2.6 Shift spaces and coding

In various of the examples of dynamical systems that we saw so far (the doubling map, the baker map and the Gauss map) we described an orbit through its *itinerary*. In this section we introduce some symbolic spaces that allow to describe the dynamics of more maps using itineraries.

Let us decompose the space X in finitely many pieces (more in general, one could consider countably many pieces), that is

$$X = P_1 \cup P_2 \cup \cdots \cup P_N.$$

If P_i are pairwise disjoint, we say that $\{P_1, \dots, P_N\}$ is a finite partition of X. Let $x \in X$. Since P_i cover X, for each $i \in \mathbb{N}$ there exists $1 \le a_i \le N$ such that $f^i(x) \in P_{a_i}$. The sequence $a_0, a_1, \dots, a_n, \dots$, is the itinerary of x with respect to $\{P_1, \dots, P_N\}$.

We can *code* the (forward) orbit $\mathcal{O}_f^+(x)$ with the sequence $\underline{a} = (a_i)_{i=0}^{\infty}$. The sequence belongs to

$$\Sigma_N^+ = \{1, \dots, N\}^{\mathbb{N}} = \{\ \underline{a} = (a_i)_{i=0}^{\infty}, \qquad 1 \le a_i \le N\},$$

that is the space of (one-sided) sequences in the digits $1, \dots, N$.

If f is invertible, for each $i \in \mathbb{Z}$ there exists $1 \le a_i \le N$ such that $f^i(x) \in P_{a_i}$ and we can code the full orbit $\mathcal{O}_f^+(x)$ with the full (past and future) itinerary $\underline{a} = (a_i)_{i=-\infty}^{\infty}$, which belongs to the space

$$\Sigma_N = \{1, \dots, N\}^{\mathbb{Z}} = \{\ \underline{a} = (a_i)_{i=-\infty}^{\infty}, \qquad 1 \le a_i \le N\},$$

that is the space of bi-sided sequences in the digits $1, \dots, N$.

In both cases, $f(x)$ is coded by the *shifted sequence*: since $f^i(f(x)) = f^{i+1}(x) \in P_{a_{i+1}}$ by definition of itinerary of x, the itinerary of $f(x)$, and hence the coding of $\mathcal{O}_f^+(f(x))$ is given by

$$\sigma^+((a_i)_{i=0}^{+\infty}) = (a_{i+1})_{i=0}^{+\infty}, \qquad \text{or, when } f \text{ is invertible, by} \qquad \sigma((a_i)_{i=-\infty}^{+\infty}) = (a_{i+1})_{i=-\infty}^{+\infty}.$$

The maps

$$\sigma^+ : \Sigma_N^+ \to \Sigma_N^+, \qquad \sigma : \Sigma_N \to \Sigma_N,$$

are known as *full* (one-sided) *shift* on N symbols and *full bi-sided shift* on N

If $\psi : X \to \Sigma_N^+$ (or $\psi : X \to \Sigma_N$ in the invertible case) is the *coding map* which assign to each point its itinerary, the previous relation shows that for all $x \in X$

$$\psi(f(x)) = \sigma^+(\psi(x)) \qquad (\text{or} \quad \psi(f(x)) = \sigma^+(\psi(x)) \quad \text{if } f \text{ is invertible}).$$

In order to give a conjugacy, though, the coding map ψ should be both injective and surjective. Thus, it is natural to ask:

(Q1) Is the coding unique?

(Q2) Do all sequences in Σ_N^+ (or in Σ_N) occur as possible itineraries?

The answer to both these questions is generally NO. In all the cases that we saw (doubling map, baker map, Gauss map) so far, all possible *finite*[1] sequences (in Σ_2^+ for the doubling map, in Σ_2 for the baker map and in countably many digits $\{1, \dots, n, \dots, \}^{\mathbb{N}}$ for the Gauss map) do occur, but as the Example 2.6.1 below shows, this is often not the case.

Example 2.6.1. Let $f : [0,1] \to [0,1]$ be the map give by

$$f(x) = \begin{cases} 2x & \text{if } 0 \le x < \frac{1}{2}, \\ x - \frac{1}{2} & \text{if } \frac{1}{2} \le x < 1. \end{cases}$$

whose graph is shown in Figure 2.1. Let $I_1 = [0, 1/2)$ and $I_2 = [1/2, 1]$. It is clear that if $x \in I_2$, then $f(x) \in I_1$. On the other hand, if $x \in I_1$, one could have either $f(x) \in I_1$ (if $x < 1/4$) or $f(x) \in I_2$ (if $1 < 4 \le x < 1/4$). Thus, one will never see two consecutive digits $2, 2$ in the itinerary, while all combinations $1, 1$, $1, 2$ and $2, 1$ can occur.

[1] Also in these examples there are countably many infinite sequences that do not occurr as itinerararies: for example, for the doubling map, if the coding partition is $[0, 1/2)$ and $[1/2, 1]$, all sequences which end with a tail of 1s do not appear as itinerary of any point.

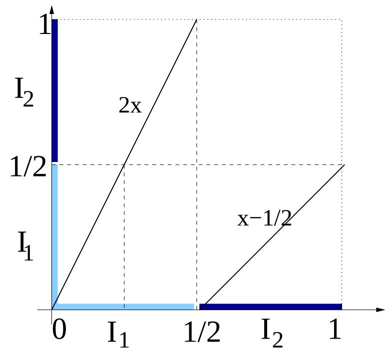


Figure 2.1: The map f in Example 2.6.1.

To be able to describe the subset of the shift that describes itineraries of this form is one of the reasons to study *subshifts of finite type* of the following form.

Definition 2.6.1. An $N \times N$ matrix is called an *transition matrix* (also called *incidence matrix*) if all entries A_{ij}, $1 \le i, j \le N$, are either 0 or 1.

One can use a matrix A to encore the information of which pairs of consecutive digits can appear in an itinerary: the digit i can be followed by the digit j if and only if the entry A_{ij} is equal to 1. More formally, we can consider the following subspaces $\Sigma_A^+ \subset \Sigma_N^+$ and $\Sigma_A \subset \Sigma_N$ of sequences

Definition 2.6.2. The shift spaces associated to a transition matrix A are:

$$\Sigma_A^+ = \{(a_i)_{i=0}^{+\infty} \in \Sigma_N^+, \quad A_{a_i a_{i+1}} = 1 \quad \text{for all } i \in \mathbb{N}\},$$
$$\Sigma_A = \{(a_i)_{i=-\infty}^{+\infty} \in \Sigma_N, \quad A_{a_i a_{i+1}} = 1 \quad \text{for all } i \in \mathbb{Z}\}.$$

Example 2.6.2. For example if A is the matrix

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

since the only zero entry is $A_{22} = 0$, the digit 2 cannot be followed by another digit 2, while all the other pairs of successive digits 12, 11 and 21 are allowed. Thus the sequences in Σ_A^+ (respectively Σ_A) are all sequences in the digits $1, 2$ (respectively all the bisided sequences) without any pairs of consecutive digits 2.

If $(a_i)_{i=0}^{+\infty} \in \Sigma_A^+$, also the shifted sequence $\sigma^+((a_i)_{i=0}^{+\infty})$ belongs to Σ_A^+, since if $A_{a_i a_{i+1}} = 1$ for all $i \in \mathbb{N}$, clearly also $A_{a_{i+1} a_{i+2}} = 1$ for all $i \in \mathbb{N}$ (in other words, if a pair of consecutive digits did not occurr in $\underline{a}$, it clearly does not occurr either in the shifted sequence). The same is true for bisided sequences: if $(a_i)_{i=-\infty}^{+\infty} \in \Sigma_A$, also the shifted sequence $\sigma((a_i)_{i=-\infty}^{+\infty}) \in \Sigma_A$. Thus, the spaces Σ_A and Σ_A^+ are *invariant* under the shift and we can consider the restriction of σ^+ and σ to this subspaces.

Definition 2.6.3. The restriction of the shift maps to

$$\sigma^+ : \Sigma_A^+ \to \Sigma_A^+, \qquad \sigma : \Sigma_A \to \Sigma_A,$$

are called a *topological Markov chain*[2] (or also a *subshift of finite type*) associated to the matrix A.

These are special examples of *subshifts*, that is restrictions of the shift to closed invariant spaces of Σ_N^+ (or Σ_N). In a topological Markov chain, the only type of restrictions on the sequences is of the form i *cannot be followed by* j, thus depend only on the previous digit[3].

[2]In probability, one studies Markov chains, which consists of a topological Markov chain with in addition a measure. We will define a measure which is invariant under the shift in one of the next lectures.

[3]More in general, one can define for example invariant spaces where certain combination of digits, also called *words* in the digits, are not allowed (for example forbidden words could be 2212 and 111, so there cannot be occurrences of 11, but not three consecutive digits 1). A subshift can be equivalently defined in terms of countably many forbidden words: no sequence in the subshift contains a forbidden word and any sequence in the complement does contain a forbidden word. If only a finite number of words are forbidden, we have a *subshift of finite type*. If the maximal lenght of forbidden words is $k+1$, the subshift is called a *k-step* subshift of finite type. Thus, topological Markov chains are 1−step subshifts of finite type.

It is very convenient to visualize sequences sequences in Σ_A as paths on a graph.

Definition 2.6.4. The graph $\mathscr{G}_A$ associated to the $N \times N$ transition matrix A is a graph with vertices are $v_1, \dots, v_N$, where v_i and v_j are connected by an arrow from v_i to v_j if and only if $A_{ij} = 1$.

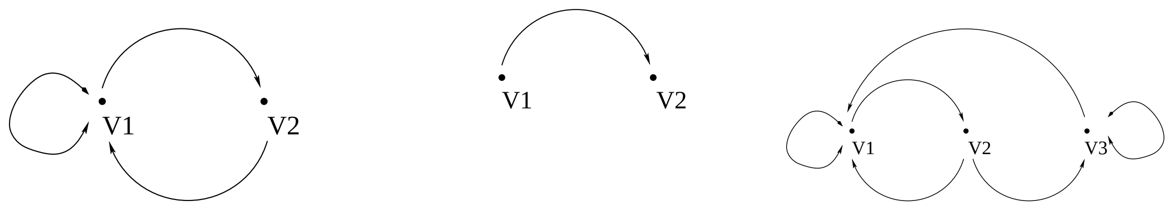
Then the following fact is immediate:

Lemma 2.6.1. *A sequence $(a_i)_{i=0}^{+\infty} \in \Sigma_N^+$ belongs to Σ_A^+ if and only if it describes a path on $\mathscr{G}_A$. Similarly a sequence $(a_i)_{i=-\infty}^{+\infty} \in \Sigma_N$ belongs to Σ_A if and only if it describes a bi-infinite path on $\mathscr{G}_A$.*

Example 2.6.3. For example for the following matrices:

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix}, \tag{2.1}$$

one obtains the graphs $\mathscr{G}_A$, $\mathscr{G}_B$ and $\mathscr{G}_C$ in Figure 2.2. A paths on $\mathscr{G}_A$ can never go through v_2 and immediately after v_2 again. Since there are no infinite paths on $\mathscr{G}_B$, we see that $\Sigma_B = \emptyset$.



(a) $\mathscr{G}_A$ (b) $\mathscr{G}_B$ (c) $\mathscr{G}_C$

Figure 2.2: The graphs associated to the transition matrices A, B, C in (2.3).

To avoid trivial cases like above, where $\Sigma_B = \emptyset$, but also to guarantee interesting dynamical properties, as we will see in the next lecture, one can ask conditions on the matrix as the following.

Definition 2.6.5. A transition matrix A is called *irredudible* if for any $1 \le i, j \le N$ there exists an $n \in \mathbb{N}$ (possibly dependent on i, j) such that the entry $(A^n)_{ij}$ of the matrix A^n obtained multiplying A by itself n times is *positive* ($(A^n)_{ij} > 0$).

A transition matrix A is called *aperiodic* (or also, in some books, *transitive*) if there exists an $n \in \mathbb{N}$ such that if for any $1 \le i, j \le N$ we have $(A^n)_{ij} > 0$.

A matrix A such that all entries $A_{ij} > 0$ is called *positive* (and we write $A > 0$). Thus, A is aperiodic if there exists a power $n \in \mathbb{N}$ such that A^n is positive.

Example 2.6.4. For example

$$A^2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}; \qquad \text{if } D = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad D^n = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}, \tag{2.2}$$

so A is irreducible and aperiodic (with $n = 2$) since all entries of A^2 are positive, while D is not irreducible, since for any n the entry $(D^n)_{21} = 0$.

Remark 2.6.1. Irreducible and aperiodic matrices can be easily recognized using the associated graph:

(1) A is *irreducible* if and only if for any two vertices v_i and v_j on $\mathscr{G}_A$ there exists a path connecting v_i to v_j;

(2) A is *aperiodic* if and only if there exists an n such that any two vertices v_i and v_j on $\mathscr{G}_A$ can be connected by a path of the same lenght n.

The proof of this remark will be given in the next section.

Example 2.6.5. The graph $\mathscr{G}_A$ in Figure 2.2 shows that A is irreducible, since all vertices can be connected: for example, v_2 can be connected to itself going through v_1. Moreover, it is aperiodic with $n = 2$, since the paths from v_2 to v_2 has lenght two and one can get paths of lenght 2 from v_1 to itself, from v_2 to v_1 and from v_1 to v_2 by repeating the loop around v_1. The graph $\mathscr{G}_B$ in Figure 2.2 shows that B is *not* irreducible, since there are no paths connecting for example v_1 to itself (or neither paths connecting v_2 to v_1 and v_2 to itself). The graph $\mathscr{G}_C$ in Figure 2.2 shows that C is irreducible, since one sees immediately that all vertices can be connected to each other.

Exercise 2.6.1. Consider the following transition matrices

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}. \tag{2.3}$$

Draw the corresponding graphs $\mathscr{G}_{A_i}$, $i = 1, 2$, associated to them. For each $i = 1, 2$ is A_i irreducible? is A_i aperiodic?

2.7 Dynamical properties of topological Markov chains

Let A be an $N \times N$ transition matrix and let $\sigma : \Sigma_A \to \Sigma_A$ be the corresponding topological Markov chain, defined in § 2.6. In this section we first show that Σ_A is a metric space and describe the open balls. Hence, $\sigma : \Sigma_A \to \Sigma_A$ is a topological dynamical system. Then we investigate the topological dynamical properties of $\sigma : \Sigma_A \to \Sigma_A$.

Metric, cylinders and balls for a shift space.

The shift spaces Σ_N^+ and Σ_N and their subshifts spaces are metric spaces with one of the following distances.

Let $\Sigma_N = \{1, \dots, N\}^{\mathbb{Z}}$ be the *full bi-sided shift space* on N symbols. For $\rho > 1$ consider the distance

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=-\infty}^{\infty} \frac{|x_k - y_k|}{\rho^{|k|}}, \qquad \text{where} \quad \underline{x} = (x_k)_{k=-\infty}^{+\infty}, \quad \underline{y} = (y_k)_{k=-\infty}^{\infty}. \tag{2.4}$$

The distance is well defined since $|x_k - y_k| \le N - 1$ for any $k \in \mathbb{N}$ (since both $x_k, y_k \in \{1, \dots, N\}$, their difference is at most $N - 1$) and the series $\sum_k 1/\rho^k$ is convergent for any $\rho > 1$, so that

$$d_\rho(\underline{x}, \underline{y}) \le \sum_{k=-\infty}^{\infty} \frac{N-1}{\rho^{|k|}} \le 2(N-1) \sum_{k=0}^{\infty} \frac{1}{\rho^k} < \infty.$$

Similarly, for the full *one-sided* shift space $\Sigma_N^+ = \{1, \dots, N\}^{\mathbb{N}}$ we can consider for any $\rho > 1$ the distance

$$d_\rho^+(\underline{x}, \underline{y}) = \sum_{k=0}^{\infty} \frac{|x_k - y_k|}{\rho^k}, \qquad \text{where} \quad \underline{x} = (x_k)_{k=0}^{\infty}, \quad \underline{y} = (y_k)_{k=0}^{\infty}. \tag{2.5}$$

Remark that if $\Sigma_A \subset \Sigma_N$ (respectively $\Sigma_A^+ \subset \Sigma_N^+$) is a subshift, the distance d_ρ in (2.4) (respectively the distance d_ρ^+ in (2.5), gives a distance on Σ_A (respectively Σ_A^+). More generally, all subshift spaces (subsets of Σ_N or Σ_N^+ invariant under the shift map σ) are metric spaces.

The following sets, called cylinders, play an essential role in the study of shift spaces:

Definition 2.7.1. A *cylinder* is a subset of Σ_N of the form

$$C_{-m,n}(a_{-m}, \dots, a_n) = \{ \ \underline{x} = (x_i)_{i=-\infty}^{+\infty} \in \Sigma_N, \ \text{ such that } \ x_i = a_i \quad \text{for all } -m \le i \le n \ \},$$

where $m, n \in \mathbb{N}$ and $a_i \in \{1, \dots, N\}$ for $-m \le i \le n$.

A cylinder is called *symmetric* if $n = m$.

Similarly a cylinder in Σ_N^+ is a subset of the form

$$C_n(a_0, \dots, a_n) = \{ \underline{x} = (x_i)_{i=0}^{+\infty} \in \Sigma_N^+, \text{ such that } x_i = a_i \text{ for all } 0 \le i \le n \},$$

where $n \in \mathbb{N}$ and $a_i \in \{1, \dots, N\}$ for $0 \le i \le n$.

Example 2.7.1. The following sequence

$$\underline{x} = \dots, x_{-m-2}, x_{-m-1}, \underbrace{a_{-m}, \dots, \overbrace{a_0}^{i=0}, \dots, a_n}_{fixed\ block}, x_{n+1}, x_{n+2} \dots$$

belongs to the cylinder $C_{-m,n}(a_{-m}, \dots, a_n)$. All points $(y_i)_{i=-\infty}^{+\infty}$ in $C_{-m,n}(a_{-m}, \dots, a_n)$ contain the fixed block of digits $a_{-m}, \dots, a_n$ centered at a_0 (for $i = 0$), while the tails can be any sequence in the digits $\{1, \dots, N\}$.

It two points belong to the same cylinder, they share a common central block of digits. Thus, it is clear that the distances in (2.4) and (2.5) are small. More is true. If ρ is chosen sufficiently large, then symmetric cylinders are *exactly* balls with respect to the distance d_ρ:

Lemma 2.7.1. *(1) If $\rho > N$ and d_ρ^+ is the distance on the one-sided shift space Σ_N^+, than for any $\epsilon = 1/\rho^n$ we have:*

$$C_{0,n}(x_0, \dots, x_n) = B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right),$$

where $\underline{x} = (x_i)_{i=0}^{\infty} \in \Sigma_N$ be a sequence which contains the central block $x_0, \dots, x_n$.

(2) If $\rho > 2N - 1$ and d_ρ is the distance on the two-sided shift space Σ_N,, than for any $\epsilon = 1/\rho^n$ we have:

$$C_{-n,n}(x_{-n}, \dots, x_n) = B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right),$$

where $\underline{x} = (x_i)_{i=-\infty}^{\infty} \in \Sigma_N$ be a sequence which contains the central block $x_{-n}, \dots, x_n$.

Proof. Let us show Part (1). Let $C_{0,n}(x_0, \dots, x_n)$ be a cylinder in $\Sigma_N^+ = \{1, \dots, N\}^{\mathbb{N}}$. Let $\underline{x} = (x_i)_{i=0}^{\infty} \in \Sigma_N^+$ be a sequence which begins with $x_0, \dots, x_n$, so that $\underline{x} \in C_{0,n}(x_0, \dots, x_n)$.

If also $\underline{y} \in C_{0,n}(x_0, \dots, x_n)$, then, since $|x_k - y_k| = 0$ for all $0 \le k \le n$, we have

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=n+1}^{\infty} \frac{|x_k - y_k|}{\rho^k} \qquad \le \frac{1}{\rho^{n+1}} \sum_{j=0}^{\infty} \frac{|x_{n+1+j} - y_{n+1+j}|}{\rho^j}.$$

Remark that, since both $x_i, y_i \in \{1, \dots, N\}$, for any $i \in \mathbb{N}$ we have $|x_i - y_i| \le N - 1$. Thus, using the formula for the sum of the geometric progression, we get

$$d_\rho(\underline{x}, \underline{y}) \le \frac{1}{\rho^{n+1}} \sum_{j=0}^{\infty} \frac{N-1}{\rho^j} = \frac{N-1}{\rho^{n+1}} \frac{1}{1 - \frac{1}{\rho}} = \frac{N-1}{\rho^n} \frac{1}{\rho - 1}.$$

Thus, if

$$\left(\frac{N-1}{\rho^n} \frac{1}{\rho - 1} \le \frac{1}{\rho^n} \quad \Leftrightarrow \quad \frac{N-1}{\rho - 1} \le 1\right) \qquad \Rightarrow \qquad \left(d_\rho(\underline{x}, \underline{y}) \le \frac{1}{\rho^n} \quad \Leftrightarrow \quad \underline{y} \in B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right)\right).$$

Thus, if $\rho \ge N$ to have the inclusion

$$C_{0,n}(x_0, \dots, x_n) \subset B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right).$$

Let us check that this condition is also sufficient to have the reverse inclusion. Assume that $\underline{y} \in B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right)$. Then, if by contradiction $\underline{y} \notin C_{0,n}(x_0, \ldots, x_n)$, there exists $0 \le j \le n$ such that $x_j \neq y_j$, so that $|x_j - y_j| \ge 1$. But then

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=0}^{\infty} \frac{|x_k - y_k|}{\rho^k} \ge \frac{|x_j - y_j|}{\rho^j} \ge \frac{1}{\rho^j} \ge \frac{1}{\rho^n}, \qquad \text{since } 0 \le j \le n.$$

Thus, $\underline{y} \notin B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right)$, which is a contradiction. Thus, we also have (without any additional assumption on ρ) the opposite inclusion

$$B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right) \subset C_{0,n}(x_0, \ldots, x_n).$$

Let us now prove Part (2) Let $C_{-n,n}(x_{-n}, \ldots, x_n)$ be a symmetric cylinder in Σ_N. Since $\underline{x} = (x_i)_{i=-\infty}^{\infty} \in \Sigma_N$ contains the central block $x_{-n}, \ldots, x_n$, the point $\underline{x} \in C_{-n,n}(x_{-n}, \ldots, x_n)$.

If also $\underline{y} \in C_{-n,n}(x_{-n}, \ldots, x_n)$, then, since $|x_k - y_k| = 0$ for all $k \in \mathbb{Z}$ with $|k| \le n$, we have

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=-\infty}^{-n-1} \frac{|x_k - y_k|}{\rho^{|k|}} + \sum_{k=n+1}^{\infty} \frac{|x_k - y_k|}{\rho^{|k|}}.$$

Remark that, since both $x_i, y_i \in \{1, \ldots, N\}$, for any $i \in \mathbb{N}$ we have $|x_i - y_i| \le N - 1$. Thus, using also the formula for the sum of the geometric progression, that gives us

$$\sum_{j=0}^{\infty} \frac{1}{\rho^j} = \frac{1}{1 - \frac{1}{\rho}} = \frac{\rho}{\rho - 1},$$

we get

$$d_\rho(\underline{x}, \underline{y}) \le 2 \sum_{k=n+1}^{\infty} \frac{N-1}{\rho^k} \le \frac{2(N-1)}{\rho^{n+1}} \sum_{j=0}^{\infty} \frac{1}{\rho^j} = \frac{2(N-1)}{\rho^{n+1}} \frac{\rho}{\rho - 1} = \frac{2(N-1)}{\rho^n} \frac{1}{\rho - 1}.$$

Thus, since

$$d_\rho(\underline{x}, \underline{y}) \le \frac{2(N-1)}{\rho^n} \frac{1}{\rho - 1} < \frac{1}{\rho^n} \quad \Leftrightarrow \quad \frac{2(N-1)}{\rho - 1} < 1 \quad \Leftrightarrow \quad \rho > 2N - 1,$$

if $\rho > 2N - 1$, we have

$$d_\rho(\underline{x}, \underline{y}) < \frac{1}{\rho^n} \quad \Leftrightarrow \quad \underline{y} \in B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right).$$

This proves that if $\rho > 2N - 1$ we have the inclusion

$$C_{-n,n}(x_{-n}, \ldots, x_n) \subset B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right).$$

Let us check the reverse inclusion. Assume that $\underline{y} \in B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right)$. Then, if by contradiction $\underline{y} \notin C_{-n,n}(x_{-n}, \ldots, x_n)$, there exists $j \in \mathbb{Z}$ with $|j| \le n$ such that $x_j \neq y_j$, so that $|x_j - y_j| \ge 1$. But then

$$d_\rho(\underline{x}, \underline{y}) = \sum_{k=-\infty}^{\infty} \frac{|x_k - y_k|}{\rho^{|k|}} \ge \frac{|x_j - y_j|}{\rho^{|j|}} \ge \frac{1}{\rho^{|j|}} \ge \frac{1}{\rho^n}, \qquad \text{since } 0 \le |j| \le n.$$

Thus, $\underline{y} \notin B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right)$, which is a contradiction. Thus, we also have (without any additional assumption on ρ) the opposite inclusion

$$B_{d_\rho}\left(\underline{x}, \frac{1}{\rho^n}\right) \subset C_{-n,n}(x_{-n}, \ldots, x_n).$$

□

Consider now a subshift $\Sigma_A \subset \Sigma_N$ determined by the transition matrix A (or the one-sided subshift $\Sigma_A^+ \subset \Sigma_N^+$).

Definition 2.7.2. A cylinder $C_{-m,n}(a_{-m},\dots,a_n)$ (where $n,m \in \mathbb{N}$ and $a_i \in \{1,\dots,N\}$ for $-m \le i \le n$) is called *admissible* if $A_{a_i,a_{i+1}} = 1$ for all $-m \le i < n$.

Similarly, a cylinder $C_{0,n}(a_0,\dots,a_n)$ (where $n \in \mathbb{N}$ and $a_i \in \{1,\dots,N\}$ for $0 \le i \le n$) is called *admissible* if $A_{a_i,a_{i+1}} = 1$ for all $-m \le i < n$.

Remark 2.7.1. If A is irreducible, a cylinder is *admissible* if and only if it is not empy. Indeed, the condition $A_{a_i,a_{i+1}} = 1$ for all $-m \le i < n$ guarantees that there is a path on $\mathscr{G}_A$ described by $a_{-m},\dots,a_n$ (that is passing in order through the vertices $v_{a_{-m}},\dots,v_{a_n}$) and since A is irreducible, one can continue this path to a biinfinite path in $\mathscr{G}_A$ (adding any admissible forward tail starting from a_n and any admissible backward tail before a_{-m}). This path belong to the cylinder and shows that it is not empy.

Number of paths and periodic points

Recall that in the previous section we interpreted the subshift $\Sigma_A \subset \{1,\dots,N\}^{\mathbb{Z}}$ as the space of bi-infinite paths on the graph $\mathscr{G}_A$ associated to the transition matrix A. The following Lemma turns out to be very helpful to study dynamical properties.

Lemma 2.7.2 (Number of paths). *For any $1 \le i,j \le N$, the number of paths of length n on $\mathscr{G}_A$ (that is, paths obtained composing n arrows) starting from the vertex v_i and ending in the vertex v_j is given by $(A^n)_{ij}$ (recall that A^n_{ij} is the (i,j) entry of the matrix A^n obtained producing A by itself n times).*

Proof. Let us prove it by induction on n. For $n = 1$, the paths of length 1 connecting v_i to v_j are simply arrows from v_i to v_j. By definition of $\mathscr{G}_A$ there is an arrow from v_i to v_j if and only if $A_{ij} = 1$, thus the statement for $n = 1$ holds.

Assume we proved it for n. Then, the number of paths of length $n+1$ starting from the vertex v_i and ending in the vertex v_j can be obtained by considering all paths of length $n+1$ starting from the vertex v_i and ending in any of the other vertices v_k, where $1 \le k \le N$ and extending each to a path of length $n+1$ ending in v_j if there is an arrow from v_k to v_j. Using that by inductive assumption the number of paths of length n starting from v_i and ending in v_k is given by A^n_{ij}, this gives

$$
\begin{aligned}
&Card\{ \text{ paths of length } n+1 \text{ from } v_i \text{ to } v_j \ \} \\
&\quad = \sum_{k=1}^{N} Card\{ \text{ paths of length } n \text{ from } v_i \text{ to } v_k \ \}\, Card\{ \text{ arrows from } v_k \text{ to } v_j \ \} \\
&\quad = \sum_{k=1}^{N} A^n_{ik} A_{kj} = A^{n+1}_{ij},
\end{aligned}
$$

where in the latter equation we simply used the definition of (i,j) entry of the product matrix $A^n A$. □

The Lemma has the following immediate Corollary on the number of periodic points of a topological Markov chain.

Corollary 2.7.1. *The cardinality of periodic points of period n for $\sigma : \Sigma_A \to \Sigma_A$ is exactly the trace $Tr(A^n)$.*

(Recall that the trace of a matrix $Tr(A) = \sum_i A_{ii}$ is the sum of the diagonal entries of A.)

Proof. If $\underline{x}$ is a periodic points of period n for $\sigma : \Sigma_A \to \Sigma_A$, then $\sigma^n(\underline{x}) = \underline{x}$, which implies that the digits of the sequence $\underline{x} = (x_i)_{i=-\infty}^{+\infty}$ have period n, that is $x_{n+i} = x_i$ for all $i \in \mathbb{Z}$. Thus, the path described by $\underline{x}$ on $\mathscr{G}_A$ is a periodic path, that repeats periodically the path starting from some v_i and coming back to the same v_i. Since the paths of length n connecting v_i to v_i are A^n_{ii} by Lemma 2.7.2, we have

$$
\begin{aligned}
&Card\{\underline{x} = (x_i)_{i=-\infty}^{+\infty} \quad \text{such that } \sigma^n(\underline{x}) = \underline{x}\} = \\
&\quad = \sum_{i=1}^{N} Card\{\underline{x} \quad \text{such that } \sigma^n(\underline{x}) = \underline{x} \quad \text{and} \quad x_0 = i\} = \sum_{i=1}^{N} A^n_{ii} = Tr(A^n),
\end{aligned}
$$

which is the desired expression. □

Topological transitivity and topological mixing

In the previous section we defined when a transition matrix A is *irreducible* and when it is *aperiodic* (also called transitive or primitive in certain books). The following equivalent definition in terms of existence of paths connecting two vertices v_i and v_j (already stated in Remark 2.6.1) follows from Lemma 2.7.2.

Corollary 2.7.2. *A transition matrix is* irreducible *if and only if for any two vertices v_i and v_j on $\mathscr{G}_A$ there exists a path connecting v_i to v_j and it is* aperiodic *if and only if there exists an n such that any two vertices v_i and v_j on $\mathscr{G}_A$ can be connected by a path of the same lenght n.*

Proof. By Lemma 2.7.2, $A^n_{ij} > 0$ if and only if there exists a path of lenght n in $\mathscr{G}_A$ from v_i to v_j. The Corollary then follows from the definition of irreducible and aperiodic. □

The dynamical significance of these definitions lie in the following Theorem.

Theorem 2.7.1. *Consider the metric space (Σ_A, d_ρ) where d_ρ is the metric defined in 2.4 and $\rho > 2N - 1$. Let $\sigma : \Sigma_A \to \Sigma_A$ be a topological Markov chain.*

(1) If the matrix A is irreducible then $\sigma : \Sigma_A \to \Sigma_A$ is topologically transitive;

(2) If A is aperiodic, then $\sigma : \Sigma_A \to \Sigma_A$ is topologically mixing.

Let us first prove a Lemma which will be used in the proof of the Theorem.

Lemma 2.7.3. *If $A^n > 0$ for some $n > 0$, then for any $m \geq n$ we also have $A^m > 0$.*

Proof. Remark first that if $A^n > 0$ for some $n > 0$, this means that for each j there exists a k_j such that $A_{k_j j} = 1$. Otherwise, if $A_{kj} = 0$ for all $1 \leq k \leq N$, then the vertex v_j cannot be reached from any other vertex v_k, so there cannot exist any path of length n reaching v_j, in contradiction with the fact that $A^n_{ij} > 0$.

Let us now prove by induction on m that $A^m > 0$ for $m \geq n$. For $m = n$ it is true by assumption. If we verified it for m, take any $1 \leq i, j \leq N$. Then, by the previous remark there exists k_j such that $A_{k_j j} = 1$. Moreover, for all the other k we have $A_{kj} \geq 0$. Hence, we get

$$A^{m+1}_{ij} = \sum_{k=1}^{N} A^m_{ik} A_{kj} \geq A^m_{ik_j} A_{k_j j} = A^m_{ik_j}$$

and $A^m_{ik_j} > 0$ since $A^m > 0$ the inductive assumption. This shows that $A^{m+1} > 0$ and concludes the proof. □

Proof of Theorem 2.7.1. Let us prove the first implication in (1). Assume that A is irreducible. We want to show that for each U, V non-empty open sets there exists $M > 0$ such that $\sigma^M(U) \cap V \neq 0$. Each open sets contains an open ball of the form $B_{d_\rho}(\underline{x}, \rho^{-k})$ for some large k and, since $\rho > 2N - 1$, by Lemma 2.7.1 each ball of this form is a non-empty symmetric cylinder. Hence, there exists two admissible cylinders

$$C_{(-k,k)}(a_{-k}, \dots, a_k) \subset U, \qquad C_{(-l,l)}(b_{-l}, \dots, b_l) \subset V. \tag{2.6}$$

Let us now construct a point $\underline{x}$ which contains both blocks of digits $a_{-k}, \dots, a_k$ and $b_{-l}, \dots, b_l$. By definition of irreducibility, taking $i = a_k$ and $j = b_{-l}$, there exists $n > 0$ such that $A^n_{a_k, b_{-l}} > 0$. This means that there exists a path of length n which connects v_{a_k} to $v_{b_{-1}}$. Let us denote by

$$y_0 = a_k, y_1, y_2, \dots, y_{n-1}, y_n = b_{-l},$$

the digits which describe this path. Clearly $A_{y_i y_{i+1}} = 1$ for all $0 \leq i \leq n - 1$. Consider a point $\underline{x} \in \Sigma_A$ such that

$$\underline{x} = \dots a_{-k}, \dots, \underbrace{a_0}_{i=0}, \dots, a_k, y_1, \dots, y_{n-1}, b_{-l}, \dots, b_l, \dots$$

(such point exist since by irreducibility we can choose a backward and a forward tail by choosing any path on $\mathscr{G}_A$ which starts from b_l (for the forward tail) or ends in a_{-k} (for the backward tail). Clearly, since $\underline{x}$ contains as central block of digits $a_{-k}, \ldots, a_k$, we have $\underline{x} \in C_{(-k,k)}(a_{-k}, \ldots, a_k) \subset U$. Moreover, if we set $M = n+k+l$, shifting the sequence $k+n+l$ times to the left, since $x_{k+n+l} = b_0$, we get

$$\sigma^M(\underline{x}) = \ldots b_{-l}, \ldots, \underbrace{b_0}_{i=0}, \ldots, b_l, \ldots,$$

so that $\sigma^M(\underline{x}) \in C_{(-l,l)}(b_{-l}, \ldots, b_l) \subset V$. This shows that

$$\underline{x} \in U \cap \sigma^{-M}(V) \neq \emptyset \qquad \Leftrightarrow \qquad \sigma^M U \cap V \neq \emptyset.$$

This shows that $\sigma : \Sigma_N \to \Sigma_N$ is topologically transitive.

Let us now prove (2). Assume that $A^n > 0$. We want to show that $\sigma : \Sigma_A \to \Sigma_A$ is topologically mixing. Let U, V be non empty open sets. We seek M_0 such that for any $M \geq M_0$ we have $\sigma^M(U) \cap V \neq \emptyset$. We can reason very similarly to part (1). Both U, V contain admissible symmetric cylinders of the form (2.6). Let $M_0 = n + k + l$. If $M \geq M_0$, then $M = m + k + l$ with $m \geq n$. Then also $A^m > 0$ by Lemma 2.7.3, so $A^m_{a_k, b_{-l}} > 0$. Thus, there exists a path of length m from v_{a_k} and $v_{b_{-l}}$, so we can construct a point in Σ_A of the form

$$\underline{x} = \ldots a_{-k}, \ldots, \underbrace{a_0}_{i=0}, \ldots, a_k, y_1, \ldots, y_{m-1}, b_{-l}, \ldots, b_l, \ldots$$

Reasoning as in Part (1), $\underline{x} \in U \cap \sigma^{-M}(V)$, so that $\sigma^M(U) \cap V \neq \emptyset$. This can be repeated for any $M \geq M_0$, showing that $\sigma : \Sigma_A \to \Sigma_A$ is topologically mixing. □

