferences with radius $\rho \leq 1$, and on every circumference there is a multiple of d of eigenvalues (precisely d on the unitary circumference). The only exception is the degenerate circle ($\rho = 0$).

Below, in fig 3.1 there are some concrete examples.

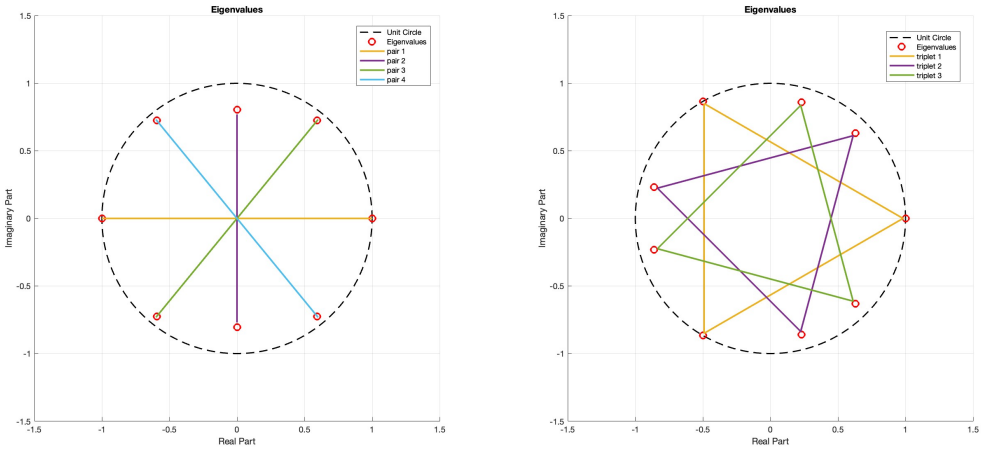


Figure 3.1: *The eigenvalues of two stochastic matrices, respectively of period two and three. The connected points represent eigenvalues belonging to the same circumference.*

Remark: Since the size of the matrix P, denoted by n, can be expressed as $n = h \cdot d + r$, where $h, r \in \mathbb{N}$ and $r < d$, it follows that $\dim(\ker(P)) \geq r$. Consequently, if $r \neq 0$, the matrix P is not invertible.

Definition 3.1.3. *$\lambda^* = 1$ is called Perron-Frobenius eigenvalue.*

Definition 3.1.4. *A discrete distribution π is called stationary distribution for P if:*

$$\pi = \pi \cdot P$$

Theorem 3.1.5. *If P is irreducible there exist a unique stationary distribution π.*

Proof. This result is well known, however, since the proof is not present in [1], we provide one.

Let n be the size of P and w be a probability density. We define the set:

$$S = \left\{ u \in \mathbb{R}^n \;\middle|\; u_i \geq 0, \sum_{i=1}^{n} u_i = 1 \right\}$$

which represents the set of discrete probability vectors of size n.

We construct the sequence:

$$u_m = \frac{1}{m} \sum_{k=0}^{m-1} uP^k$$

Note that since uP^k is a discrete probability vector for each k, the vector u_m, being a convex combination of such vectors, is itself a probability vector.

Since S is a closed and bounded subset of $\mathbb{R}^n$, it is compact.

Thus, the sequence $\{u_m\}$ has a subsequence $\{u_{m_k}\}$ that converges to a unique limit in S.

This limit π is stationary, meaning:

$$\pi P = \pi.$$

in fact:

$$\lim_{k\to\infty} \left(u_{m_k} P - u_{m_k} \right) = \left(\lim_{k\to\infty} u_{m_k} \right) P - \lim_{k\to\infty} u_{m_k} = \lim_{k\to\infty} \left(\left(\frac{1}{m_k} \sum_{\ell=0}^{m_k-1} uP^\ell \right) P - \frac{1}{m_k} \sum_{\ell=0}^{m_k-1} uP^\ell \right)$$

Which is a Mengoli Series, hence

$$\lim_{k\to\infty} \left(\frac{1}{m_k} (uP^{m_k} - u) \right) = \underline{0}$$

since $1/m_k$ is infinitesimal and $(uP^{m_k} - u)$ is bounded.

Thus, the limit $\lim_{k\to\infty} u_{m_k} = \pi$ satisfies:

$$\pi P = \pi.$$

As for unicity: Recall that for any square matrix P, both the algebraic and geometric multiplicities of each eigenvalue coincide with those of its transpose P^T.

Suppose for contradiction that there is another stationary distribution $\nu^T \neq \pi$. Then, by transposing the equation, ν is a right eigenvector of P^T corresponding to the eigenvalue 1. By theorem 3.1.2, the eigenspace of P^T associated to eigenvalue 1 is one-dimensional.

Hence ν^T must be a scalar multiple of π. Since both ν^T and π are probability distributions (they sum to 1), the only way they can be a scalar multiple of each other is if that factor equals 1. Thus $\nu^T = \pi$, contradicting our assumption that $\nu^T \neq \pi$.

Therefore, π is the unique stationary distribution of P.

□

3.2. Decomposing Transition Dynamics: Factorization and Stability

From now on, unless stated otherwise, we will always **assume P to be diagonalizable**. The reasoning behind this assumption will become clear later. In short, for diagonalizable matrices, the n-step transition probability for any state can be expressed as a linear combination of the eigenvalues raised to the nth power. By contrast, if P is not diagonalizable, then in the most general case every power up to n may appear in the expression, thus greatly complicating the algebra. As explained more thoroughly in the appendix, however, this assumption is not particularly restrictive.

We now introduce some foundational topics on diagonalizability; for a more rigorous and comprehensive discussion see [4].

Theorem 3.2.1. *Let $P \in \mathbb{C}^{n\times n}$ be a diagonalizable matrix. Then there exists an invertible matrix T such that:*

$P = TDT^{-1}$

where D is a diagonal matrix whose entries on the diagonal are the eigenvalues of P.

Corollary 3.2.2. *Let $P = TDT^{-1}$ be the diagonal decomposition of an $n \times n$ matrix P. Then, for any integer $k \geq 1$, the k-th power of P is given by:*

$$P^k = TD^kT^{-1}$$

Proof. Starting from $P = TDT^{-1}$:

$$P = TDT^{-1} \Rightarrow P^k = (TDT^{-1})^k = TDT^{-1} \cdot TDT^{-1} \cdots TDT^{-1} = TD^kT^{-1}$$

□

Therefore, the entry $(P^k)_{i,j}$ of the matrix P^k is given by:

$$(P^k)_{i,j} = \sum_{l=1}^{n} T_{i,l}\lambda_l^k \left(T^{-1}\right)_{l,j}$$

In particular, for the diagonal entries ($i = j$), we have:

$$(P^k)_{j,j} = \sum_{l=1}^{n} T_{j,l} \lambda_l^k \left(T^{-1}\right)_{l,j}$$

For a more rigorous discussion about diagonalizability and more we indicate [4].

Hence, for a stochastic matrix P, by expanding the matrix product we are able to express the probability of return in k steps from a state $j \in E$ as a linear combination of the eigenvalues of P each raised to the k:

$$(P^k)_{j,j} = \sum_{l=1}^{n} T_{j,l} \lambda_l^k \left(T^{-1}\right)_{l,j} = \sum_{l=1}^{n} \left(T_{j,l} \cdot T_{l,j}^{-1}\right) \cdot \lambda_l^k$$

Note that $T_{j,l} \cdot T_{l,j}^{-1}$ is simply a number.

From Theorem 3.1.2 and previous observations, we know that the **non-null eigenvalues** of P take the form:

$$\lambda_k = \rho \cdot e^{i \cdot \left(\frac{2\pi k}{d} + \theta\right)}, \quad k = 0, \ldots, d-1,$$

where:

- $\rho \in (0, 1]$ is the modulus of the eigenvalue
- $\theta \in [0, 2\pi)$ is the phase shift
- $\frac{2\pi k}{d}$ ensures the eigenvalues are evenly spaced on a circumference of radius ρ.

We introduce the class of eigenvalues associated with a given ρ and θ as:

$$\Lambda_{\rho,\theta} := \bigcup_{k=0}^{d-1} \lambda_k = \bigcup_{k=0}^{d-1} \rho \cdot e^{i \cdot \left(\frac{2\pi k}{d} + \theta\right)}.$$

This grouping highlights the structure of the eigenvalues in a d-periodic matrix:

- Each group $\Lambda_{\rho,\theta}$ corresponds to a circumference of radius ρ in the complex plane, centered at the origin.
- The eigenvalues in $\Lambda_{\rho,\theta}$ are **evenly spaced** by an angle of $\frac{2\pi}{d}$, forming a regular polygon with d vertices on the circumference, shifted by θ with respect to the roots of unity.

Remark: To fully characterize an eigenvalue class, specifying the modulus ρ alone is not sufficient. There may be multiple groups of eigenvalues with the same modulus but

different phase shifts.

The spectrum $\sigma(P)$ of the transition matrix P can then be written as:

$$\sigma(P) = \bigcup_{j=1}^{h} \Lambda_{\rho_j,\theta_j} \cup \bigcup_{j=1}^{r} \{0\}, \quad \text{with } \Lambda_{\rho,\theta} := \bigcup_{k=0}^{d-1} \rho \cdot e^{i\cdot\left(\frac{2\pi k}{d}+\theta\right)}.$$

for a certain h equal to the number of classes of non-null eigenvalues of P.
We are now allowed to write the explicit expression for $p_{j,j}^{(k)}$, according to the partition made, as:

$$p_{j,j}^{(k)} = \sum_{s=1}^{h}\sum_{t=0}^{d-1} \rho_s^k \cdot \alpha_{s,t} \cdot e^{ki\cdot(\frac{2\pi t}{d}+\theta_s)} + \sum_{t=1}^{r} \alpha_{0,t} \cdot 0^k = p_{j,j}^{(k)} = \sum_{s=1}^{h} \rho_s^k e^{ki\theta_s} \cdot \left[\sum_{t=0}^{d-1} \alpha_{s,t} \cdot e^{ki\cdot(\frac{2\pi t}{d})}\right]$$

where $\alpha_{s,t}$ are the coefficients uniquely associated to the eigenvalue of P in the class $\Lambda_{\rho_s,\theta_s}$ with phase $\frac{2\pi t}{d} + \theta_s$.
Is there a more significative way to rearrange this seemingly uninformative expression? We now introduce a final definition and are then ready to prove our result.

Definition 3.2.3. *Given $n \in \mathbb{N}$, we define simple cycle of size n, the directed graph corresponding to the transition matrix:*

$$C_n = \begin{bmatrix} \mathbb{O}_{(n-1,1)} & \mathbb{I}_{(n-1,n-1)} \\ \mathbb{I}_{(1,1)} & \mathbb{O}_{(1,n-1)} \end{bmatrix}$$

where $\mathbb{O}_{a\times b}$ is the null matrix of size $a \times b$ and $\mathbb{I}$ is the identity matrix. This Graph corresponds to a oriented "necklace" of size n.

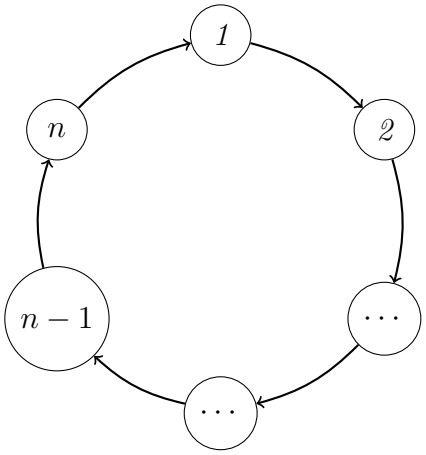


***Figure 3.2**: graph representation of a simple cycle of size n.*

In compact notation:

$$\{C_n\}_{i,j} = \delta_{(i,(j-1) \bmod n)}$$

This matrix is n-periodic, and thus, due to Theorem 3.1.2 its eigenvalues are the nth roots of unity.

Theorem 3.2.4. *Let P be an irreducible, diagonalizable d-periodic transition matrix, j a generic state and $k \in \mathbb{N}$, then, $\forall s \leq h$, there exists $\alpha_s \in \mathbb{C}$ such that*

$$\alpha_{s,t} = \alpha_s, \quad \text{for } t = 0, ..., d-1,$$

where $\alpha_{s,t}$ is the coefficient associated to the eigenvalue with phase $\frac{2\pi t}{d} + \theta_s$ in the class $\Lambda_{\rho_s,\theta_s}$ in the expression for $p_{j,j}^{(k)}$ mentioned above.

Proof. We start by writing explicitly the expression:

$$p_{j,j}^{(k)} = \sum_{s=1}^{h}\sum_{t=0}^{d-1} \rho_s^k \cdot \alpha_{s,t} \cdot e^{ki\cdot(\frac{2\pi t}{d}+\theta_s)} + \sum_{t=1}^{r} \alpha_{0,t} \cdot 0^k = p_{j,j}^{(k)} = \sum_{s=1}^{h} \rho_s^k e^{ki\theta_s} \cdot \left[\sum_{t=0}^{d-1} \alpha_{s,t} \cdot e^{ki\cdot(\frac{2\pi t}{d})}\right]$$

What we are trying to prove is that $\alpha_{s,t}$ does not depend on t.
We start by considering the specific case where the transition matrix corresponds to the simple cycle of size d (C_d), thus the eigenvalues correspond to the dth roots of unity. In this case we know that $p_{j,j}^{(k)} = 0$ if k is not a multiple of d and is equal to one otherwise. In order to determine the value of $\underline{\alpha}^T = (\alpha_1, \alpha_2, \ldots, \alpha_d)$, we set up the following linear system ($M\underline{\alpha} = \underline{b}$):

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^k & \lambda_2^k & \cdots & \lambda_d^k \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{d-1} & \lambda_2^{d-1} & \cdots & \lambda_d^{d-1} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_k \\ \vdots \\ \alpha_d \end{pmatrix} = \begin{pmatrix} 1 = p_{j,j}^{(0)} \\ 0 = p_{j,j}^{(1)} \\ 0 = p_{j,j}^{(k)} \\ \vdots \\ 0 = p_{j,j}^{(d-1)} \end{pmatrix}.$$

We now notice that M is a Vandermonde matrix of distinct elements over $\mathbb{C}$, hence $det(M) \neq 0$ and there exists at most one solution $\underline{\alpha}$. It is easy to check that $\underline{\alpha} = \underline{b}_1/d$ satisfies the system.
Note that a system of the form $M\underline{\alpha} = \underline{b}$ where M is a $d \times d$ Vandermonde matrix and b is of the form $[b_1, 0, ..., 0]$ always admits $\underline{\alpha} = \underline{b}_1/d$ as a solution. The idea in the following

steps is to inductively set up linear systems of this form for every circumference.

Now we consider the case where $1 = \rho_1 > \rho_2 > \cdots > \rho_s > \cdots > \rho_h$

$$p_{j,j}^{(k)} = \sum_{s=1}^{h} \rho_s^k \cdot [\sum_{t=0}^{d-1} \alpha_{s,t} \cdot e^{ki\cdot(\frac{2\pi t}{d}+\theta_s)}]$$

If $c \to \infty$, where $k = c \cdot d + \tau$ for some fixed $\tau < d$, we find $p_{j,j}^{(k)} = \sum_{t=0}^{d-1} \alpha_{1,t} \cdot e^{\tau i\cdot(\frac{2\pi t}{d}+\theta_1)}$
This is the case previously tackled of the simple cycle, hence we know that $\alpha_{1,t} = \alpha_1 \; \forall t \in \{0, 1, .., d-1\}$
Now we substitute the result found into the equation above and we find:

$$p_{j,j}^{(k)} = \rho_1^k \cdot \alpha_1 \cdot \sum_{t=0}^{d-1} e^{\tau i\cdot(\frac{2\pi t}{d}+\theta_1)} + \sum_{s=2}^{h} \rho_s^k \cdot [\sum_{t=0}^{d-1} \alpha_{s,t} \cdot e^{\tau i\cdot(\frac{2\pi t}{d}+\theta_s)}]$$

Having established equality for the top layer ($\rho_1 = 1$), we proceed by "peeling off" this layer, namely consider:

$$p_{j,j}^{(k)} - \rho_1^k \alpha_1 \sum_{t=0}^{d-1} e^{ik(\frac{2\pi t}{d}+\theta_1)} = \rho_2^k \cdot \sum_{s=2}^{h} (\frac{\rho_s}{\rho_2})^k \cdot [\sum_{t=0}^{d-1} \alpha_{s,t} \cdot e^{\tau i\cdot(\frac{2\pi t}{d}+\theta_s)}]$$

We now consider:

$$q_{j,j}^{(k)} := \frac{p_{j,j}^{(k)} - \rho_1^k \alpha_1 \sum_{t=0}^{d-1} e^{ik(\frac{2\pi t}{d}+\theta_1)}}{\rho_2^k} = \sum_{s=2}^{h} (\frac{\rho_s}{\rho_2})^k \cdot [\sum_{t=0}^{d-1} \alpha_{s,t} \cdot e^{\tau i\cdot(\frac{2\pi t}{d}+\theta_s)}]$$

Notice that, by the same reasoning as before, (fixing τ and sending $c \to \infty$) we observe that $\frac{p_{j,j}^{(k)} - \rho_1^k \alpha_1 \sum_{t=0}^{d-1} e^{ik(\frac{2\pi t}{d}+\theta_1)}}{\rho_2^k}$ satisfies the properties required to set up the previous Vandermonde system with $\underline{b} = [q_{j,j}^{c\cdot b}, q_{j,j}^{c\cdot b+1}, ..., q_{j,j}^{c\cdot b+\tau}, ..., q_{j,j}^{c\cdot b+d-1}]$ as right-hand side. In fact, if we now focus on $(\frac{\rho_2}{\rho_2})^k \cdot \sum_{t=0}^{d-1} \alpha_{2,t} \cdot e^{ik(\frac{2\pi t}{d}+\theta_2)}$, we notice that we are back to the case addressed above with index 1 instead of 2, hence:

$$\alpha_{2,0} = \alpha_{2,1} = \cdots = \alpha_{2,d-1}.$$

Repeating this argument inductively for each ρ_s, we find that all coefficients within each class $\Lambda_{\rho_s,\theta_s}$ are equal. Hence, for every s:

$$\alpha_{s,0} = \alpha_{s,1} = \cdots = \alpha_{s,d-1}.$$

Lastly, we consider the case where there exists $\Lambda_{\rho_1,\theta_1}$, $\Lambda_{\rho_2,\theta_2}$ with $\rho_1 = \rho_2 = \rho$ (There are two classes sharing the same modulus but different phase shifts). In this case, by the same reasoning as before, we can consider only this layer, ignoring the contribute of the inner layers (via the technique of analyzing the limit behaviour). Now we have $2d$ distinct eigenvalues and we can apply the Vandermonde argument again on the matrix M with kth-row $\lambda_{1,1}^k\ \lambda_{1,2}^k \cdots \lambda_{1,d}^k \quad \lambda_{2,1}^k\ \lambda_{2,2}^k \cdots \lambda_{2,d}^k$ and $b = q_{j,j}^{(0)}\ q_{j,j}^{(1)}\ \cdots q_{j,j}^{(2d-1)}$.

□

Our theorem (3.2.4) has shown that the coefficients $\alpha_{s,t}$ depend solely on s. Consequently, this allows for a compact factorization of the expression for $p_{j,j}^{(k)}$:

$$p_{j,j}^{(k)} = \sum_{s=1}^{h} \rho_s^k \cdot \alpha_s \cdot \left[\sum_{t=0}^{d-1} e^{ki\cdot\left(\frac{2\pi t}{d}+\theta_s\right)}\right] = \left[\sum_{s=1}^{h} \rho_s^k \cdot \alpha_s \cdot e^{ki\theta_s}\right] \cdot \left[\sum_{t=0}^{d-1} e^{ki\cdot\frac{2\pi t}{d}}\right].$$

This factorization reveals that the expression can be split into two distinct components:

1. **The second term:**
$$\Phi_j(k) = \sum_{t=0}^{d-1} e^{ki\cdot\frac{2\pi t}{d}}.$$
This term corresponds to the spectral structure of a simple cycle (eigenvalues distributed only on the circumference of radius 1). It is also the term that prevents the chain distribution to converge to the stationary, as 1 is the only value of ρ for which ρ^k does not tend to 0 if k tends to $+\infty$.
Properties of $\Phi_j(k)$: The term $\Phi_j(k)$ exhibits an interesting periodic structure. Specifically:
$$\Phi_j(k) = \begin{cases} 0 & \text{if } k \not\equiv 0 \pmod d, \\ d & \text{if } k \equiv 0 \pmod d. \end{cases}$$
This behavior stems from the fact that $\Phi_j(k)$ represents the sum of d-th roots of unity, which cancel out unless k is a multiple of d. Note that $\Phi_j(k)$ does not depend on the state j. In fact, it is easy to check that, if we considered another state j' , the expression of $\Phi_{j'}(k)$ would have been the same since it only encodes the periodic properties which are the same for all the states in the chain. For this reason we will omit the subindex j when referring to $\Phi(k)$.

2. **The first term:**
$$\Psi_j(k) = \sum_{s=1}^{h} \rho_s^k \cdot \alpha_s \cdot e^{ki\theta_s},$$

where $|\rho_s e^{i\theta_s}| \leq 1$. This term governs the decay and oscillatory behavior of the transition probabilities over k. As will be seen later, it is the reason why the chain does not exhibit purely simple cycle behaviour.

Non-Vanishing properties of $\Psi_j(k)$:
To compute the period effectively, $\Psi_j(k)$ must not vanish at every index where $\Phi(k) = d$. Intuitively, if $\Psi_j(k)$ were to vanish at all such moments, the probability $p_{j,j}^{(k)}$ would be zero, contradicting the requirement for a well-defined period.

For the period to be exactly d, there must exist two distinct coprime indices k_1 and k_2, such that:

$$\Phi(k_1 \cdot d) = \Phi(k_2 \cdot d) = d,$$

(always true) and:

$$\Psi_j(k_1), \Psi_j(k_2) \neq 0.$$

If such indices do not exist, the effective period could be a multiple of d, rather than d itself, contradicting the assumption of P being a d-periodic matrix. Note that $\Psi_j(k)$ depends on the coefficients α_s, which in turn depend on the state j that we used to derive the factorization. Consequently, we will always include the subscript j.

Now we prove a stronger claim than the one concerning the non-vanishing properties of $\Psi_j(k)$:

Theorem 3.2.5. *For every state j there exists $k^* \in \mathbb{N}$ (dependent on j) such that for every $k > k^*$, then $|\Psi_j(k)| > 0$, which means that $|\Psi|$ is eventually positive.*

Proof.

$$\Psi_j(k) = \sum_{s=1}^{h} \rho_s^k \cdot \alpha_s \cdot e^{ki\theta_s} = \alpha_1 + \sum_{s=2}^{h} \rho_s^k \cdot \alpha_s \cdot e^{ki\theta_s}$$

where $\alpha_1 = \pi_j$ is the coefficient corresponding to $\Lambda_{1,0}$ (Perron-Frobenius Eigenvalue) and π is the stationary distribution.

We will later show that $\Psi_j(k) \to \alpha_1$. Hence, we first need to prove that $\pi_i > 0 \; \forall i \in E$.

Since π is a discrete distribution, we know that $\exists j \in E$ such that $\pi_j > 0$.
Furthermore, since P is irreducible, we have $j \to i \;\; \Rightarrow \exists n^* \in \mathbb{N}$ such that $p_{j,i}^{(n^*)} > 0$.

At the same time:

$$\pi = \pi \cdot P^{n^*} \quad \Rightarrow \quad \pi_i = \sum_{l=1}^{n} \pi_l \cdot (P^{n^*})_{l,i}.$$

Using the above, we deduce:

$$\pi_i > \pi_j \cdot p_{j,i}^{(n^*)} > 0.$$

We want to show that $\Psi_j(k) \to \alpha_1$.

$$\Psi_j(k) \to \alpha_1 \quad \Leftrightarrow \quad \mathrm{Re}(\Psi_j(k)) \to \alpha_1$$

since $\mathbb{R}^+ \ni p_{j,j}^{(k)} = \Psi_j(k) \cdot \Phi(k)$, and since $\Phi(k) \in \mathbb{R} \Rightarrow \mathrm{Im}(\Psi_j(k)) = 0$.
Hence, it is sufficient to study $\mathrm{Re}(\Psi_j(k))$

$$|\mathrm{Re}(\Psi_j(k))| = \left|\alpha_1 + \sum_{s=2}^{h} \rho_s^k \cdot \alpha_s \cdot \cos(k\theta_s)\right| \geq \alpha_1 + \sum_{s=2}^{h} \left[\max_{l\in[2;h]} \rho_l\right]^k \cdot 1 \cdot (-1) = \alpha_1 - (h-2) \cdot \rho_{MAX}^k$$

but, since $0 < \rho_{MAX} < 1$, if $k > k^* := \left\lceil \log_{\rho_{MAX}} \left(\frac{\alpha_1}{h-2}\right) \right\rceil \Rightarrow |\mathrm{Re}(\Psi_j(k))| > 0 \Rightarrow \Psi_j(k) > 0$

□

This last theorem shows that the partial period eventually stabilizes. While Theorem 1.2.5 provided optimal upper bounds, the same concept reframed into the $(\Psi(k), \Phi(k))$ formalism shows that within $2n-1$ exist two coprime k_1, k_2 that make $\Psi(k)$ not vanish, offering a complementary perspective to close the circle.

List of Symbols

Symbol	Description
Γ	Oriented Graph
$\mathcal{N}(\Gamma)$	Nodes of Γ
$\mathcal{E}(\Gamma)$	Edges of Γ
$i \to j$	state i leads to state j
$i \leftrightarrow j$	state i communicates with state j
E	State Space
P	Transition matrix of a Markov Chain
p_{ij}	Probability of transition from state i to state j in 1 step
$p_{ij}^{(k)}$	Probability of transition from state i to state j in k steps
$\mathcal{H}_i$	Communication class of state i
$d(i)$	Period of state i
$d_k(i)$	kth Partial Period of state i
$d_k(\mathcal{H}_i)$	kth Partial Period of $\mathcal{H}_i$
λ	Eigenvalue
$\Lambda_{\rho,\theta}$	Eigenvalue class of radius ρ and angle shift θ

Bibliography

[1] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag, New York, 1999. doi: 10.1007/978-1-4757-3124-8. URL http://dx.doi.org/10.1007/978-1-4757-3124-8.

[2] S. Jukna. The pigeonhole principle. In *Extremal Combinatorics*, Texts in Theoretical Computer Science. An EATCS Series. Springer, Berlin, Heidelberg, 2011. doi: 10.1007/978-3-642-17364-6_4. URL https://doi.org/10.1007/978-3-642-17364-6_4.

[3] R. Kannan. The frobenius problem. In C. E. Veni Madhavan, editor, *Foundations of Software Technology and Theoretical Computer Science*, volume 405 of *Lecture Notes in Computer Science*. Springer, Berlin, Heidelberg, 1989. doi: 10.1007/3-540-52048-1_47. URL https://doi.org/10.1007/3-540-52048-1_47.

[4] E. Schlesinger. *Algebra Lineare e Geometria*. Zanichelli, Bologna, Italy, 2nd edition, 2017. ISBN 978-88-08-52069-2. URL https://www.zanichelli.it/ricerca/prodotti/algebra-lineare-e-geometria-schlesinger.

[5] E. Seneta. *Non-Negative Matrices and Markov Chains*. Springer, New York, 2 edition, 1981. doi: 10.1007/0-387-32792-4. URL https://doi.org/10.1007/0-387-32792-4.

Appendix A: On Diagonalizability of Irreducible Matrices

In Chapter 3 we assumed to work with diagonalizable stochastic matrices. However, it is important to highlight that non-diagonalizable stochastic matrices do exist. In particular, there exist stochastic matrices non-diagonalizable due to eigenspaces associated with eigenvalues other than 0, namely the cases in which the key results of the last chapter do not hold. The aim of this section is to provide a concrete method to construct such matrices and, perhaps, convince the reader of their rarity. Below we show an example:

$$P = \begin{bmatrix} \frac{48}{90} & \frac{21}{90} & \frac{21}{90} \\ \frac{41}{90} & \frac{38}{90} & \frac{11}{90} \\ \frac{1}{90} & \frac{31}{90} & \frac{58}{90} \end{bmatrix}$$

We do not burden the discussion with a verification of its non-diagonalizability, nor would it be useful to do so, as the reader will become convinced of the non-diagonalizability of this matrix–and infinitely many others—by learning how to construct them directly.

Constructing a Non-Diagonalizable 3×3 Matrix

Definition 3.2.6. *A **Jordan block** $J_k(\lambda)$ of size k corresponding to an eigenvalue λ is a $k \times k$ upper-triangular matrix where the diagonal entries are λ, the superdiagonal entries are 1, and all other entries are 0:*

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \lambda & 1 \\ 0 & \cdots & \cdots & 0 & \lambda \end{pmatrix}$$

Theorem 3.2.7. *Let $P \in \mathbb{C}^{n \times n}$. Then there exists an invertible matrix T such that:*

$P = TJT^{-1}$

where J is the Jordan canonical form of P, which is a block-diagonal matrix consisting of Jordan blocks $J_k(\lambda)$ corresponding to the eigenvalues λ of P.

We can now start the construction.

Let $\underline{x}$ be the uniform distribution vector:

$$\underline{x} = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}$$

and let $\underline{u}^T$ be the row vector of ones:

$$\underline{u}^T = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix}.$$

We can complete these vectors to form two matrices L and R by introducing vectors $\underline{y}, \underline{z}$ (column vectors) and $\underline{v}^T, \underline{w}^T$ (row vectors), such that:

$$L = \begin{pmatrix} \underline{x} & | & \underline{y} & | & \underline{z} \end{pmatrix}, \quad R = \begin{pmatrix} \underline{u}^T \\ \underline{v}^T \\ \underline{w}^T \end{pmatrix}.$$

These matrices are constructed to satisfy the relations:

$$LR = RL = I_3,$$

where I_3 is the 3×3 identity matrix.

Expanding the product explicitly, we have:

$$RL = \begin{pmatrix} \underline{u}^T \\ \underline{v}^T \\ \underline{w}^T \end{pmatrix} \begin{pmatrix} \underline{x} & | & \underline{y} & | & \underline{z} \end{pmatrix}.$$

Computing each entry using the inner product notation $\langle \cdot, \cdot \rangle$, we obtain:

$$RL = \begin{pmatrix} \langle \underline{u}, \underline{x} \rangle & \langle \underline{u}, \underline{y} \rangle & \langle \underline{u}, \underline{z} \rangle \\ \langle \underline{v}, \underline{x} \rangle & \langle \underline{v}, \underline{y} \rangle & \langle \underline{v}, \underline{z} \rangle \\ \langle \underline{w}, \underline{x} \rangle & \langle \underline{w}, \underline{y} \rangle & \langle \underline{w}, \underline{z} \rangle \end{pmatrix}.$$

Imposing $RL = I_3$, we obtain the following system of equations:

$$\begin{cases} \langle \underline{u}, \underline{x} \rangle = 1, & \langle \underline{u}, \underline{y} \rangle = 0, & \langle \underline{u}, \underline{z} \rangle = 0, \\ \langle \underline{v}, \underline{x} \rangle = 0, & \langle \underline{v}, \underline{y} \rangle = 1, & \langle \underline{v}, \underline{z} \rangle = 0, \\ \langle \underline{w}, \underline{x} \rangle = 0, & \langle \underline{w}, \underline{y} \rangle = 0, & \langle \underline{w}, \underline{z} \rangle = 1. \end{cases}$$

Now, we are ready to define the transition matrix P as:

$$P = L\tilde{J}R,$$

where $\tilde{J}$ is the "almost-Jordan" matrix:

$$\tilde{J} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & \mu & \lambda \end{pmatrix},$$

with λ, μ being real numbers.
expanding we get:

$$P = L\,\tilde{J}\,R = \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} u^T \\ \lambda\, v^T \\ \mu\, v^T + \lambda\, w^T \end{pmatrix}$$
$$= x\,u^T \;+\; \lambda\, y\, v^T \;+\; \mu\, z\, v^T \;+\; \lambda\, z\, w^T.$$

Given the construction, we must show three properties of $P = L\,\tilde{J}\,R$ for sufficiently small real parameters λ and μ:

- *P is non-diagonalizable with eigenvalue λ having a Jordan block.*
- *P is non-negative.*
- *P is a stochastic matrix.*

1. Non-diagonalizability and the Jordan block for λ.

Recall

$$\tilde{J} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & \mu & \lambda \end{pmatrix}$$

We now notice that $\tilde{J}$ can be decomposed in standard Jordan form as SJS^{-1} in the following:

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \mu^{-1} \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & \mu & 0 \end{pmatrix}$$

hence P can be decomposed as TJT^{-1} where $T = LS$, since $L = R^{-1}$.

2. Non-Negativity of P.

We write the expanded form

$$P \;=\; L\,\tilde{J}\,R \;=\; x\,u^T \;+\; \lambda\, y\, v^T \;+\; \mu\, z\, v^T \;+\; \lambda\, z\, w^T,$$

By construction xu^T is positive and thus, if λ, μ are chosen sufficiently small, all the entries remain non-negative.

3. P is stochastic.

A matrix is *row-stochastic* when each of its rows sums to 1. Equivalently if $Pu = u$ since, by the definition of the matrix product, $(Pu)_j$ is equal to the sum of the jth row. We recall that if $P = TJT^{-1}$, then the jth column of T is a right eigenvector associated to the jth eigenvalue. expanding the product $T = LS$ we find

$$T = \begin{pmatrix} \underline{x} & | & \underline{y} & | & \underline{z} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & \mu^{-1} \\ 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \underline{x} & | & \underline{z} & | & \mu^{-1} \cdot \underline{y} \end{pmatrix}$$

Thus, we see that x is a right eigenvector with eigenvalue 1, and since it is a scalar multiple of u, u is also a right eigenvector of eigenvalue 1.