On the Structure of Multilayer Cellular Neural Networks: Complexity between Two Layers

Jung-Chao Ban[*]
Department of Applied Mathematics, National Dong Hwa University
Hualian 97401, Taiwan, ROC

Chih-Hung Chang[†]
Department of Applied Mathematics, National University of Kaohsiung
Kaohsiung 81148, Taiwan, ROC
[*]*jcban@mail.ndhu.edu.tw*
[†]*Corresponding author: chchang@nuk.edu.tw.*

Let $\mathbf{Y}$ be the solution space of an *n*-layer cellular neural network, and let $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$ be the hidden spaces, where $1 \leq i, j \leq n$. ($\mathbf{Y}^{(n)}$ is called the output space.) The classification and the existence of factor maps between two hidden spaces that reach the same topological entropies are investigated in [1]. This paper elucidates the existence of factor maps between those hidden spaces carrying distinct topological entropies. For either case, the Hausdorff dimensions dim $\mathbf{Y}^{(i)}$ and dim $\mathbf{Y}^{(j)}$ can be calculated. Furthermore, the dimensions of $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$ are related by the factor map between them.

1. Introduction

Multilayer cellular neural networks (MCNNs) are large aggregates of analog circuits presenting themselves as arrays of identical cells that are locally coupled. MCNNs have been widely applied in studying the signal propagation between neurons, as well as in image processing, pattern recognition and information technology [2–11]. A one-dimensional MCNN is realized as

$$\frac{dx_i^{(\ell)}}{dt} = -x_i^{(\ell)} + \sum_{|k| \leq d} a_k^{(\ell)} y_{i+k}^{(\ell)} + \sum_{|k| \leq d} b_k^{(\ell)} u_{i+k}^{(\ell)} + z^{(\ell)}, \tag{1}$$

for some $d \in \mathbb{N}$, $1 \leq \ell \leq n \in \mathbb{N}$, $i \in \mathbb{Z}$, where

$$\begin{aligned} u_i^{(\ell)} &= y_i^{(\ell-1)} \text{ for } 2 \leq \ell \leq n, \\ u_i^{(1)} &= u_i, \ x_i^{(\ell)}(0) = x_{i,0}^{(\ell)}, \end{aligned} \tag{2}$$

and

$$y = f(x) = \frac{1}{2}(|x + 1| - |x - 1|) \tag{3}$$

is the output function.

The stationary solutions $\bar{x} = (\bar{x}_i^{(\ell)})$ of equation (1) are essential for understanding the system, and their outputs $\bar{y}_i^{(\ell)} = f(\bar{x}_i^{(\ell)})$ are called *output patterns*. A *mosaic solution* $(\bar{x}_i^{(\ell)})$ is a stationary solution satisfying $|\bar{x}_i^{(\ell)}| > 1$ for all i, ℓ, and the output of a mosaic solution is called a mosaic output pattern. Mosaic solutions are crucial for studying the complexity of equation (1) due to their asymptotical stability [12–23]. In an MCNN system, the "status" of each cell is taken as an input for a cell in the next layer, except for those cells in the last layer. The results that can be recorded are the output of the cells in the last layer. Since the phenomena that can be observed are only the output patterns of the nth layer, the nth layer of equation (1) is called the *output layer*, while the other $n - 1$ layers are called *hidden layers*.

We remark that except from mosaic solutions exhibiting key features of MCNNs, mosaic solutions themselves are constrained by the so-called "separation property" [24, 25]. This makes the investigation more difficult. Furthermore, the output patterns of mosaic solutions of an MCNN can be treated as a cellular automaton. For a discussion of systems satisfying constraints and cellular automata, readers are referred to Wolfram's celebrated book [26]. (Constrained systems are discussed in Chapter 5.)

Suppose $\mathbf{Y}$ is the solution space of an MCNN. For $\ell = 1, 2, \ldots, n$, let

$$\mathbf{Y}^{(\ell)} = \{\cdots y_{-1}^{(\ell)} y_0^{(\ell)} y_1^{(\ell)} \cdots\}$$

be the space that consists of patterns in the ℓth layer of $\mathbf{Y}$, and let $\phi^{(\ell)} : \mathbf{Y} \to \mathbf{Y}^{(\ell)}$ be the projection map. Then $\mathbf{Y}^{(n)}$ is called the *output space* and $\mathbf{Y}^{(\ell)}$ is called the (ℓ^{th}) *hidden space* for $\ell = 1, 2, \ldots, n - 1$. It is natural to ask whether there exists a relation between $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$ for $1 \le i \ne j \le n$. Take $n = 2$ for instance; the existence of a map connecting $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ that commutes with $\phi^{(1)}$ and $\phi^{(2)}$ means the *decoupling* of the solution space $\mathbf{Y}$. More precisely, if there exists $\pi_{12} : \mathbf{Y}^{(1)} \to \mathbf{Y}^{(2)}$ such that $\pi_{12} \circ \phi^{(1)} = \phi^{(2)}$, then π_{12} enables the investigation of structures between the output space and hidden space. A serial work is contributed for this purpose.

At the very beginning, Ban et al. [27] demonstrated that the output space $\mathbf{Y}^{(n)}$ is topologically conjugated to a one-dimensional sofic shift. This result is differentiated from earlier research that indicated that

the output space of a one-layer cellular neural network (CNN) without input is topologically conjugated to a Markov shift (also known as a shift of finite type). Some unsolved open problems, either on the mathematical or on the engineering side, have drawn interest since then. An analogous argument asserts that every hidden space $\mathbf{Y}^{(\ell)}$ is also topologically conjugated to a sofic shift for $1 \le \ell \le n-1$, and the solution space $\mathbf{Y}$ is topologically conjugated to a subshift of finite type. More than that, the topological entropy and dynamical zeta function of $\mathbf{Y}^{(\ell)}$ and $\mathbf{Y}$ are capable of calculation. This is a novel phenomenon, the asymmetry of topological entropy. It is known that a nonempty insertive and extractive language $\mathcal{L}$ is regular if and only if $\mathcal{L}$ is the language of a sofic shift; namely, $\mathcal{L} \subseteq \bigcup_{i\ge 0} B_i(X)$ for some sofic shift X, where

$$B_i(X) = \{x_1 x_2 \ldots x_i : (x_k)_{k\in\mathbb{Z}} \in X\}.$$

Therefore, elucidating sofic shifts is equivalent to the investigation of regular languages. Readers are referred to [28] and the references therein for more details about the illustration of languages and sofic shifts.

Followed by [1], the classification of the hidden and output spaces is revealed for those spaces reaching the same topological entropy. Notably, the study of the existence of $\pi_{ij} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(j)}$ for some i, j is equivalent to illustrating whether there is a map connecting two sofic shifts. Mostly it is difficult to demonstrate the existence of such maps. The authors have provided a systematic strategy for determining whether there exists a map between $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$. More than that, the explicit expression of π_{ij} is unveiled whenever there is a factor-like matrix E (defined later).

The present paper, as a continuation of [1, 27], is devoted to investigating the Hausdorff dimensions of the output and hidden spaces. We emphasize that in this elucidation, those spaces need not attain the same topological entropy. In addition to examining the existence of maps between $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$ (for the case where the topological entropies of two spaces are distinct), the complexity of the geometrical structure is discussed. The Hausdorff dimension of a specified space is an icon that unveils the geometrical structure and helps describe the complexity.

Furthermore, aside from the existence of factor maps between $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$, the correspondence of the Hausdorff dimension is of interest to this study. Suppose there exists a factor map $\pi_{ij} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(j)}$: then the Hausdorff dimensions of $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$ are related under some additional conditions (see Theorems 2 and 3). More explicitly, it is now known that in many examples the calculation of the Hausdorff dimen-

sion of a set is closely related to the maximal measures (defined later) of its corresponding symbolic dynamical system (cf. [29, Theorem 13.1], for instance). Theorems 2 and 3 also indicate that the Hausdorff dimension of $\mathbf{Y}^{(j)}$ is the quotient of the measure-theoretic entropy $h_{\pi_{ij}\nu^{(i)}}(\mathbf{Y}^{(j)})$ and the metric of $\mathbf{Y}^{(j)}$, where $\nu^{(i)}$ is the maximal measure of $\mathbf{Y}^{(i)}$. Notably, such a result relies on whether the push-forward measure $\pi_{ij}\nu^{(i)}$ of $\nu^{(i)}$ under the factor map π_{ij} remains a maximal measure. We propose a methodology so that all the conditions are verifiable, and the Hausdorff dimension $\dim\mathbf{Y}^{(\ell)}$ can be formulated accurately for $1 \le \ell \le n$.

Figure 1 illustrates the fractal sets of the hidden and output spaces (namely, $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$) of a two-layer CNN. It is seen that $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are two entirely different spaces. Aside from calculating the Hausdorff dimensions of these spaces, it is interesting to investigate whether there is a map connecting $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$, and how $\dim\mathbf{Y}^{(1)}$ is related to $\dim\mathbf{Y}^{(2)}$. See Example 2 for more details.
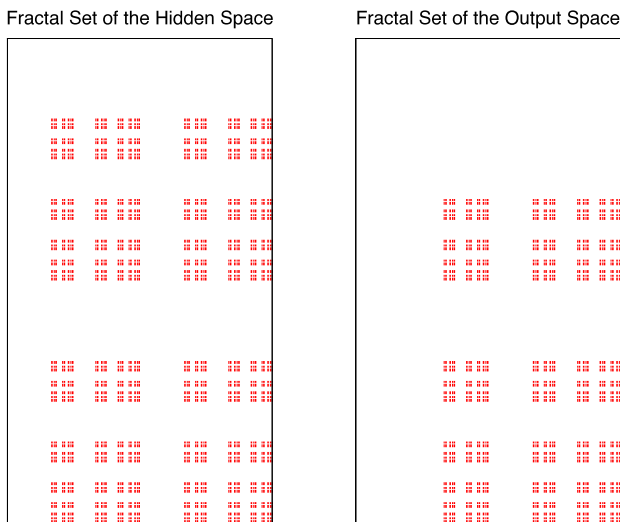

Fractal Set of the Hidden Space   Fractal Set of the Output Space

Figure 1. The fractal sets of the hidden and output spaces of an MCNN with templates $\left[a^{(1)}, a_r^{(1)}, z^{(1)}\right] = [2.9, 1.7, 0.1]$ and $\left[a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)}\right] = [-0.3, -1.2, 0.7, 2.3, 0.9]$. Each fractal set is a subspace of $[0, 1]\times[0, 1]$ and is derived from the expansion $\Phi^{(i)}(x) = \left(\Sigma_{k\ge 0}x_k \left/ \left(m_i^{k+1}\right)\right., \Sigma_{k\le 0}x_k \left/ \left(m_i^{|k|+1}\right)\right.\right)$ for $x \in \mathbf{Y}^{(i)}$, $m_i = |\mathcal{A}(\mathbf{Y}^{(i)})|$, and $i = 1, 2$. It is seen that the expansion map $\Phi^{(i)} : \mathbf{Y}^{(i)} \to [0, 1]\times[0, 1]$ is one-to-one almost everywhere and hence does not make an impact on the discussion of the Hausdorff dimension. The figures come from repeating nine operations based on the basic set of admissible local patterns. See Example 2 for more discussion.

In the meantime, we want to mention some further issues that are related to our elucidation and that have attracted widespread attention recently. One of them is the investigation of the so-called *sofic measure* or *hidden Markov measure*. Let μ be a Markov measure on $\mathbf{Y}$. The push-forward measure $\phi^{(\ell)}\mu$, defined by $\left(\phi^{(\ell)}\mu\right)(O) = \mu\left(\left(\phi^{(\ell)}\right)^{-1}O\right)$ for all Borel set O in $\mathbf{Y}^{(\ell)}$, is called a sofic measure or a hidden Markov measure. There have been many papers about sofic measures written in the past decades. An important question is under what conditions the push-forward measure of a Markov measure remains a Markov measure. To be more specific, we are interested in which properties a sofic measure would satisfy. This elucidation focuses on the study of the measures on the hidden/output space. Recalling that the hidden/output space is a factor of the solution space, it follows that the investigation of the measures on the hidden/output space is equivalent to the investigation of sofic measures. We propose a methodology to verify when a sofic measure is reduced to a Markov measure. In this case, the explicit form of a maximal measure and the Hausdorff dimension of the hidden/output space are formulated. For more discussion of the hidden Markov measures, the reader is referred to [28, 30, 31] and the references therein.

It is known that the tiling problem is undecidable. As an application, it is of interest to investigate the decidability of the language of a sofic shift, which can be realized as a hidden or output space of an MCNN. The related work is in preparation.

The rest of this investigation is organized as follows. A brief recall of [1, 27] and some definitions and notations are given in Section 2. The main theorems (Theorems 2 and 3) for two-layer CNNs are also stated therein. Section 3 analyzes the existence of factor maps that connect two spaces and the hidden Markov measures. The proofs of the main theorems are illustrated there. Some examples are given in Section 4. We generalize Theorems 2 and 3 to general MCNNs in Section 5. Figure 8 provides the flow chart of the present investigation. Section 6 is saved for the conclusion and further problems.

2. Main Results and Preliminaries

Due to this paper being a continuation of [1], the upcoming section intends to give a brief review of [1] and illustrates the main results of our study. For the self-containment of the present investigation, we recall some definitions and known results for symbolic dynamical systems and MCNNs. The reader is referred to [1, 27, 32] and the references therein for more details.

▌ 2.1 Multilayer Cellular Neural Networks

Since an elucidation of two-layer CNNs is essential for the study of MCNNs, we compare MCNNs to two-layer CNNs and focus on them in the rest of this paper unless otherwise stated. A two-layer CNN is realized as

$$
\begin{cases}
\dfrac{dx_i^{(1)}}{dt} = -x_i^{(1)} + \sum_{|k|\le d} a_k^{(1)} y_{i+k}^{(1)} + \sum_{|\ell|\le d} b_\ell^{(1)} u_{i+\ell}^{(1)} + z^{(1)}, \\[2ex]
\dfrac{dx_i^{(2)}}{dt} = -x_i^{(2)} + \sum_{|k|\le d} a_k^{(2)} y_{i+k}^{(2)} + \sum_{|\ell|\le d} b_\ell^{(2)} u_{i+\ell}^{(2)} + z^{(2)},
\end{cases}
\tag{4}
$$

for some $d \in \mathbb{N}$, and $u_i^{(2)} = y_i^{(1)}$ for $i \in \mathbb{Z}$; $\mathbb{N}$ denotes the positive integers and $\mathbb{Z}$ denotes the integers. The prototype of equation (4) is

$$
\frac{dx_i}{dt} = -x_i + \sum_{|k|\le d} a_k y_{i+k} + \sum_{|\ell|\le d} b_\ell u_{i+\ell} + z.
$$

Here $A = [-a_d, \ldots, a_d]$, $B = \left[-b_d, \ldots, b_d\right]$ are the *feedback* and *controlling templates*, respectively. z is the *threshold*, and $y_i = f(x_i) = 1/2(|x_i + 1| - |x_i - 1|)$ is the output of x_i. The quantity x_i represents the state of the cell at i for $i \in \mathbb{Z}$. The output of a stationary solution $\overline{x} = (\overline{x}_i)_{i\in\mathbb{Z}}$ is called an output pattern. A *mosaic solution* $\overline{x}$ satisfies $|\overline{x}_i| > 1$, and its corresponding pattern $\overline{y}$ is called a *mosaic output pattern*. Consider the mosaic solution $\overline{x}$, the necessary and sufficient condition for state "+" at cell C_i, that is, $\overline{y}_i = 1$, is

$$
a - 1 + z > -\left(\sum_{0<|k|\le d} a_k \overline{y}_{i+k} + \sum_{|\ell|\le d} b_\ell u_{i+\ell} \right),
\tag{5}
$$

where $a = a_0$. Similarly, the necessary and sufficient condition for state "$-$" at cell C_i, that is, $\overline{y}_i = -1$, is

$$
a - 1 - z > \sum_{0<|k|\le d} a_k \overline{y}_{i+k} + \sum_{|\ell|\le d} b_\ell u_{i+\ell}.
\tag{6}
$$

For simplicity, we denote $\overline{y}_i$ by y_i and rewrite the output patterns $y_{-d}\cdots y_0 \cdots y_d$ coupled with input $u_{-d}\cdots u_0 \cdots u_d$ as

$$
\boxed{\begin{matrix} y_{-d}\cdots y_{-1} y_0 y_1 \cdots y_d \\ u_{-d}\cdots u_{-1} u_0 u_1 \cdots u_d \end{matrix}} \equiv y_{-d}\cdots y_d \diamond u_{-d}\cdots u_d \in \{-1, 1\}^{\mathbb{Z}_{(2d+1)\times 2}}.
\tag{7}
$$

Let

$$
V^n = \left\{ v \in \mathbb{R}^n : v = (v_1, v_2, \ldots, v_n), \text{ and } |v_i| = 1, 1 \le i \le n \right\},
$$

where $n = 4d + 1$; equations (5) and (6) can be rewritten in a compact form by introducing the following notation.

Denote $\alpha = (a_{-d}, \ldots, a_{-1}, a_1, \ldots, a_d)$, $\beta = (b_{-d}, \ldots, b_d)$. Then, α can be used to represent A', the surrounding template of A without a center, and β can be used to represent the template B. The basic set of admissible local patterns with "+" state in the center is defined as

$$\mathcal{B}(+, A, B, z) = \{v \diamond w \in V^n : a - 1 + z > -(\alpha \cdot v + \beta \cdot w)\}, \tag{8}$$

where "$\cdot$" is the inner product in Euclidean space. Similarly, the basic set of admissible local patterns with "−" state in the center is defined as

$$\mathcal{B}(-, A, B, z) = \{v' \diamond w' \in V^n : a - 1 - z > \alpha \cdot v + \beta \cdot w\}. \tag{9}$$

Furthermore, the admissible local patterns induced by (A, B, z) can be denoted by

$$\mathcal{B}(A, B, z) = (\mathcal{B}(+, A, B, z), \mathcal{B}(-, A, B, z)). \tag{10}$$

It is shown that the parameter space can be partitioned into finite equivalent subregions; that is, two sets of parameters induce identical basic sets of admissible local patterns if they belong to the same partition in the parameter space. Moreover, the parameter space of an MCNN is also partitioned into finite equivalent subregions [27].

Suppose a partition of the parameter space is determined; that is, the templates

$$A^{(\ell)} = [a_{-d}^{(\ell)}, \ldots, a_d^{(\ell)}], \ B^{(\ell)} = [b_{-d}^{(\ell)}, \ldots, b_d^{(\ell)}], \ z^{(\ell)} \ell = 1, 2$$

are given. A stationary solution

$$\mathbf{x} = \mathbf{x}^{(2)} \diamond \mathbf{x}^{(1)} = \begin{pmatrix} x_i^{(2)} \\ x_i^{(1)} \end{pmatrix}_{i \in \mathbb{Z}}$$

is called mosaic if $|x_i^{(\ell)}| > 1$ for $\ell = 1, 2$ and $i \in \mathbb{Z}$. The output

$$\mathbf{y} = \mathbf{y}^{(2)} \diamond \mathbf{y}^{(1)} = \begin{pmatrix} y_i^{(2)} \\ y_i^{(1)} \end{pmatrix}_{i \in \mathbb{Z}}$$

of a mosaic solution $\mathbf{x}$ is called a mosaic pattern.

Suppose $\mathcal{B}$ is the basic set of admissible local patterns of an MCNN. Since equation (4) is spatial homogeneous, that is, the templates of equation (4) are fixed for each cell, the *solution space* $\mathbf{Y} \subseteq \{-1, 1\}^{\mathbb{Z}_{\infty \times 2}}$ is determined by $\mathcal{B}$ as

$$\mathbf{Y} = \{\mathbf{y}^{(2)} \diamond \mathbf{y}^{(1)} : y_{i-d}^{(2)} \cdots y_i^{(2)} \cdots y_{i+d}^{(2)} \diamond y_{i-d}^{(1)} \cdots y_i^{(1)} \cdots y_{i+d}^{(1)} \in \mathcal{B} \text{ for } i \in \mathbb{Z}\}.$$

Moreover, the *output space* $\mathbf{Y}^{(2)}$ and the *hidden space* $\mathbf{Y}^{(1)}$ are defined by

$$\mathbf{Y}^{(2)} = \left\{ \mathbf{y} \in \{-1, 1\}^{\mathbb{Z}} : \mathbf{y} \diamond \mathbf{u} \in \mathbf{Y} \text{ for some } u \right\}$$

and

$$\mathbf{Y}^{(1)} = \left\{ \mathbf{u} \in \{-1, 1\}^{\mathbb{Z}} : \mathbf{y} \diamond \mathbf{u} \in \mathbf{Y} \text{ for some } \mathbf{y} \right\},$$

respectively. In [1, 27], the authors demonstrated that $\mathbf{Y}$ is a *shift of finite type* (SFT) and $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$ are both sofic shifts. In general, for $i = 1, 2$, the factor $\phi^{(i)} : \mathbf{Y} \to \mathbf{Y}^{(i)}$ is not even a finite-to-one surjective map. Furthermore, for $i = 1, 2$, there is a *covering space* $W^{(i)}$ of $\mathbf{Y}^{(i)}$ and a finite-to-one factor $\phi^{(i)} : W^{(i)} \to \mathbf{Y}^{(i)}$, with $W^{(i)}$ being an SFT. (We abuse the finite-to-one factor $\phi^{(i)}$ rather than $\hat{\phi}^{(i)}$ to ease the use of notation.) For a topological space Y, we say that X is a covering space of Y if there exists a continuous onto map $\phi : X \to Y$ that is locally homeomorphic.

A quantity that describes the complexity of a system is *topological entropy*. Suppose X is a shift space. Denote $\Gamma_k(X)$ the cardinality of the collection of words of length k. The topological entropy of X is then defined by

$$h(X) = \lim_{k \to \infty} \frac{\log \Gamma_k(X)}{k}.$$

Whenever the hidden space $\mathbf{Y}^{(1)}$ and the output space $\mathbf{Y}^{(2)}$ reach the same topological entropy, $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are *finite shift equivalent* (FSE) [1]. Herein two spaces X and Y are FSE if there is a triple (Z, ϕ_X, ϕ_Y) such that Z is an SFT and $\phi_X : Z \to X$, $\phi_Y : Z \to Y$ are both finite-to-one factors. Ban et al. [1] asserted that the existence of a *factor-like* matrix helps in determining whether or not there is a map between $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$. A non-negative $m \times n$ integral matrix E is called factor-like if, for each fixed row, the summation of all entries is equal to 1.

Proposition 1. (See [1, Proposition 3.15, Theorem 3.17].) Let $T^{(i)}$ be the transition matrix of $W^{(i)}$ for $i = 1, 2$. Suppose E is a factor-like matrix such that $T^{(i)} E = E T^{(\bar{i})}$; then there is a map $\pi : W^{(i)} \to W^{(\bar{i})}$ that preserves topological entropy, where $i + \bar{i} = 3$. Furthermore, if $\phi^{(i)}$ is a conjugacy, then there is a map $\bar{\pi} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(\bar{i})}$ that preserves topological entropy.

Proposition 1 implies a criterion for the existence of maps between $W^{(1)}$, $W^{(2)}$ and $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$ when $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are FSE. Figure 2 illustrates a triangular structure between $W^{(i)}$ and $\mathbf{Y}^{(i)}$. The structures of

the hidden and output spaces are related if the dashed lines can be replaced by solid lines. Some natural questions follow immediately.
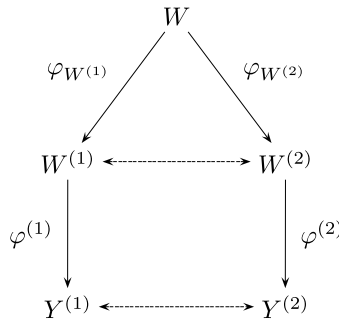

Figure 2. For the case that $h(\mathbf{Y}^{(1)}) = h(\mathbf{Y}^{(2)})$, we get a triangular structure. Whether the dashed lines can be replaced by solid lines implies whether the structures of the hidden and output spaces are related.

Problem 1. Suppose $\pi : W^{(i)} \to W^{(\bar{i})}$ exists.

1. Let μ be a Markov measure on $W^{(i)}$. Is $\pi\mu$ a Markov measure on $W^{(\bar{i})}$, where $\pi\mu := \mu \circ \pi^{-1}$ is the push-forward measure of μ?

2. Suppose π is surjective. For each Markov measure μ' on $W^{(\bar{i})}$, does there exist a μ on $W^{(i)}$ such that $\pi\mu = \mu'$?

3. How is the Hausdorff dimension dim $W^{(\bar{i})}$ related to the Hausdorff dimension dim $W^{(i)}$?

Problem 1 considers whether or not a topological map connects the measures and the Hausdorff dimension of two spaces. Notably, $W^{(1)}$, $W^{(2)}$ are topological Markov chains. It gets more complicated when investigating the hidden and output spaces.

Problem 2. Suppose $\bar{\pi} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(\bar{i})}$ exists.

1. Let ν be a maximal measure on $\mathbf{Y}^{(i)}$. Is $\bar{\pi}\nu$ a maximal measure on $\mathbf{Y}^{(\bar{i})}$?

2. Suppose $\bar{\pi}$ is surjective. For each Markov measure ν' on $\mathbf{Y}^{(\bar{i})}$, does there exist a ν on $\mathbf{Y}^{(i)}$ such that $\bar{\pi}\nu = \nu'$?

3. Is the Hausdorff dimension dim $\mathbf{Y}^{(\bar{i})}$ related to the Hausdorff dimension dim $\mathbf{Y}^{(i)}$?

▌ 2.2 Shift Spaces and Hausdorff Dimension

In this subsection, we recall some definitions and properties of shift spaces and the Hausdorff dimension for the reader's convenience. The detailed information is referred to in [29, 32]. Let $\mathcal{A}$ be a finite set with cardinality $|\mathcal{A}| = n$, which we consider to be an alphabet

of symbols. Without the loss of generality, we usually take $\mathcal{A} = \{0, 1, \ldots, n-1\}$. The full $\mathcal{A}$-shift $\mathcal{A}^{\mathbb{Z}}$ is the collection of all bi-infinite sequences with entries from $\mathcal{A}$. More precisely,

$$\mathcal{A}^{\mathbb{Z}} = \{\alpha = (\alpha_i)_{i\in\mathbb{Z}} : \alpha_i \in \mathcal{A} \text{ for all } i \in \mathbb{Z}\}.$$

The shift map σ on the full shift $\mathcal{A}^{\mathbb{Z}}$ is defined by

$$\sigma(\alpha)_i = \alpha_{i+1} \text{ for } i \in \mathbb{Z}.$$

A *shift space* X is a subset of $\mathcal{A}^{\mathbb{Z}}$ such that $\sigma(X) \subseteq X$. $\mathcal{A}^{\mathbb{Z}}$ is a compact metric space endowed with the metric

$$d(x, y) = \sum_{i\in\mathbb{Z}} \frac{|x_i - y_i|}{n^{|i|+1}}, \quad x, \ y \in \mathcal{A}^{\mathbb{Z}}.$$

Two specific types of shift spaces that are related to our investigation are subshifts of finite type and sofic shifts. First we introduce the former. For each $k \in \mathbb{N}$, let

$$\mathcal{A}_k = \{w_0 w_1 \ldots w_{k-1} : w_i \in \mathcal{A}, 0 \le i \le k-1\}$$

denote the collection of words of length k and let $\mathcal{A}_0$ denote the empty set. A *cylinder* $I \subset \mathcal{A}^{\mathbb{Z}}$ is

$$I = \{x \in \mathcal{A}^{\mathbb{Z}} : x_i x_{i+1} \ldots x_{i+k-1} = \omega_0 \omega_1 \ldots \omega_{k-1}\},$$

for some $i \in \mathbb{Z}, k \in \mathbb{N}$, and $\omega_0 \omega_1 \ldots \omega_{k-1} \in \mathcal{A}_k$. (Sometimes we also write $I = [\omega_0, \omega_1, \ldots, \omega_{k-1}]$.) If X is a shift space and there exists $L \ge 0$ and $\mathcal{F} \subseteq \bigcup_{0\le k\le L} \mathcal{A}_k$ such that

$$X = \{(\alpha_i)_{i\in\mathbb{Z}} : \alpha_i \alpha_{i+1} \ldots \alpha_{i+k-1} \notin \mathcal{F} \text{ for } k \le L, i \in \mathbb{Z}\},$$

then we say that X is an SFT. The SFT is *L-step* if words in $\mathcal{F}$ have length at most $L+1$.

Notably, it is known that without the loss of generality, SFTs can be defined by 0, 1 transition matrices. For instance, let T be an $n \times n$ matrix with rows and columns indexed by $\mathcal{A}$ and entries from $\{0, 1\}$. Then

$$X = \{x \in \mathcal{A}^{\mathbb{Z}} : T(x_i, x_{i+1}) = 1 \text{ for all } i \in \mathbb{Z}\}$$

is a one-step SFT. (It is also known as a *topological Markov chain* by Parry.) A topological Markov chain is called irreducible/mixing if its transition matrix is irreducible/mixing.

An extended concept of SFTs is called *sofic shifts*. A sofic shift is a subshift that is the image of an SFT under a factor map. Suppose X and Y are two shift spaces. A *factor map* is a continuous onto map $\pi : X \to Y$ such that $\pi \circ \sigma_X = \sigma_Y \circ \pi$. A one-to-one factor map is called a *topological conjugacy*. A sofic shift is *irreducible* if it is the image of an irreducible SFT.

In Section 2.1 we mentioned that the topological entropy illustrates the complexity of the topological behavior of a system. Aside from the topological entropy, the Hausdorff dimension characterizes its geometrical structure. The concept of the Hausdorff dimension generalizes the notion of the dimension of a real vector space and helps to distinguish the difference of measure zero sets. We recall the definition of the Hausdorff dimension for the reader's convenience.

Given $\epsilon > 0$, an ϵ-cover $\{U_i\}$ of X is a cover such that the diameter of U_i is less than ϵ for each i. Hence

$$\mathcal{H}^s(X) = \liminf_{\epsilon \to 0} \sum_{i=1}^{\infty} \delta(U_i)^s, \tag{11}$$

where $\delta(U_i)$ denotes the diameter of U_i. The Hausdorff dimension of X is defined by

$$\dim X = \inf\{s : \mathcal{H}^s(X) = 0\}. \tag{12}$$

For subsets that are invariant under a dynamical system, we can pose the problem of the Hausdorff dimension of an invariant measure. To be precise, let us consider a map $g : X \to X$ with invariant probability measure μ. The stochastic properties of g are related to the topological structure of X. A relevant quantitative characteristic, which can be used to describe the complexity of the topological structure of X, is the Hausdorff dimension of the measure μ. The Hausdorff dimension of a probability measure μ on X is defined by

$$\dim \mu = \inf\{\dim Z : Z \subset X \text{ and } \mu(Z) = 1\}.$$

μ is called a *measure of full Hausdorff dimension* (MFHD) if $\dim \mu = \dim X$. An MFHD is used for the investigation of the Hausdorff dimension $\dim X$, and the computation of the Hausdorff dimension of an MFHD corresponds to the computation of the *measure-theoretic entropy*, an analogous quantity since it is the topological entropy that illustrates the complexity of a physical system of X with respect to the MFHD [29, 33]. This causes the discussion of measure-theoretic entropy to play an important role in this elucidation.

Given a shift space (X, σ), we denote by $\mathcal{M}(X)$ the set of σ-invariant Borel probability measures on X. Suppose P is an irreducible stochastic matrix and p a stochastic row vector such that $pP = p$; that is, the summation of entries in each row of P is 1, and the summation of the entries of p is 1. Notably such p is unique due to the irreducibility of P. Define a 0, 1 matrix T by $T(i, j) = 1$ if and only if $P(i, j) > 0$. (The matrix T is sometimes known as the *incidence matrix* of P.) Denote the space of *right-sided* SFT X_T^+ by

$$X_T^+ = \{x \in \mathcal{A}^{\mathbb{Z}^+} : T(x_i, x_{i+1}) = 1 \text{ for all } i \geq 0\}.$$

It is seen that X_T^+ is embedded as a subspace of X_T. The metric on X_T^+ is endowed with

$$d^+(x, y) = \sum_{i \geq 0} \frac{|x_i - y_i|}{n^{i+1}}, \ x, y \in X_T^+.$$

Then (p, P) defines an invariant measure μ^+ on X_T^+ as

$$\mu^+([\omega_0, \omega_1, \dots, \omega_{k-1}]) = p(\omega_0)P(\omega_0, \omega_1)\dots P(\omega_{k-2}, \omega_{k-1})$$

for each cylinder set $I^+ = [\omega_0, \omega_1, \dots, \omega_{k-1}] \subset X_T^+$ by the Kolmogorov extension theorem. Moreover, a measure μ^+ on X_T^+ is a Markov measure if and only if it is determined by a pair (p, P) as above.

Similarly to the above, we define the *left-sided* SFT X_T^- by

$$X_T^- = \{x \in \mathcal{A}^{\mathbb{Z}^-} : T(x_{-i}, x_{-i+1}) = 1 \text{ for all } i \in \mathbb{N}\}.$$

Then X_T^- is a subspace of X_T, and the metric on X_T^- is endowed with

$$d^-(x, y) = \sum_{i \leq 0} \frac{x_i - y_i}{n^{-i+1}}, \ x, y \in X_T^-.$$

Let Q be the transpose of P, and q be the stochastic row vector such that $qQ = q$. Then (q, Q) defines an invariant measure μ^- on X_T^- as

$$\mu^-([\omega_{-k+1}, \dots, \omega_{-1}, \omega_0]) = q(\omega_0)Q(\omega_0, \omega_{-1})\dots Q(\omega_{-k+2}, \omega_{-k+1})$$

for each cylinder set $I^- = [\omega_{-k+1}, \dots, \omega_{-1}, \omega_0] \subset X_T^-$.

Notice that given a cylinder $I = [\omega_{-\ell+1}, \dots, \omega_0, \dots, \omega_{k-1}] \subset X_T$, $\ell, k \geq 1$, I can be identified with the direct product $I^+ \times I^-$, where $I^+ = [\omega_0, \dots, \omega_{k-1}] \subset X_T^+$ and $I^- = [\omega_{-\ell+1}, \dots, \omega_0] \subset X_T^-$. Furthermore, (p, P) defines an invariant measure μ on X_T as $\mu(I) \approx \mu^+(I^+)\mu^-(I^-)$ for any cylinder $I \subset X_T$. To be precise, there exist positive constants A_1 and A_2 such that for integers $k, \ell \geq 0$, and any cylinder $I = [\omega_{-\ell+1}, \dots, \omega_0, \dots, \omega_{k-1}] \subset X_T$,

$$A_1 \leq \frac{\mu(I)}{\mu^+(I^+)\mu^-(I^-)} \leq A_2. \tag{13}$$

Combining equation (13) with the fact that every cylinder $I \in X_T$ is identified with $I^+ \times I^-$ implies that the study of the measure-theoretic entropy of one-sided subspace X_T^+ / X_T^- is significant for investigating the measure-theoretic entropy of X_T. What is more, the computation of the Hausdorff dimension of X is closely related to the computation of the Hausdorff dimension of X_T^+ / X_T^-. The reader is referred to [34] for more details.

Now we are ready to introduce the general definition of the measure-theoretic entropy. Given a shift space $X \subseteq \mathcal{A}^{\mathbb{Z}}$ and an invariant

probability measure μ on X, the measure-theoretic entropy of X with respect to μ is given by

$$h_\mu(X) = -\lim_{n\to\infty} \frac{1}{n} \sum_{I\in X_n} \mu(I) \log \mu(I),$$

where X_n denotes the collection of cylinders of length n in X. The concepts of the measure-theoretic and topological entropies are connected by the *variational principle*:

$$h(X) = \sup\{h_\mu(X) : \mu \in \mathcal{M}(X)\}.$$

μ is called a *measure of maximal entropy* (also known as maximal measure) if $h_\mu(X) = h(X)$. Notably, suppose $X_T^+/X_T^-/X_T$ is an SFT determined by T, which is the incidence matrix of an irreducible stochastic matrix P. It is well-known that the Markov measure $\mu^+/\mu^-/\mu$, derived from the pair (p, P), is the *unique* measure of maximal entropy.

▍ 2.3 Results

This subsection is devoted to illustrating the main results of the present elucidation. First we recall a well-known result.

Theorem 1. (See [34, Theorem 4.1.7].) Suppose $\phi : X \to Y$ is a one-block factor map between mixing SFTs, and X has positive entropy. Then either:

1. ϕ is uniformly bounded-to-one.

2. ϕ has no diamond.

3. $h(X) = h(Y)$.

Or:

4. ϕ is uncountable-to-one on some point.

5. ϕ has a diamond.

6. $h(X) > h(Y)$.

A *diamond* for $\phi : X \to Y$ is a pair of distinct points in X differing in only a finite number of coordinates with the same image under ϕ (see Figure 3). Theorem 1 reveals that the investigation of the existence of diamonds is equivalent to the study of infinite-to-one factor maps.

Without the loss of generality, we may assume that every factor map ϕ is a one-block code. That is, there exists $\Phi : \mathcal{A}(X) \to \mathcal{A}(Y)$ such that $\phi(x)_i = \Phi(x_i)$ for $i \in \mathbb{Z}$. Theorem 1, in other words, indicates that every factor map is either finite-to-one or infinite-to-one. In [1], the authors investigated those finite-to-one factor maps. The infinite-to-

one factor maps are examined in this study. Once a factor map exists, we can use it to formulate the Hausdorff dimensions of these spaces.
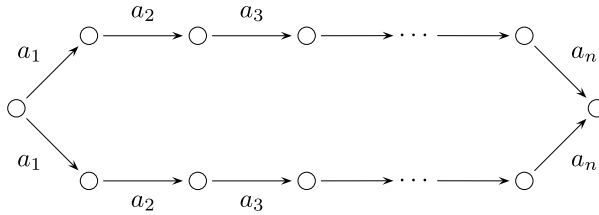

Figure 3. That a factor map $\phi: X \to Y$ has a diamond implies that there exists a pair of distinct points in X differing in only finitely many coordinates with the same image under ϕ. It is named after the shape of its labeled graph representation.

We start with considering the case that $\mathbf{Y}^{(1)}$ is finitely shift equivalent to $\mathbf{Y}^{(2)}$. That two spaces are FSE implies that a factor map between them, if it exists, is finite-to-one. Let X be a shift space. A point $x \in X$ is said to be *doubly transitive* if, for every $k \in \mathbb{N}$ and word w in X, there exist $\bar{\ell} < 0 < \ell$ with $|\bar{\ell}|, \ell > k$ such that

$$x_{\bar{\ell}-|w|+1}...x_{\bar{\ell}} = w \text{ and } x_{\ell}...x_{\ell+|w|-1} = w.$$

Suppose $\phi: X \to Y$ is a factor map. If there is a positive integer K such that every doubly transitive point of Y has exactly K preimages under ϕ, such K is called the *degree* of ϕ, and we define $d_{\phi} = K$ [32].

Let $w = w_1...w_n$ be a word in Y. For $1 \le i \le n$, define $d_{\phi}^*(w, i)$ to be the number of alphabets at coordinate i in the preimages of w. In other words,

$$d_{\phi}^*(w, i) = \\ \sharp\{u_i \in \mathcal{A}(X) : u = u_1...u_n \in X_n, \Phi(u_i) = w_i \text{ for } 1 \le i \le n\}.$$

Denote

$$d_{\phi}^* = \min\{d_{\phi}^*(w, i) : w \in B(Y), 1 \le i \le |w|\},$$

where $B(Y)$ indicates the collection of words in Y.

Definition 1. We say that $w \in B(Y)$ is a magic word if $d_{\phi}^*(w, i) = d_{\phi}^*$ for some i. Such an index i is called a magic coordinate.

We say a factor map ϕ has a *synchronizing word* if there is a finite block $y_1y_2...y_n \in \mathcal{A}_n(Y)$ such that each element in $\phi^{-1}(y_1y_2...y_n)$ admits the same terminal entry. A finite-to-one factor map ϕ has a synchronizing word indicating that the push-forward measure of a mea-

sure of maximal entropy under a finite-to-one factor map is still a measure of maximal entropy. The following is our first main result.

Theorem 2. Suppose the hidden space $\mathbf{Y}^{(1)}$ and the output space $\mathbf{Y}^{(2)}$ are FSE. Let $W^{(i)}$ be irreducible with finite-to-one factor map $\phi^{(i)} : W^{(i)} \to \mathbf{Y}^{(i)}$ for $i = 1, 2$. If $\phi^{(i)}$ has a synchronizing word, then

1. There is a one-to-one correspondence between $\mathcal{M}_{\max}(W^{(i)})$ and $\mathcal{M}_{\max}(\mathbf{Y}^{(i)})$, where $\mathcal{M}_{\max}(X)$ indicates the set of measures of maximal entropy.

2. Let $m_i = |\mathcal{A}(W^{(i)})|$, $n_i = |\mathcal{A}(\mathbf{Y}^{(i)})|$, and $\mu^{(i)}$ be a maximal measure of $W^{(i)}$. Then

$$\dim W^{(i)} = \frac{h_{\mu^{(i)}}(W^{(i)})}{\log m_i} = 2\frac{h_{\mu^{(i),+}}(W^{(i),+})}{\log m_i}$$

and

$$\dim \mathbf{Y}^{(i)} = \frac{h_{\phi^{(i)}\mu^{(i)}}(\mathbf{Y}^{(i)})}{\log n_i} = \frac{h_{\mu^{(i)}}(W^{(i)})}{\log n_i} = 2\frac{h_{\mu^{(i),+}}(W^{(i),+})}{\log n_i},$$

   where $\mu^{(i),+}$ is the maximal measure of the right-sided subspace $W^{(i),+}$ of $W^{(i)}$.

3. Suppose $\pi : W^{(i)} \to W^{(\bar{i})}$ is a factor map and $\nu^{(i)} = \phi^{(i)}\mu^{(i)}$ for $i = 1, 2$, where $i + \bar{i} = 3$. If

$$\dim \mathbf{Y}^{(i)} = \frac{h_{\nu^{(i)}}(\mathbf{Y}^{(i)})}{\log n_i},$$

   then

$$\dim \mathbf{Y}^{(\bar{i})} = \frac{h_{\bar{\pi}\nu^{(i)}}(\mathbf{Y}^{(\bar{i})})}{\log n_{\bar{i}}} = \frac{h_{\nu^{(\bar{i})}}(\mathbf{Y}^{(\bar{i})})}{\log n_{\bar{i}}}$$

   for some $\bar{\pi}$.

In contrast with the map connecting $\bar{\pi} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(\bar{i})}$, if it exists, being finite-to-one when $h(\mathbf{Y}^{(1)}) = h(\mathbf{Y}^{(2)})$, Theorem 1 indicates that $\bar{\pi}$ must be infinite-to-one for the case where $h(\mathbf{Y}^{(1)}) \neq h(\mathbf{Y}^{(2)})$. Intuitively, the number of infinite-to-one factor maps is much larger than the number of finite-to-one factor maps. Computer-assisted examination gives affirmative results for MCNN [35].

Suppose $\phi : X \to Y$ is a factor map and $h(X) \neq h(Y)$. Intuitively, there is a maximal measure in Y with infinite preimage. It is natural to ask whether these preimages are isomorphic to one another. The isomorphism of two measures demonstrates that their measure-theoretic entropies coincide with the same value. In [36], Boyle and Tun-

cel indicated that any two Markov measures associated with the same image are isomorphic to each other if ϕ is a *uniform factor*. We say that ϕ is uniform if $\phi\mu \in \mathcal{M}_{\max}(Y)$ for every $\mu \in \mathcal{M}_{\max}(X)$. ϕ is a uniform factor indicating $\dim Y$ is related to $\dim X$.

Theorem 3. Assume that $h(\mathbf{Y}^{(1)}) \neq h(\mathbf{Y}^{(2)})$. Let $W^{(i)}$ be irreducible with finite-to-one factor map $\phi^{(i)} : W^{(i)} \to \mathbf{Y}^{(i)}$ for $i = 1, 2$.

1. Suppose $\pi : W^{(i)} \to W^{(\bar{i})}$ is a uniform factor. Let $m_i = |\mathcal{A}(W^{(i)})|$, $n_i = |\mathcal{A}(\mathbf{Y}^{(i)})|$, and $\mu^{(i)}$ be a maximal measure of $W^{(i)}$. If

    $$\dim W^{(i)} = \frac{h_{\mu^{(i)}}(W^{(i)})}{\log m_i},$$

    then

    $$\dim W^{(\bar{i})} = \frac{h_{\mu^{(\bar{i})}}(W^{(\bar{i})})}{\log m_{\bar{i}}} = \frac{h_{\pi\mu^{(i)}}(W^{(\bar{i})})}{\log m_{\bar{i}}}.$$

    Furthermore, suppose $h(\mathbf{Y}^{(i)}) > h(\mathbf{Y}^{(\bar{i})})$ and $\phi^{(i)}$ has a synchronizing word; then

2. There exists a factor map $\bar{\pi} : \mathcal{M}_{\max}(\mathbf{Y}^{(i)}) \to \mathcal{M}_{\max}(\mathbf{Y}^{(\bar{i})})$.

3. If

    $$\dim \mathbf{Y}^{(i)} = \frac{h_{\gamma^{(i)}}(\mathbf{Y}^{(i)})}{\log n_i},$$

    then

    $$\dim \mathbf{Y}^{(\bar{i})} = \frac{h_{\bar{\pi}\gamma^{(i)}}(\mathbf{Y}^{(\bar{i})})}{\log n_{\bar{i}}}.$$

We postpone the proof of Theorems 2 and 3 to Sections 3.3 and 3.4. In the meantime, we will introduce the factor maps between the solution, hidden, and output spaces.

3. Existence of Factors

The existence of factor maps plays an important role in the proof of Theorems 2 and 3. First we focus on whether or not a factor map between two spaces exists, and if it exists, the possibility of finding out an explicit form.

3.1 Classification of Solution Spaces

To clarify the discussion, we consider a simplified case. A simplified MCNN (SMCNN) is unveiled as

$$\begin{cases} \dfrac{dx_i^{(1)}}{dt} = -x_i^{(1)} + a^{(1)}y_i^{(1)} + a_r^{(1)}y_{i+1}^{(1)} + z^{(1)}, \\[4mm] \dfrac{dx_i^{(2)}}{dt} = -x_i^{(2)} + a^{(2)}y_i^{(2)} + a_r^{(2)}y_{i+1}^{(2)} + b^{(2)}u_i^{(2)} + b_r^{(2)}u_{i+1}^{(2)} + z^{(2)}. \end{cases} \tag{14}$$

Suppose

$$\mathbf{y} = \begin{pmatrix} \cdots y_{-1}^{(2)} y_0^{(2)} y_1^{(2)} \cdots \\ \cdots y_{-1}^{(1)} y_0^{(1)} y_1^{(1)} \cdots \end{pmatrix}$$

is a mosaic pattern. For $i \in \mathbb{Z}$, $y_i^{(1)} = 1$ if and only if

$$a^{(1)} + z^{(1)} - 1 > -a_r^{(1)}y_{i+1}^{(1)}. \tag{15}$$

Similarly, $y_i^{(1)} = -1$ if and only if

$$a^{(1)} - z^{(1)} - 1 > a_r^{(1)}y_{i+1}^{(1)}. \tag{16}$$

The same argument asserts

$$a^{(2)} + z^{(2)} - 1 > -a_r^{(2)}y_{i+1}^{(2)} - \left(b^{(2)}u_i^{(2)} + b_r^{(2)}u_{i+1}^{(2)}\right) \tag{17}$$

and

$$a^{(2)} - z^{(2)} - 1 > a_r^{(2)}y_{i+1}^{(2)} + \left(b^{(2)}u_i^{(2)} + b_r^{(2)}u_{i+1}^{(2)}\right) \tag{18}$$

are the necessary and sufficient conditions for $y_i^{(2)} = -1$ and $y_i^{(2)} = 1$, respectively. Note that the quantity $u_i^{(2)}$ in equations (17) and (18) satisfies $\left|u_i^{(2)}\right| = 1$ for each i. Define $\xi_1 : \{-1, 1\} \to \mathbb{R}$ and $\xi_2 : \{-1, 1\}^{\mathbb{Z}_{3\times 1}} \to \mathbb{R}$ by

$$\xi_1(w) = a_r^{(1)}w, \quad \xi_2(w_1, w_2, w_3) = a_r^{(2)}w_1 + b^{(2)}w_2 + b_r^{(2)}w_3.$$

Set

$$\mathcal{B}^{(1)} = \left\{ \boxed{y^{(1)}y_r^{(1)}} : y^{(1)}, y_r^{(1)} \in \{-1, 1\} \text{ satisfy} \right.$$
$$\left. \text{equations } (15) \text{ and } (16) \right\},$$

$$\mathcal{B}^{(2)} = \left\{ \boxed{\begin{matrix} y^{(2)}y_r^{(2)} \\ u^{(2)}u_r^{(2)} \end{matrix}} : y^{(2)}, y_r^{(2)}, u^{(2)}, u_r^{(2)} \in \{-1, 1\} \text{ satisfy} \right.$$
$$\left. \text{equations } (17) \text{ and } (18) \right\}.$$

That is,

$$\boxed{y^{(1)}y_r^{(1)}} \in \mathcal{B}^{(1)} \Leftrightarrow \begin{cases} a^{(1)} + z^{(1)} - 1 > -\xi_1(y_r^{(1)}), & \text{if } y^{(1)} = 1; \\ a^{(1)} - z^{(1)} - 1 > \xi_1(y_r^{(1)}), & \text{if } y^{(1)} = -1. \end{cases}$$

$$\boxed{\begin{matrix} y^{(2)}y_r^{(2)} \\ u^{(2)}u_r^{(2)} \end{matrix}} \in \mathcal{B}^{(2)} \Leftrightarrow \begin{cases} a^{(2)} + z^{(2)} - 1 > -\xi_2(y_r^{(2)}, u^{(2)}, u_r^{(2)}), & \text{if } y^{(2)} = 1; \\ a^{(2)} - z^{(2)} - 1 > \xi_2(y_r^{(2)}, u^{(2)}, u_r^{(2)}), & \text{if } y^{(2)} = -1. \end{cases}$$

The set of admissible local patterns $\mathcal{B}$ of equation (14) is then

$$\mathcal{B} = \left\{ \boxed{\begin{matrix} yy_r \\ uu_r \end{matrix}} : \boxed{\begin{matrix} yy_r \\ uu_r \end{matrix}} \in \mathcal{B}^{(2)} \text{and} \boxed{uu_r} \in \mathcal{B}^{(1)} \right\}.$$

The authors indicated in [1] that there exist 139 968 regions in the parameter space $\mathcal{P}$ of SMCNNs such that any two sets of templates that are located in the same region imply the same solution spaces. The partition of the parameter space is determined as follows.

Since

$$y^{(1)}, y^{(2)}, y_r^{(1)}, y_r^{(2)}, u^{(2)}, u_r^{(2)} \in \{-1, 1\},$$
$$a^{(1)} + z^{(1)} - 1 = -\xi_1(y_r^{(1)})$$

and

$$a^{(1)} + z^{(1)} - 1 = \xi_1(y_r^{(1)})$$

partition the $a^{(1)}z^{(1)}$-plane into nine regions. The "order" of lines $a^{(1)} + z^{(1)} - 1 = (-1)^\ell \xi_1(y_r^{(1)})$, $\ell = 0, 1$ comes from the sign of $a_r^{(1)}$. Thus the parameter space $\{(a^{(1)}, a_r^{(1)}, z^{(1)})\}$ is partitioned into $2 \times 9 = 18$ regions. Similarly,

$$a^{(2)} + z^{(2)} - 1 > -\xi_2(y_r^{(2)}, u^{(2)}, u_r^{(2)})$$

and

$$a^{(2)} + z^{(2)} - 1 > \xi_2(y_r^{(2)}, u^{(2)}, u_r^{(2)})$$

partition the $a^{(2)}z^{(2)}$-plane into 81 regions. The order of

$$a^{(2)} + z^{(2)} - 1 > (-1)^\ell \xi_2(y_r^{(2)}, u^{(2)}, u_r^{(2)})$$

can be uniquely determined according to the following procedures.

1. The signs of $a_r^{(2)}, b^{(2)}, b_r^{(2)}$.

2. The magnitude of $a_r^{(2)}, b^{(2)}, b_r^{(2)}$.

3. The competition between the parameters with the largest magnitude and the others. In other words, suppose $m_1 > m_2 > m_3$ represent $|a_r^{(2)}|$, $|b^{(2)}|$, $|b_r^{(2)}|$. We need to determine whether $m_1 > m_2 + m_3$ or $m_1 < m_2 + m_3$.

This partitions the parameter space $\{(a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)})\}$ into $8 \times 6 \times 2 \times 81 = 7776$ regions. Hence the parameter space $\mathcal{P}$ is partitioned into $81 \times 7776 = 139\,968$ equivalent subregions.

Since the solution space $\mathbf{Y}$ is determined by the basic set of admissible local patterns, these local patterns play an essential role for investigating SMCNNs. Substitute mosaic patterns -1 and 1 as symbols $-$ and $+$, respectively. Define the ordering matrix of $\{-, +\}^{\mathbb{Z}_{2\times 2}}$ by

$$
\mathbb{X} =
\begin{pmatrix}
\begin{smallmatrix} - \\ - \end{smallmatrix} &
\begin{smallmatrix} - \\ + \end{smallmatrix} &
\begin{smallmatrix} + \\ - \end{smallmatrix} &
\begin{smallmatrix} + \\ + \end{smallmatrix} \\[4pt]
\end{pmatrix}
$$

$$
\mathbb{X} =
\begin{smallmatrix} - \\ - \end{smallmatrix}
\begin{smallmatrix} - \\ + \end{smallmatrix}
\begin{smallmatrix} + \\ - \end{smallmatrix}
\begin{smallmatrix} + \\ + \end{smallmatrix}
\left(
\begin{matrix}
\begin{smallmatrix} -- \\ -- \end{smallmatrix} & \begin{smallmatrix} -- \\ -+ \end{smallmatrix} & \begin{smallmatrix} -+ \\ -- \end{smallmatrix} & \begin{smallmatrix} -+ \\ -+ \end{smallmatrix} \\[4pt]
\begin{smallmatrix} -- \\ +- \end{smallmatrix} & \begin{smallmatrix} -- \\ ++ \end{smallmatrix} & \begin{smallmatrix} -+ \\ +- \end{smallmatrix} & \begin{smallmatrix} -+ \\ ++ \end{smallmatrix} \\[4pt]
\begin{smallmatrix} +- \\ -- \end{smallmatrix} & \begin{smallmatrix} +- \\ -+ \end{smallmatrix} & \begin{smallmatrix} ++ \\ -- \end{smallmatrix} & \begin{smallmatrix} ++ \\ -+ \end{smallmatrix} \\[4pt]
\begin{smallmatrix} +- \\ +- \end{smallmatrix} & \begin{smallmatrix} +- \\ ++ \end{smallmatrix} & \begin{smallmatrix} ++ \\ +- \end{smallmatrix} & \begin{smallmatrix} ++ \\ ++ \end{smallmatrix}
\end{matrix}
\right)
= (x_{pq})_{1 \le p, q \le 4}.
$$

Notably, each entry in $\mathbb{X}$ is a 2×2 pattern, since $\mathcal{B}$ consists of 2×2 local patterns. Suppose that $\mathcal{B}$ is given. The transition matrix $T \equiv T(\mathcal{B}) \in \mathcal{M}_4(\{0, 1\})$ is defined by

$$
T(p, q) = \begin{cases} 1, & \text{if } x_{pq} \in \mathcal{B}; \\ 0, & \text{otherwise.} \end{cases}
$$

Let $\mathcal{A} = \{\alpha_0, \alpha_1, \alpha_2, \alpha_3\}$, where

$$
\alpha_0 = --, \quad \alpha_1 = -+, \quad \alpha_2 = +- \text{ and } \alpha_3 = ++.
$$

Define $\mathcal{L}^{(1)}, \mathcal{L}^{(2)}$ by

$$
\mathcal{L}^{(1)}(y_0 y_1 \diamond u_0 u_1) = u_0 u_1 \text{ and } \mathcal{L}^{(2)}(y_0 y_1 \diamond u_0 u_1) = y_0 y_1,
$$

respectively. It is known that T determines a graph, while $(T, \mathcal{L}^{(i)})$ determines a labeled graph for $i = 1, 2$. As we mentioned in Section 2, the transition matrix T determines the solution space $\mathbf{Y}$; T does not describe the hidden and output spaces $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$, though. Instead, $\mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}$ are illustrated by the symbolic transition matrices. The symbolic transition matrix $S^{(i)}$ is defined by

$$
S^{(i)}(p, q) = \begin{cases} \alpha_j, & \text{if } T(p, q) = 1 \text{ and } \mathcal{L}^{(i)}(x_{pq}) = \alpha_j \text{ for some } j; \\ \emptyset, & \text{otherwise.} \end{cases} \tag{19}
$$

Herein $\emptyset$ means there exists no local pattern in $\mathcal{B}$ related to its corresponding entry in the ordering matrix. A labeled graph is called *right-*

resolving if for every fixed row of its symbolic transition matrix, the multiplicity of each symbol is 1. With a little abuse of notations, $\mathbf{Y}^{(i)}$ can be described by $S^{(i)}$, which is right-resolving for $i = 1, 2$. Let $T^{(i)}$ be the incidence matrix of $S^{(i)}$; that is, $T^{(i)}$ is of the same size of $S^{(i)}$ and is defined by

$$T^{(i)}(p, q) = \begin{cases} 1, & \text{if } S^{(i)}(p, q) \neq \emptyset; \\ 0, & \text{otherwise.} \end{cases}$$

Then $W^{(i)}$ is determined by $T^{(i)}$ for $i = 1, 2$. The reader is referred to [1, 27] for more details.

3.2 Sofic Measures and Linear Representable Measures

Theorems 2 and 3 investigate the Hausdorff dimensions of $W^{(i)}$ and $\mathbf{Y}^{(i)}$ and see if they are related. The proof relies on two essential ingredients: the existence of maximal measures and factor maps. The upcoming subsection involves the former, while the latter is discussed in the next two subsections separately.

Let X and Y be subshifts and $\phi : X \to Y$ be a factor map. Suppose μ is a Markov measure on X; then $\phi\mu$ is called a *sofic measure* (also known as a *hidden Markov measure*, see [28]). Let $B \in \mathbb{R}^{m \times m}$ be an irreducible matrix with spectral radius ρ_B and positive right eigenvector r; the *stochasticization* of B is the stochastic matrix

$$\mathbb{B} := \text{stoch}(B) = \frac{1}{\rho_B} D^{-1} B D,$$

where D is the diagonal matrix with diagonal entries $D(i, i) = r(i)$. A measure μ on X is called *linear representable* with dimension m if there exists a triple (x, P, y) with x being a $1 \times m$ row vector, y being a $m \times 1$ column vector, and $P = (P_i)_{i \in \mathcal{A}(X)}$, where $P_i \in \mathbb{R}^{m \times m}$ such that for all $I = [i_0, \ldots, i_{n-1}] \in X_n$, the measure μ can be characterized as the following form:

$$\mu([I]) = x P_I y = x P_{i_0} P_{i_1} \ldots P_{i_{n-1}} y.$$

The triple (x, P, y) is called the *linear representation* of the measure μ. The reader is referred to [28] for more details.

Proposition 2. (See [28, Theorem 4.20].) Let X be an irreducible SFT with transition matrix $T \in \mathbb{R}^{m \times m}$ and $\phi : X \to Y$ be a one-block factor map. Let $\mathbb{T} = \text{stoch}(T)$ and l be the probability left eigenvector of $\mathbb{T}$. Then

1. The Markov measure μ on X is the linear representable measure with respect to the triple $(l, P, \mathbf{1}_m)$, where $\mathbf{1}_m$ is the column vector, with each entry being 1 and $P = (\mathbb{T}_i)_{i \in \mathcal{A}(X)}$, for which

$$P_I = \mathbb{T}_{i_0} \ldots \mathbb{T}_{i_{n-1}} \text{ for all } I = [i_0, \ldots, i_{n-1}] \in X_n,$$

here $\mathbb{T}_k(i, j) = \mathbb{T}(i, j)$ if $j = k$ and $\mathbb{T}_k(i, j) = 0$ otherwise.

2. The push-forward measure $\nu = \phi\mu$ is linear representable with respect to the triple $(l, Q, \mathbf{1}_m)$, where Q is generated by $(Q_j)_{j \in \mathcal{A}(Y)} = (\mathbb{T}_j)_{j \in \mathcal{A}(Y)}$, for which $\mathbb{T}_k(u, v) = \mathbb{T}(u, v)$ if $\phi(v) = k$ and $\mathbb{T}_k(u, v) = 0$ otherwise.

In the following we propose a criterion to determine whether a sofic measure is actually a Markov measure. The procedure of the criterion is systematic and verifiable, which makes our method practical. Suppose the factor map $\phi : X \to Y$ is a one-block code. For $j \in \mathcal{A}(Y)$, define

$$E_j = \{i : \phi(i) = j\} \text{ and } e_j = \sharp E_j.$$

For each $j_1 j_2 \in Y_2$, let $N_{j_1 j_2} \in \mathbb{R}^{e_{j_1} \times e_{j_2}}$ be defined by

$$N_{j_1 j_2}(p, q) = \begin{cases} 1, & pq \in X_2; \\ 0, & \text{otherwise,} \end{cases}$$

where $p \in E_{j_1}$, $q \in E_{j_2}$. Set $N = (N_{j_1 j_2})$ if $e_{j_1} = e_{j_2}$ for all $j_1 j_2 \in Y_2$. Otherwise, we enlarge the dimension of $N_{j_1 j_2}$ by inserting "pseudo vertices" so that $N_{j_1 j_2}$ is a square matrix.

We say that N satisfies the *Markov condition of order k* if there exists a nontrivial subspace $\{V_J\}_{J \in Y_{k+1}}$ such that for each $J \in Y_{k+1}$ there exists $m_{J(0,k-1),J(1,k)}$ such that $V_{J(0,k-1)} N_J = m_{J(0,k-1)}, J(1, k) V_{J(1,k)}$; here $J(i_1, i_2) = j_{i_1} j_{i_1+1} \ldots j_{i_2}$ with $J = j_1 j_2 \ldots j_{k+1}$. For simplification, we say that N satisfies the Markov condition if N satisfies the Markov condition of order k for some $k \in \mathbb{N}$.

At this point, a further question arises, suppose N satisfies the Markov condition—what kind of Markov measure is ν?

To answer this question, we may assume $m(J(0, k - 1), J(1, k)) \in \mathbb{R}$ such that

$$V_{J(0,k-1)} N_{J(0,k)} = m(J(0, k - 1), J(1, k)) V_{J(1,k)}. \tag{20}$$

In [30], the authors illustrated what kind of Markov measure ν is.

Theorem 4. (See [30, Theorem 4].) If N satisfies the Markov condition of order k, then Y is an SFT. Furthermore, ν is the unique maximal measure of Y with transition matrix $M = [m(J, J')]_{J,J' \in Y_k}$.

To clarify the construction of N and Theorem 4, we introduce an example that was initiated by Blackwell.

Example 1. (Blackwell [37].) Let $\mathcal{A}(X) = \{1, 2, 3\}$, $\mathcal{A}(Y) = \{1, 2\}$, and the one-block map $\Phi : \mathcal{A}(X) \to \mathcal{A}(Y)$ be defined by $\Phi(1) = 1$, $\Phi(2) = \Phi(3) = 2$. Let $\phi : X \to Y$ be the factor induced from Φ, and the transition matrix of X be

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

This factor has been proven [28, Example 2.7] to be Markovian. Here we use Theorem 4 to give a criterion for this property. Since $E_1 = \{1\}$ and $E_2 = \{2, 3\}$, we see that $\mathbf{m} = e_2 = 2$, and an extra pseudo vertex is needed for E_1. For these reasons we introduce the new symbols, and the corresponding sets $\hat{E}_1$ and $\hat{E}_2$ are as follows.

$$D = \{1, 2\} \times \{1, 2\} = \{(1, 1), (1, 2), (2, 1), (2, 2)\},$$
$$\hat{E}_1 = \{1 = (1, 1), (2, 1)\}, \quad \hat{E}_2 = \{2 = (1, 2), 3 = (2, 2)\}.$$

Therefore,

$$B = \begin{pmatrix} N_{11} & N_{12} \\ N_{21} & N_{22} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

$$N_{11} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad N_{12} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix},$$
$$N_{21} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad N_{22} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Taking $V_1 = (1\,0)$ and $V_2 = (1\,1)$, we can easily check that $N = (N_{ij})_{i,j=1}^2$ satisfies the Markov condition of order 1. Thus Theorem 4 is applied to show that the factor is a Markov map.

▌ 3.3 Proof of Theorem 2

Proposition 1 asserts that the existence of a factor-like matrix for $T^{(1)}$, $T^{(2)}$ together with the topological conjugacy of $\phi^{(i)}$ implies there is a map $\bar{\pi} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(\bar{i})}$ that preserves topological entropy, where $i = 1, 2$ and $i + \bar{i} = 3$. A natural question is whether or not we can find a map connecting $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ under the condition neither $\phi^{(1)}$ nor $\phi^{(2)}$ is topologically conjugated. The answer is affirmative. First we define the product of scalar and alphabet.

Definition 2. Suppose $\mathcal{A}$ is an alphabet set. Let $\mathbf{A}$ be the free Abelian additive group generated by $\mathcal{A} \cup \{\emptyset\}$; here $\emptyset$ is the identity element. For $\mathbf{k} \in \mathbb{Z}$, $\mathbf{a} \in \mathcal{A} \cup \{\emptyset\}$, we define a commutative operator $*$ by

$$\mathbf{a} * \mathbf{k} = \mathbf{k} * \mathbf{a} = \begin{cases} \mathbf{ka}, & \text{if } \mathbf{a} \neq \emptyset \text{ and } \mathbf{k} \neq 0; \\ \emptyset & \text{otherwise.} \end{cases}$$

Suppose S is an $m \times n$ symbolic matrix and A is an $n \times k$ integral matrix. The product $S * A$ is defined by $(S * A)(p, q) = \sum_{i=1}^{n} S(p, i) * A(i, q)$ for $1 \leq p \leq m$, $1 \leq q \leq k$. For simplicity we denote $S * A$ by SA. Similarly, we can define $A * S$ and denote by AS for $m \times n$ integral matrix A and $n \times k$ symbolic matrix S.

The following proposition, which is an extension of Proposition 1, can be verified with a little modification of the proof of Proposition 3.15 in [1]. Hence we omit the detail.

Proposition 3. Let $S^{(i)}$ be the symbolic transition matrix of $\mathbf{Y}^{(i)}$ for $i = 1, 2$. Suppose E is a factor-like matrix such that $S^{(i)}E = ES^{(\bar{i})}$; then there exist maps $\pi : W^{(i)} \to W^{(\bar{i})}$ and $\bar{\pi} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(\bar{i})}$ that both preserve topological entropy, where $i + \bar{i} = 3$.

A factor map ϕ is *almost invertible* if every doubly transitive point has exactly one preimage. Lemma 1 shows that the existence of a synchronizing word is a necessary and sufficient criterion to determine whether ϕ is almost invertible.

Lemma 1. Suppose $\phi : X \to Y$ is a one-block factor map. Then ϕ is almost invertible if and only if ϕ has a synchronizing word.

Proof. If ϕ is almost invertible, then $d_\phi^* = 1$. Let w be a magic word and i be a magic coordinate. In other words, $d_\phi(w, i) = 1$. The fact that ϕ is right-resolving implies that $d_\phi(w, |w|) = 1$. Hence w is a synchronizing word.

On the other hand, suppose w is a synchronizing word. ϕ being right-resolving indicates $d_\phi(wa, |wa|) = 1$ for some a such that $wa \in B(Y)$. That is, wa is a magic word and $d_\phi^* = 1$. Therefore, ϕ is almost invertible. $\square$

The proof of the first statement of Theorem 2 is done by Lemma 1 and the following theorem.

Theorem 5. (See [38, Theorem 3.4].) Suppose $\phi : X \to Y$ is a factor map and X is an irreducible SFT. If ϕ is almost invertible, then $\phi : \mathcal{M}_{\max}(X) \to \mathcal{M}_{\max}(Y)$ is a bijection. Moreover, $h_\mu(X) = h_{\phi\mu}(Y)$ for $\mu \in \mathcal{M}_{\max}(X)$.

Next we continue the proof of Theorem 2.

Fix $i \in \{1, 2\}$. Recall that the metric $d^{(i)} : W^{(i)} \times W^{(i)} \to \mathbb{R}$ is given by

$$d^{(i)}(x, y) = \sum_{j \in \mathbb{Z}} \frac{|x_j - y_j|}{m_i^{|j|+1}}$$

for $x, y \in W^{(i)}$, where $m_i = \left| \mathcal{A}\left(W^{(i)}\right) \right|$.

To formulate the explicit form of the Hausdorff dimensions of the hidden and output spaces, we introduce the following from Pesin's well-known work.

Theorem 6. (See [29, Theorems 13.1 and 22.2].) Let (X, σ) be a shift space with $|\mathcal{A}(X)| = m$, and $0 < \lambda_1, \dots, \lambda_m < 1$. Suppose d is a metric defined on X. If there exist $K_1, K_2 > 0$ such that

$$K_1 \prod_{j=0}^{n_2} \lambda_{i_j} < \mathrm{diam}\left[i_0, \dots, i_{n_2}\right] < K_2 \prod_{j=0}^{n_2} \lambda_{i_j},$$

$$K_1 \prod_{j=0}^{n_1} \lambda_{i_j} < \mathrm{diam}\left[i_{-n_1}, \dots, i_0\right] < K_2 \prod_{j=0}^{n_1} \lambda_{i_j},$$

for any cylinder $I = \left[i_{-n_1}, \dots, i_{n_2}\right]$, $n_1, n_2 \geq 0$, then

$$\dim X = - \frac{h_{\mu_\lambda}(X)}{\int_X \log \lambda_{i_0} d\mu_\lambda} = -2 \frac{h_{\mu_\lambda^\pm}(X)}{\int_X \log \lambda_{i_0} d\mu_\lambda^\pm},$$

where μ_λ is a maximal measure on X and $\mu_\lambda^\pm$ is a maximal measure on the right-sided subspace X^+/left-sided subspace X^-.

Suppose $\mu^{(i)}$ is a maximal measure of $W^{(i)}$. For any cylinder $I = \left[i_{-n_1}, \dots, i_{n_2}\right]$, the diameters of $\left[i_0, \dots, i_{n_2}\right]$ and $\left[i_{-n_1}, \dots, i_0\right]$ are $1 / \left(m_i^{n_2+1}\right)$ and $1 / \left(m_i^{n_1+1}\right)$, respectively. Let $K_1 = 1$, $K_2 = 3$, and $\lambda_1 = \dots = \lambda_{m_i} = 1 / (m_i)$, apply Theorem 6, and we have

$$\dim W^{(i)} = - \frac{h_{\mu^{(i)}}\left(W^{(i)}\right)}{\int_{W^{(i)}} \log \frac{1}{m_i} d\mu^{(i)}} = \frac{h_{\mu^{(i)}}\left(W^{(i)}\right)}{\log m_i} = 2 \frac{h_{\mu^{(i),\pm}}\left(W^{(i)}\right)}{\log m_i}.$$

Moreover, the one-to-one correspondence between $\mathcal{M}_{\max}\left(W^{(i)}\right)$ and $\mathcal{M}_{\max}\left(\mathbf{Y}^{(i)}\right)$ demonstrates that

$$\dim \mathbf{Y}^{(i)} = \sup\left\{ \frac{h_\nu\left(\mathbf{Y}^{(i)}\right)}{\log n_i} : \nu \text{ is invariant on } \mathbf{Y}^{(i)}\right\} =$$

$$\frac{h_{\phi^{(i)}\mu^{(i)}}\left(\mathbf{Y}^{(i)}\right)}{\log n_i} = \frac{h_{\mu^{(i)}}\left(W^{(i)}\right)}{\log n_i}.$$

The last equality comes from Theorem 5. This completes the proof of Theorem 2 part 2.

Observe that

$$\dim \mathbf{Y}^{(i)} = \frac{h_{\nu^{(i)}}\left(\mathbf{Y}^{(i)}\right)}{\log n_i}$$

indicates $\nu^{(i)}$ is a maximal measure on $\mathbf{Y}^{(i)}$. Since $W^{(i)}$ is irreducible, the maximal measure $\mu^{(i)}$ is unique. Hence there is a bijection $\overline{\pi} : \mathcal{M}_{\max}\left(\mathbf{Y}^{(i)}\right) \to \mathcal{M}_{\max}\left(\mathbf{Y}^{(\bar{i})}\right)$ such that $\overline{\pi}\nu^{(i)} = \nu^{(\bar{i})}$. Therefore,

$$\dim \mathbf{Y}^{(\bar{i})} = \frac{h_{\overline{\pi}\nu^{(i)}}\left(\mathbf{Y}^{(\bar{i})}\right)}{\log n_{\bar{i}}} = \frac{h_{\nu^{(\bar{i})}}\left(\mathbf{Y}^{(\bar{i})}\right)}{\log n_{\bar{i}}}.$$

This completes the proof of Theorem 2.

▎ 3.4 Proof of Theorem 3

Whether there exists a factor map connecting two spaces is always a concerning issue. In general, it is difficult to construct or to say such factor maps exist for a given pair of spaces. Proposition 3 proposes a methodology for constructing a connection between two spaces. Notably a map constructed via Proposition 3 preserves topological entropy. In other words, it only works for those spaces reaching the same topological entropy if we restrict the factor maps. In this subsection, we turn our attention to the factor maps connecting spaces with unequal topological entropies.

Similarly to the proof of Theorem 2, demonstrating Theorem 3 relies mainly on the existence of a factor map. Instead of $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$, we start with examining whether there is a factor map from $W^{(i)}$ to $W^{(\bar{i})}$; note here that $i + \bar{i} = 3$.

Theorem 7. Suppose $W^{(1)}$ and $W^{(2)}$ are irreducible with $h\left(W^{(1)}\right) \neq h\left(W^{(2)}\right)$. Suppose $h\left(W^{(i)}\right) > h\left(W^{(\bar{i})}\right)$, where $i + \bar{i} = 3$. Then there exists an infinite-to-one map $\pi : W^{(i)} \to W^{(\bar{i})}$ if one of the following is satisfied.

1. $h(W^{(i)}) = h(\mathbf{Y})$, and there is a factor-like matrix F such that $T^{(i)}F = FT$, where T is the transition matrix of $\mathbf{Y}$.

2. $h(W^{(i)}) < h(\mathbf{Y})$.

Remark 1. (*i*) Suppose X, Y are two irreducible SFTs with $h(X) > h(Y)$. In [34], Kitchens showed that if there is an infinite-to-one factor map from X^+ to Y^+, then there exists an infinite-to-one factor map $\pi : X \to Y$. This reduces the investigation of Theorem 7 to the exis-

tence of an infinite-to-one map between the right-sided subspaces of
$W^{(1)}$ and $W^{(2)}$.

(*ii*) Theorem 7 reveals the existence of an infinite-to-one map be-
tween the hidden and output spaces whenever these two spaces hit dif-
ferent topological entropies; however, there are an infinite number of
such maps in general. In addition, it is difficult to find the explicit
form of an infinite-to-one map. This is an important issue and is still
open in the field of symbolic dynamical systems. It helps for the inves-
tigation of MCNNs if a methodology can be proposed to find a con-
crete expression of an infinite-to-one map.

The following corollary comes immediately after Theorem 7.

Corollary 1. Under the same assumption of Theorem 7. Suppose further-
more that $\left|\mathcal{A}\left(W^{(i)}\right)\right| \geq \left|\mathcal{A}(\mathbf{Y})\right|$ and 7.1 is satisfied. Then $\pi : W^{(i)} \to W^{(\bar{i})}$
is an infinite-to-one factor map.

Suppose X is a shift space. Let $P(X)$ denote the collection of peri-
odic points in X and let $P_n(X)$ be the set of periodic points with
period n. Given two shifts X and Y, let $q_n(X)$ and $q_n(Y)$ be the cardi-
nality of $\bigcup_{k \geq n} P_k(X)$ and $\bigcup_{k \geq n} P_k(Y)$, respectively. If $q_n(X) \leq q_n(Y)$ for
$n \geq 1$, then we call it an *embedding periodic point condition* and
write it as $P(X) \hookrightarrow P(Y)$. Theorem 8 asserts a necessary and sufficient
condition concerning the existence of an injective map between X
and Y.

Theorem 8. (Embedding theorem.) Suppose X and Y are irreducible
SFTs. There is an embedding map $\phi : X \to Y$ if and only if $h(X) < h(Y)$
and $P(X) \hookrightarrow P(Y)$.

An additional question is the existence of a factor map between X
and Y. Like the embedding periodic point condition, the *factor peri-
odic point condition* indicates that for every $x \in P_n(X)$, there exists a
$y \in P_m(Y)$ such that m is a factor of n and is denoted by $P(X) \searrow P(Y)$.

Theorem 9. (See [34, Theorem 4.4.5].) Suppose X and Y are irreducible
SFTs. There exists an infinite-to-one factor code $\phi : X \to Y$ if and only
if $h(X) > h(Y)$ and $P(X) \searrow P(Y)$.

Proof. (Proof of Theorem 7.) Without the loss of generality, we may
assume that $h\left(W^{(1)}\right) < h\left(W^{(2)}\right)$. It suffices to demonstrate there is an in-
finite-to-one map from $W^{(2),+}$ to $W^{(1),+}$, due to the observation in Re-
mark 1(*i*). For ease of notation, the spaces in the upcoming proof are
referred to as right-sided subspaces.

Suppose that condition 7.1 is satisfied. The existence of factor-like
matrix F such that $T^{(2)}F = FT$ implies there is a map $\Phi^{(2)} : W^{(2)} \to \mathbf{Y}$.

Recall that the graph representation $G^{(1)}$ of $W^{(1)}$ is obtained by applying subset construction to $(G, \mathcal{L}^{(1)})$. Without the loss of generality, we assume that $G^{(1)}$ is essential. That is, every vertex in $G^{(1)}$ is treated as an initial state of one edge and as a terminal state of another. Suppose $w = w_1 \ldots w_n$ is a cycle in G. If the initial state $i(w_k)$ of w_k is a vertex in $G^{(1)}$ for $k = 1, \ldots, n$, then w is also a cycle in $G^{(1)}$.

Assume k is the only index where either $i(w_k)$ or $t(w_k)$ is not a vertex in $G^{(1)}$, where $t(e)$ denotes the terminal vertex of the edge e. First we consider that only one of these two vertices is not in $G^{(1)}$. For the case that $i(w_k)$ is not a vertex in $G^{(1)}$, there is a vertex, say v_k, in $G^{(1)}$ so that v_k is a grouping vertex that contains $i(w_k)$. (In fact, each vertex in $G^{(1)}$ is the grouping of one or more vertices in G, and so is $G^{(2)}$. The reader is referred to [27] for more details.) Hence there is an edge $\overline{w}_{k-1}$ in $G^{(1)}$ such that $i(\overline{w}_{k-1}) = i(w_{k-1})$ and $t(\overline{w}_{k-1}) = v_k$. In other words, there is an edge in $G^{(1)}$ that can be related to w_{k-1}. Moreover, there is an edge $(v_k, t(w_k))$ in $G^{(1)}$ if $t(w_k)$ is a vertex in $G^{(1)}$. Hence there is a cycle in $G^{(1)}$ that corresponds to w. The case that $t(w_k)$ is not a vertex in $G^{(1)}$ can be conducted in an analogous discussion. For the case that both the initial and terminal states of w_k are not in $G^{(1)}$, combining the above demonstration implies there is a new vertex v_{k+1} and two new edges $e_k = (v_k, v_{k+1})$, $e_{k+1} = (v_{k+1}, t(w_{k+1}))$ in $G^{(1)}$. That is, there is still a cycle in $G^{(1)}$ that corresponds to w.

Repeating the process if necessary, it is seen that for every cyclic path in G with length n, there is an associated cyclic path in $G^{(1)}$ with length m, and m divides n. Theorem 9 asserts there exists an infinite-to-one factor $\Phi^{(1)} : \mathbf{Y} \to W^{(1)}$. Let $\pi = \Phi^{(1)} \circ \Phi^{(2)}$. Then π is an infinite-to-one map from $W^{(2)} \to W^{(1)}$ by Theorem 1.

Next, for another case, suppose that condition 7.2 is satisfied. It suffices to demonstrate the existence of an embedding map from $W^{(2)}$ to $\mathbf{Y}$. The elucidation of the existence of a map from $W^{(2)}$ to $\mathbf{Y}$ can be performed via a similar but converse argument, as with the discussion of $\Phi^{(1)}$. Hence we omit the details. Since the graph representation $G^{(2)}$ of $W^{(2)}$ comes from applying subset construction to $(G, \mathcal{L}^{(2)})$, it can be verified that every periodic point in $W^{(2)}$ corresponds to a cyclic path in $G^{(2)}$, and for every cyclic path in $G^{(2)}$ we can illustrate a cyclic path in G. Theorem 8 demonstrates the existence of an embedding map $\overline{\Phi}^{(2)} : W^{(2)} \to \mathbf{Y}$.

This completes the proof. $\square$

Once we demonstrate the existence of a factor map $\pi : W^{(i)} \to W^{(\bar{i})}$, the proof of Theorem 3 can be performed using methods similar to the proof of Theorem 2. Hence we skip the proof. Instead, it is interesting to find if there is a criterion to determine whether π is uniform.

Theorem 10. Suppose N, obtained from π as defined in Section 3.3, satisfies the Markov condition of order k. Define

$$M = \left[m_{J(0,k-1),J(1,k)} \right]_{J \in W^{(1)}_{k+1}};$$

here $m_{J(0,k-1),J(1,k)}$ is defined by equation (20). Then π is uniform if and only if

$$\rho_M = \frac{\rho^{(2)}}{\rho^{(1)}}, \tag{21}$$

where $\rho^{(i)}$ is the spatial radius of the transition matrix $T^{(i)}$ of $W^{(i)}$ for $i = 1, 2$.

Theorem 10 is obtained with a little modification of the proof of Proposition 6.1 in [36], thus we omit it here. The following corollary comes immediately from Theorem 10.

Corollary 2. Let N be defined as above. Suppose N satisfies the Markov condition and equation (21) holds. Then $\bar{\pi}\nu^{(i)} \in \mathcal{M}_{\max}\left(\mathbf{Y}^{(\bar{i})}\right)$ if $\nu^{(i)} \in \mathcal{M}_{\max}\left(\mathbf{Y}^{(i)}\right)$. Furthermore, if

$$\dim \mathbf{Y}^{(i)} = \frac{h_{\nu^{(i)}}\left(\mathbf{Y}^{(i)}\right)}{\log n_i},$$

then

$$\dim \mathbf{Y}^{(\bar{i})} = \frac{h_{\bar{\pi}\nu^{(i)}}\left(\mathbf{Y}^{(\bar{i})}\right)}{\log n_{\bar{i}}}.$$

4. Examples

Example 2. Suppose the templates of an SMCNN are given by the following:

$$\left[a^{(1)}, a_r^{(1)}, z^{(1)} \right] = \left[2.9, \ 1.7, \ 0.1 \right]$$
$$\left[a^{(2)}, a_r^{(2)}, b^{(2)}, z^{(2)} \right] = \left[-0.3, \ -1.2, \ 0.7, \ 2.3, \ 0.9 \right].$$

Then the basic set of admissible local patterns is

$$
\mathcal{B} = \left\{ \begin{array}{|cc|} \hline - & + \\ - & - \\ \hline \end{array}, \begin{array}{|cc|} \hline - & + \\ + & - \\ \hline \end{array}, \begin{array}{|cc|} \hline + & - \\ - & + \\ \hline \end{array}, \begin{array}{|cc|} \hline + & - \\ + & + \\ \hline \end{array}, \begin{array}{|cc|} \hline + & + \\ - & + \\ \hline \end{array}, \begin{array}{|cc|} \hline + & + \\ + & + \\ \hline \end{array} \right\}.
$$

The transition matrix T of the solution space $\mathbf{Y}$ is

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

and the symbolic transition matrices of the hidden and output spaces are

$$
S^{(1)} = \begin{pmatrix} \emptyset & \alpha_1 \\ \alpha_2 & \alpha_3 \end{pmatrix} \text{ and } S^{(2)} = \begin{pmatrix} \emptyset & \alpha_1 & \emptyset \\ \alpha_2 & \emptyset & \alpha_3 \\ \alpha_2 & \emptyset & \alpha_3 \end{pmatrix},
$$

respectively. Figure 1 shows that $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are two different spaces. The topological entropy of $\mathbf{Y}^{(i)}$ is related to the spectral radius of the incidence of $S^{(i)}$. An easy computation implies $h(\mathbf{Y}^{(1)}) = h(\mathbf{Y}^{(2)}) = \log g$, where $g = \left(1 + \sqrt{5}\right) \big/ 2$ is the golden mean.

Let

$$
E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}.
$$

Then $S^{(2)}E = ES^{(1)}$. Proposition 3 indicates that there exist factor maps $\pi : W^{(2)} \to W^{(1)}$ and $\bar{\pi} : \mathbf{Y}^{(2)} \to \mathbf{Y}^{(1)}$. More precisely, let

$$
\mathcal{A}(W^{(1)}) = \{x_1, x_2\}, \qquad \mathcal{A}(W^{(2)}) = \{x_1', x_2', x_3'\};
$$
$$
\mathcal{A}(\mathbf{Y}^{(1)}) = \{y_1, y_2\}, \qquad \mathcal{A}(\mathbf{Y}^{(2)}) = \{y_1', y_2', y_3'\}.
$$

Then

$$
\pi(x_1') = x_1, \qquad \pi(x_2' = \pi(x_3') = x_2;
$$
$$
\bar{\pi}(y_1') = y_1, \qquad \bar{\pi}(y_2') = \bar{\pi}(y_3') = y_2.
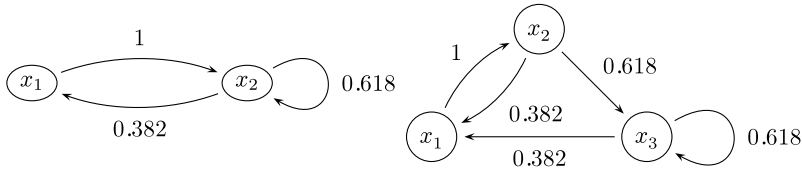$$

See Figure 4.

Figure 4. The graph representation of the hidden and output spaces of Example 2. The number on the edge is the transition probability. The left side represents $\mathbf{Y}^{(1)}$ and the right side represents $\mathbf{Y}^{(2)}$.

Suppose $\mathcal{A}(\mathbf{Y}) = \{z_1, z_2, z_3, z_4\}$; the factor map $\psi : \mathbf{Y} \to \mathbf{Y}^{(1)}$ is given by

$$\psi(z_1) = \psi(z_3) = y_1,$$
$$\psi(z_2) = \psi(z_4) = y_2.$$

Set $N = (N_{ij})_{i \leq i, j \leq 2}$ and L_1, L_2 by

$$N_{11} = N_{21} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad N_{12} = N_{22} = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$$

and $L_1 = (0\ 1)$, $L_2 = (gg)$, respectively. A straightforward calculation demonstrates that

$$L_1 N_{11} = 0 \cdot L_1,\ L_1 N_{12} =$$
$$g^{-1} \cdot L_2,\ L_2 N_{11} = g \cdot L_1,\ L_2 N_{22} = 1 \cdot L_2.$$

That is, N satisfies the Markov condition of order 1. Theorem 4 indicates that $\mathbf{Y}^{(1)}$ is an SFT with the unique maximal measure of entropy $\nu^{(1)}$, and $\nu^{(1),+} = (p_{\mathbf{Y}^{(1)}}, P_{\mathbf{Y}^{(1)}})$, where

$$p_{\mathbf{Y}^{(1)}} = \left(\frac{2-g}{3-g}, \frac{1}{3-g} \right)$$

and

$$P_{\mathbf{Y}^{(1)}} = \begin{pmatrix} 0 & 1 \\ 2-g & g-1 \end{pmatrix} = \text{stoch}(M),$$

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

Applying Theorem 2, we have

$$\dim W^{(1)} = \dim \mathbf{Y}^{(1)} = 2\frac{h_{\gamma^{(1)},+}\left(\mathbf{Y}^{(1)}\right)}{\log 2} =$$

$$\frac{2}{(g-3)\log 2}\left((2-g)\log(2-g)+(g-1)\log(g-1)\right) = 2\frac{\log g}{\log 2}.$$

On the other hand,

$$S^{(2)}S^{(2)} = \begin{pmatrix} \alpha_1\alpha_2 & \emptyset & \alpha_1\alpha_3 \\ \alpha_3\alpha_2 & \alpha_2\alpha_1 & \alpha_3\alpha_3 \\ \alpha_3\alpha_2 & \alpha_2\alpha_1 & \alpha_3\alpha_3 \end{pmatrix}$$

implies that every word of length 3 in $\mathbf{Y}^{(2)}$ is a synchronizing word. That is, $\mathbf{Y}^{(2)}$ is topologically conjugated to $W^{(2)}$. Since the unique maximal measure of $W^{(2)}$ is $\mu^{(2)}$ with $\mu^{(2),+} = (p_{W^{(2)}}, P_{W^{(2)}})$, where

$$p_{W^{(2)}} = \left(\frac{2-g}{3-g}, \frac{2-g}{3-g}, \frac{g-1}{3-g}\right),$$

$$P_{W^{(2)}} = \begin{pmatrix} 0 & 1 & 0 \\ 2-g & 0 & g-1 \\ 2-g & 0 & g-1 \end{pmatrix}.$$

Theorem 2 suggests that

$$\dim W^{(2)} = 2\frac{h_{\mu^{(2)},+}\left(W^{(2)}\right)}{\log 3} = 2\frac{\log g}{\log 3},$$

and

$$\dim \mathbf{Y}^{(2)} = 2\frac{h_{\phi^{(2)}\mu^{(2)},+}\left(\mathbf{Y}^{(2)}\right)}{\log 2} = 2\frac{h_{\mu^{(2)},+}\left(W^{(2)}\right)}{\log 2} = 2\frac{\log g}{\log 2}.$$

$\pi\mu^{(2)} = \mu^{(1)}$ can be verified without difficulty, thus we omit the details.

Example 3. Suppose the template of the first layer is the same as in Example 2, and

$$\left[a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)}\right] = \left[-0.1, -1.1, 2.1, -1.4, 0.9\right].$$

The basic set of admissible local patterns of the solution space $\mathbf{Y}$ is

$$\mathcal{B} = \left\{ \begin{array}{|c|} \hline - + \\ - + \\ \hline \end{array}, \begin{array}{|c|} \hline - - \\ - + \\ \hline \end{array}, \begin{array}{|c|} \hline + - \\ + - \\ \hline \end{array}, \begin{array}{|c|} \hline + + \\ + - \\ \hline \end{array}, \begin{array}{|c|} \hline + - \\ + + \\ \hline \end{array}, \begin{array}{|c|} \hline + - \\ - - \\ \hline \end{array} \right\}.$$

The transition matrix T of the solution space $\mathbf{Y}$ is

$$T = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

After careful examination, the hidden and output spaces are both mixing with symbolic transition matrices

$$S^{(1)} = \begin{pmatrix} \varnothing & \varnothing & \alpha_1 \\ \alpha_0 & \varnothing & \alpha_1 \\ \varnothing & \alpha_2 & \varnothing \end{pmatrix}, \; S^{(2)} = \begin{pmatrix} \varnothing & \varnothing & \alpha_1 & \varnothing \\ \alpha_2 & \varnothing & \varnothing & \varnothing \\ \varnothing & \alpha_3 & \varnothing & \alpha_2 \\ \varnothing & \varnothing & \alpha_1 & \varnothing \end{pmatrix}.$$

See Figure 5. $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$ are FSE since $h(\mathbf{Y}^{(1)}) = h(\mathbf{Y}^{(2)}) = \log \rho$, where $\rho \approx 1.3247$ satisfies $\rho^3 - \rho - 1 = 0$. Let

$$E = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Notably, $T^{(2)}E = ET^{(1)}$ and there exists no factor-like matrix F such that $S^{(2)}F = FS^{(1)}$ or $S^{(1)}F = FS^{(2)}$. It follows from $S^{(1)}$ that every word of length 2 in $\mathbf{Y}^{(1)}$ is a synchronizing word. Hence $\mathbf{Y}^{(1)} \cong W^{(1)}$. The unique maximal measure of entropy for $W^{(1),+}$ is $\mu^{(1),+} = (p_{W^{(1)}}, P_{W^{(1)}})$, where $p_{W^{(1)}} = (0.1770, 0.4115, 0.4115)$ and

$$P_{W^{(1)}} = \begin{pmatrix} 0 & 0 & 1 \\ 0.4302 & 0 & 0.5698 \\ 0 & 1 & 0 \end{pmatrix}.$$

Hence

$$\dim W^{(1)} = 2 \frac{h_{\mu^{(1),+}}(W^{(1)})}{\log 3} \approx 0.5119,$$

and

$$\dim \mathbf{Y}^{(1)} = 2 \frac{h_{\phi^{(1)}\mu^{(1),+}}(\mathbf{Y}^{(1)})}{\log 2} = 2 \frac{h_{\mu^{(1),+}}(W^{(1)})}{\log 2} \approx 0.8114.$$

Unlike Example 2, it can be verified (with or without computer assistance) that $\mathbf{Y}^{(2)}$, rather than an SFT, is a strict sofic shift since there exists no $k \in \mathbb{N}$ such that every word of length k is a synchronizing
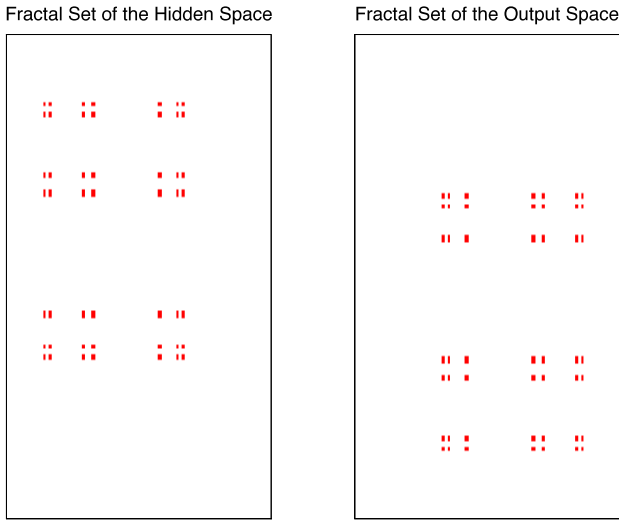
Fractal Set of the Hidden Space   Fractal Set of the Output Space

Figure 5. The fractal sets of the hidden and output spaces of Example 3. The templates are given by $\left[a^{(1)}, a_r^{(1)}, z^{(1)}\right] = [2.9, 1.7, 0.1]$ and $\left[a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)}\right] = [-0.1, -1.1, 2.1, -1.4, 0.9]$. The output space $\mathbf{Y}^{(2)}$ is a strict sofic shift rather than an SFT. Meanwhile, the hidden space $\mathbf{Y}^{(1)}$ is an SFT.

word in $\mathbf{Y}^{(2)}$. Nevertheless, there is a synchronizing word of length 2 (i.e., $\alpha_3 = ++$). Theorem 2.1 indicates that there is a one-to-one correspondence between $\mathcal{M}_{\max}(W^{(2)})$ and $\mathcal{M}_{\max}(\mathbf{Y}^{(2)})$. Since the unique maximal measure of $W^{(2),+}$ is $\mu^{(2),+} = (p_{W^{(2)}}, P_{W^{(2)}})$, where $p_{W^{(2)}} = (0.1770, 0.1770, 0.4115, 0.2345)$ and

$$
P_{W^{(2)}} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0.4302 & 0 & 0.5698 \\ 0 & 0 & 1 & 0 \end{pmatrix},
$$

we have

$$
\dim W^{(2)} = 2\,\frac{h_{\mu^{(1),+}}\left(W^{(1)}\right)}{\log 4} \approx 0.4057,
$$

and

$$
\dim \mathbf{Y}^{(2)} = 2\,\frac{h_{\phi^{(2)}\mu^{(2),+}}\left(\mathbf{Y}^{(2)}\right)}{\log 2} = 2\,\frac{h_{\mu^{(2),+}}\left(W^{(2)}\right)}{\log 2} \approx 0.8114.
$$

Example 4. Suppose the template of the first layer is the same as in Example 2, and

$$\left[a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)}\right] = \left[1.3, -1.2, 0.7, 2.3, 0.8\right].$$

Then the basic set of admissible local patterns is

$$\mathcal{B} = \left\{ \boxed{\begin{matrix}-- \\ --\end{matrix}}, \boxed{\begin{matrix}-+ \\ --\end{matrix}}, \boxed{\begin{matrix}-+ \\ +-\end{matrix}}, \boxed{\begin{matrix}+- \\ -+\end{matrix}}, \boxed{\begin{matrix}+- \\ ++\end{matrix}}, \boxed{\begin{matrix}++ \\ --\end{matrix}}, \boxed{\begin{matrix}++ \\ -+\end{matrix}}, \boxed{\begin{matrix}++ \\ ++\end{matrix}} \right\}.$$

The transition matrix T of the solution space $\mathbf{Y}$,

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

suggests that $\mathbf{Y}$ is mixing. It is not difficult to see that the symbolic transition matrices of the hidden and output spaces are

$$S^{(1)} = \begin{pmatrix} \alpha_0 & \alpha_1 \\ \alpha_2 & \alpha_3 \end{pmatrix}$$

and

$$S^{(2)} = \begin{pmatrix} \alpha_0 & \emptyset & \alpha_1 & \emptyset & \emptyset \\ \emptyset & \emptyset & \alpha_1 & \emptyset & \emptyset \\ \emptyset & \alpha_2 & \emptyset & \alpha_3 & \emptyset \\ \emptyset & \emptyset & \emptyset & \alpha_3 & \alpha_2 \\ \alpha_0 & \emptyset & \alpha_1 & \emptyset & \emptyset \end{pmatrix},$$

respectively. See Figure 6 for the fractal sets of $\mathbf{Y}^{(1)}$ and $\mathbf{Y}^{(2)}$.

Obviously $\mathbf{Y}^{(1)}$ is a full 2-shift. It is remarkable that $\phi^{(1)}\mu^{(1)}$ is not a Markov measure. The unique maximal measure for $W^{(1),+}$ (also for $\mathbf{Y}^{(1),+}$) is the uniform Bernoulli measure $\mu^{(1),+} = \left(1/2, 1/2\right)$. Therefore,

$$\dim W^{(1)} = \dim \mathbf{Y}^{(1)} = 2 \frac{h_{\mu^{(1),+}}\left(W^{(1)}\right)}{\log 2} = 2.$$

Since $h\left(W^{(2)}\right) = \log \rho$, where $\rho \approx 1.8668$ satisfies $\rho^4 - 2\rho^3 + \rho - 1 = 0$, the factor map $\pi \colon W^{(1)} \to W^{(2)}$ must be infinite-to-one if it exists. The fact $W^{(2)}$ has two fixed points, which can be seen from $T^{(2)}$, asserts that there exists an infinite-to-one factor map $\pi \colon W^{(1)} \to W^{(2)}$ by Theorem 9. However, it is difficult to find the explicit form of π.
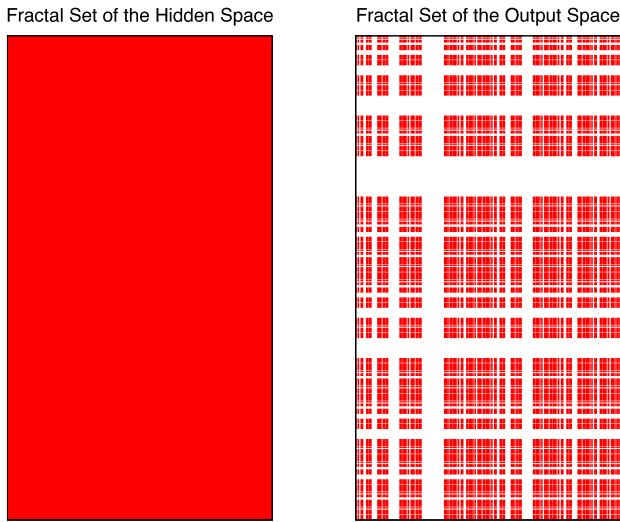
Fractal Set of the Hidden Space   Fractal Set of the Output Space

Figure 6. The fractal sets of the hidden and output spaces of Example 4. The templates are given by $\left[a^{(1)}, a_r^{(1)}, z^{(1)}\right] = [2.9, 1.7, 0.1]$ and $\left[a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)}\right] = [1.3, -1.2, 0.7, 2.3, 0.8]$. It is seen that the hidden space $\mathbf{Y}^{(1)}$ is the unit square $[0, 1] \times [0, 1]$, and so is $W^{(1)}$. Moreover, there are infinite-to-one factor maps $\pi : W^{(1)} \to W^{(2)}$ and $\pi : \mathbf{Y}^{(1)} \to \mathbf{Y}^{(2)}$.

Since the unique maximal measure of $W^{(2),+}$ is $\mu^{(2),+} = (p_{W^{(2)}}, P_{W^{(2)}})$ with $p_{W^{(2)}} = (0.1888, 0.0658, 0.2294, 0.3524, 0.1636)$ and

$$
P_{W^{(2)}} = \begin{pmatrix}
0.5357 & 0 & 0.4643 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0.2870 & 0 & 0.7130 & 0 \\
0 & 0 & 0 & 0.5357 & 0.4643 \\
0.5357 & 0 & 0.4643 & 0 & 0
\end{pmatrix},
$$

the Hausdorff dimension of $W^{(2)}$ is

$$
\dim W^{(2)} = 2 \frac{h_{\mu^{(2),+}}\left(W^{(2)}\right)}{\log 5} \approx 0.7758.
$$

Since $W^{(2)}$ is mixing, we have

$$
\dim \mathbf{Y}^{(2)} = 2 \frac{h_{\nu^{(2),+}}\left(\mathbf{Y}^{(2)}\right)}{\log 2} =
$$
$$
2 \frac{h_{\phi^{(2)}\mu^{(2),+}}\left(\mathbf{Y}^{(2)}\right)}{\log 2} = 2 \frac{h_{\mu^{(2),+}}\left(W^{(2)}\right)}{\log 2} \approx 1.8012.
$$

As a conclusion, in the present example, an infinite-to-one factor map is associated with a different Hausdorff dimension.

Example 5. Suppose the template of the first layer is the same as in Example 2, and

$$\left[a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)}\right] = [0.7, -1.1, 2.1, -1.4, 1.7].$$

The basic set of admissible local patterns of the solution space **Y** is

$$\mathcal{B} = \left\{ \begin{smallmatrix} - & + \\ - & + \end{smallmatrix}, \begin{smallmatrix} - & - \\ - & + \end{smallmatrix}, \begin{smallmatrix} + & - \\ + & - \end{smallmatrix}, \begin{smallmatrix} + & + \\ + & - \end{smallmatrix}, \begin{smallmatrix} + & - \\ + & + \end{smallmatrix}, \begin{smallmatrix} + & - \\ - & - \end{smallmatrix}, \begin{smallmatrix} + & + \\ + & + \end{smallmatrix} \right\}.$$

The transition matrix T of the solution space **Y** is

$$T = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

A straightforward examination shows that the hidden and output spaces are both mixing with symbolic transition matrices

$$S^{(1)} = \begin{pmatrix} \emptyset & \emptyset & \alpha_1 \\ \alpha_0 & \emptyset & \alpha_1 \\ \emptyset & \alpha_2 & \alpha_3 \end{pmatrix}, \; S^{(2)} = \begin{pmatrix} \emptyset & \alpha_2 & \alpha_3 \\ \alpha_1 & \emptyset & \emptyset \\ \emptyset & \alpha_0 & \alpha_3 \end{pmatrix}.$$

$h(\mathbf{Y}^{(1)}) = \log \rho$ and $h(\mathbf{Y}^{(2)}) = \log g$, where $\rho \approx 1.8393$ satisfies $\rho^3 - \rho^2 - \rho - 1 = 0$. See Figure 7.

Since $W^{(2)}$ has a fixed point, Theorem 9 implies there is an infinite-to-one factor map $\pi: W(1) \to W^{(2)}$. The unique maximal measure of $W^{(1),+}$ is $\mu^{(1),+} = (p_{W^{(1)}}, P_{W^{(1)}})$ with $p_{W^{(1)}} = (0.0994, 0.2822, 0.6184)$ and

$$P_{W^{(1)}} = \begin{pmatrix} 0 & 0 & 1 \\ 0.3522 & 0 & 0.6478 \\ 0 & 0.4563 & 0.5437 \end{pmatrix}.$$

This suggests

$$\dim W^{(1)} = 2 \frac{h_{\mu^{(1),+}}(W^{(1)})}{\log 3} \approx 1.1094.$$

The symbolic transition matrix $S^{(1)}$ asserts that every word of length 2 in $\mathbf{Y}^{(1)}$ is a synchronizing word, hence $\mathbf{Y}^{(1)}$ is topologically conjugated to $W^{(1)}$ and

$$\dim \mathbf{Y}^{(1)} = 2 \frac{h_{\nu^{(1),+}}(\mathbf{Y}^{(1)})}{\log 2} = 2 \frac{h_{\phi^{(1)}\mu^{(1),+}}(\mathbf{Y}^{(1)})}{\log 2} \approx 1.7582.$$
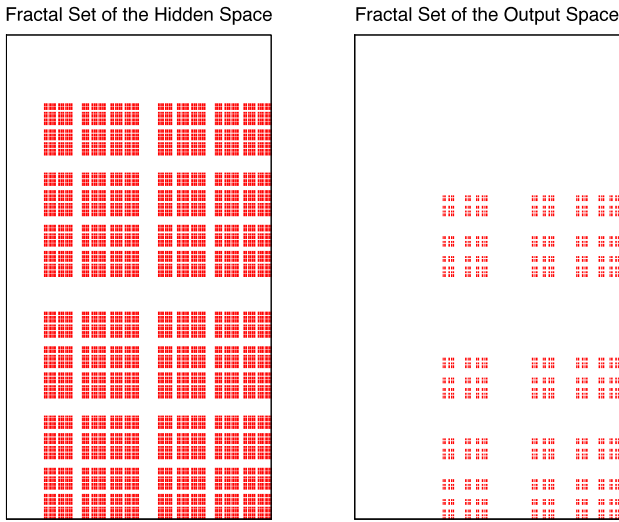
Fractal Set of the Hidden Space   Fractal Set of the Output Space

Figure 7. The fractal sets of the hidden and output spaces of Example 5. The templates are given by $\left[a^{(1)}, a_r^{(1)}, z^{(1)}\right] = [2.9, 1.7, 0.1]$ and $\left[a^{(2)}, a_r^{(2)}, b^{(2)}, b_r^{(2)}, z^{(2)}\right] = [0.7, -1.1, 2.1, -1.4, 1.7]$. It is demonstrated that there is an infinite-to-one factor map $\pi\colon W^{(1)} \to W^{(2)}$, and $\mathbf{Y}^{(1)}$, $\mathbf{Y}^{(2)}$ are strictly sofic.

On the other hand, it is verified that the unique maximal measure of $W^{(2),+}$ is $\mu^{(2),+} = (p_{W^{(2)}}, P_{W^{(2)}})$ with

$$p_{W^{(2)}} = \left(\frac{2-g}{3-g}, \frac{2-g}{3-g}, \frac{g-1}{3-g}\right)$$

and

$$P_{W^{(2)}} = \begin{pmatrix} 0 & 2-g & g-1 \\ 1 & 0 & 0 \\ 0 & 2-g & g-1 \end{pmatrix}.$$

Since every word of length 2 in $\mathbf{Y}^{(2)}$ is a synchronizing word, we have

$$\dim W^{(2)} = 2\,\frac{h_{\mu^{(2),+}}\left(W^{(2)}\right)}{\log 3} = 2\,\frac{\log g}{\log 3} \approx 0.8760,$$

and

$$\dim \mathbf{Y}^{(2)} = 2\,\frac{h_{\nu^{(2),+}}\left(\mathbf{Y}^{(2)}\right)}{\log 2} = 2\,\frac{h_{\phi^{(2)}\mu^{(2),+}}\left(\mathbf{Y}^{(2)}\right)}{\log 2} \approx 1.3884.$$

5. Relation between the Hausdorff Dimensions of Two Hidden Spaces

Theorems 2 and 3 can be extended to two spaces that are induced from a general n-layer cellular neural network equation (1) similar to the discussion in previous sections. Hence we illustrate the results without providing a detailed argument. The solution space $\mathbf{Y}$ of equation (1) is determined by

$$\mathcal{B} \equiv \mathcal{B}\big(A^{(1)}, \ldots, A^{(n)}, B^{(1)}, \ldots, B^{(n)}, z^{(1)}, \ldots, z^{(n)}\big) =$$
$$\left\{ \begin{bmatrix} y_{-d}^{(n)} \cdots y_{-1}^{(n)} y_0^{(n)} y_1^{(n)} \cdots y_d^{(n)} \\ \vdots \\ y_{-d}^{(2)} \cdots y_{-1}^{(2)} y_0^{(2)} y_1^{(2)} \cdots y_d^{(2)} \\ y_{-d}^{(1)} \cdots y_{-1}^{(1)} y_0^{(1)} y_1^{(1)} \cdots y_d^{(1)} \end{bmatrix} \right\} \subseteq \{-1, 1\}^{\mathbb{Z}_{(2d+1) \times n}}.$$

For $1 \leq \ell \leq n$, set

$$\mathcal{L}^{(\ell)}\big(y_{-d}^{(n)} \cdots y_d^{(n)} \diamond \cdots \diamond y_{-d}^{(1)} \cdots y_d^{(1)}\big) = y_{-d}^{(\ell)} \cdots y_d^{(\ell)}.$$

The hidden space $\mathbf{Y}^{(\ell)}$ is then defined by $\mathcal{L}^{(\ell)}$ as before. (For simplicity, we also call $\mathbf{Y}^{(n)}$ a hidden space instead of the output space.) Similarly, $\mathbf{Y}^{(\ell)}$ is a sofic shift with respect to a right-resolving finite-to-one factor map $\phi^{(\ell)} : W^{(\ell)} \to \mathbf{Y}^{(\ell)}$ and an SFT $W^{(\ell)}$. Furthermore, $W^{(\ell)}$ can be described by the transition matrix $T^{(\ell)}$, while $\mathbf{Y}^{(\ell)}$ can be completely described by the symbolic transition matrix $S^{(\ell)}$.

For $1 \leq i, j \leq n$, without the loss of generality, we assume that $h(\mathbf{Y}^{(i)}) \geq h(\mathbf{Y}^{(j)})$ and $\mathcal{A}(\mathbf{Y}^{(i)}) \geq \mathcal{A}(\mathbf{Y}^{(j)})$.

Proposition 4. Suppose $h(\mathbf{Y}^{(i)}) = h(\mathbf{Y}^{(j)})$. If there exists a factor-like matrix E such that $S^{(i)}E = ES^{(j)}$, then there are finite-to-one factor maps $\pi_{ij} : W^{(i)} \to W^{(j)}$ and $\bar{\pi}_{ij} : \mathbf{Y}^{(i)} \to \mathbf{Y}^{(j)}$. For the case where $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$ attain distinct topological entropies, there is an infinite-to-one factor map $\pi_{ij} : W^{(i)} \to W^{(j)}$ if $|\mathcal{A}(W^{(i)})| > |\mathcal{A}(\mathbf{Y})|$ and there exists a factor-like matrix F such that $T^{(i)}F = FT$.

The relation of the Hausdorff dimensions of $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$, if they exist, is organized as follows.

Theorem 11. Suppose $W^{(i)}$ and $W^{(j)}$ are irreducible SFTs, and there exists a factor map $\pi_{ij} : W^{(i)} \to W^{(j)}$. (See Figure 8 for a flow chart.)

Case I. $\mathbf{Y}^{(i)}$, $\mathbf{Y}^{(j)}$ share the same topological entropy.

1. There is a one-to-one correspondence between $\mathcal{M}_{\max}(W^{(\ell)})$ and $\mathcal{M}_{\max}(\mathbf{Y}^{(\ell)})$, where $\ell = i, j$.
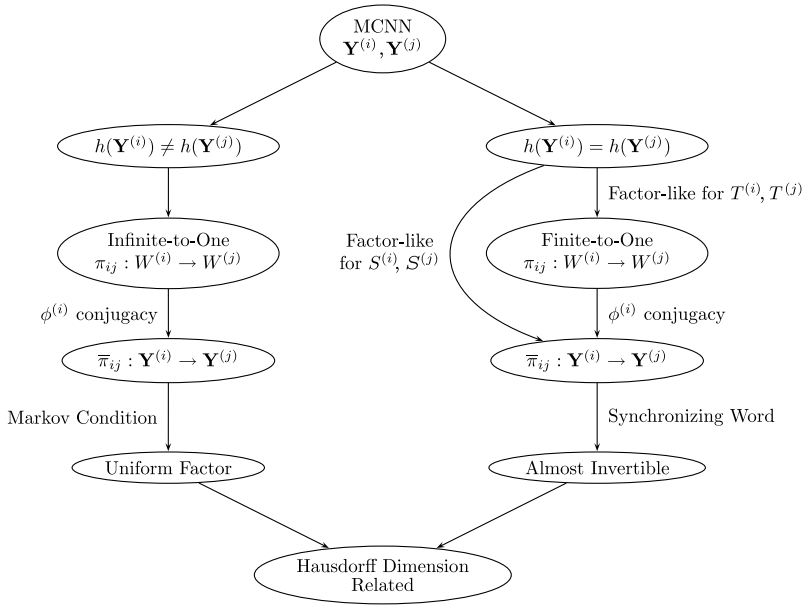
Figure 8. The flow chart of the existence of factor maps for arbitrary two hidden spaces.

2. Let $m_\ell = |\mathcal{A}(W^{(\ell)})|$, $n_\ell = |\mathcal{A}(Y^{(\ell)})|$, and $\mu^{(\ell)}$ be a maximal measure of $W^{(\ell)}$. If $\phi^{(i)}$ has a synchronizing word, then

$$\dim W^{(\ell)} = \frac{h_{\mu^{(\ell)}}(W^{(\ell)})}{\log m_\ell} \quad \text{and} \quad \dim Y^{(\ell)} = \frac{h_{\mu^{(\ell)}}(W^{(\ell)})}{\log n_\ell}.$$

3. Suppose $\nu^{(\ell)} = \phi^{(\ell)}\mu^{(\ell)}$. If

$$\dim Y^{(i)} = \frac{h_{\gamma^{(i)}}(Y^{(i)})}{\log n_i},$$

then

$$\dim Y^{(j)} = \frac{h_{\overline{\pi}\nu^{(i)}}(Y^{(j)})}{\log n_j} = \frac{h_{\gamma^{(i)}}(Y^{(j)})}{\log n_j}$$

for some $\overline{\pi}$.

Case II. $Y^{(i)}, Y^{(j)}$ are associated with distinct topological entropies.

1. Suppose $\pi_{ij} : W^{(i)} \to W^{(j)}$ is a uniform factor. If

$$\dim W^{(i)} = \frac{h_{\mu^{(i)}}(W^{(i)})}{\log m_i},$$

then

$$\dim W^{(j)} = \frac{h_{\mu^{(j)}}(W^{(j)})}{\log m_j} = \frac{h_{\pi\mu^{(i)}}(W^{(j)})}{\log m_j}.$$

2. If $\phi^{(i)}$ has a synchronizing word, then there exists a factor map $\pi: \mathcal{M}_{\max}(\mathbf{Y}^{(i)}) \to \mathcal{M}_{\max}(\mathbf{Y}^{(\bar{i})})$.

3. If

$$\dim \mathbf{Y}^{(i)} = \frac{h_{\gamma^{(i)}}(\mathbf{Y}^{(i)})}{\log n_i},$$

then

$$\dim \mathbf{Y}^{(j)} = \frac{h_{\bar{\pi}\gamma^{(i)}}(\mathbf{Y}^{(j)})}{\log n_j}.$$

6. Conclusion and Further Discussion

This investigation elucidates whether there is a factor map π (resp. $\bar{\pi}$) connecting $W^{(i)}$ and $W^{(j)}$ (resp. $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$). If a factor map does exist, the push-forward measure of a maximal measure is also a maximal measure, provided the factor map is either finite-to-one or uniform. Moreover, the Hausdorff dimensions of two spaces are thus related. Topological entropy provides a medium to make the discussion clearer.

When $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$ are FSE, the existence of a factor-like matrix asserts the existence of factor map $\bar{\pi}$. With the assistance of computer programs, we can rapidly determine if there exists a factor-like matrix for a given multilayer cellular neural network (MCNN). Moreover, the factor map $\bar{\pi}$ can be expressed in an explicit form. For most of the cases, there is no factor-like matrix for $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$.

Problem 3. Suppose there is a factor map between $\mathbf{Y}^{(i)}$ and $\mathbf{Y}^{(j)}$. Is $\dim \mathbf{Y}^{(i)}$ related to $\dim \mathbf{Y}^{(j)}$? Or, equivalently, is there a one-to-one correspondence between $\mathcal{M}_{\max}(\mathbf{Y}^{(i)})$ and $\mathcal{M}_{\max}(\mathbf{Y}^{(j)})$?

A partial result of Problem 3 is the existence of synchronizing words. Lemma 1 demonstrates that if $\phi^{(i)}/\phi^{(j)}$ has a synchronizing word, then $\phi^{(i)}/\phi^{(j)}$ is almost invertible. This implies a one-to-one correspondence between $\mathcal{M}_{\max}(\mathbf{Y}^{(i)})$ and $\mathcal{M}_{\max}(\mathbf{Y}^{(j)})$.

Problem 4. How large is the portion of almost invertible maps in the collection of factor maps?

If $h(\mathbf{Y}^{(i)}) \neq h(\mathbf{Y}^{(j)})$, on the other hand, we propose a criterion for the existence of factor maps. We will not find the explicit form of the factor map.

Problem 5. Can we find some methodology so that we can write down the explicit form of a factor map if it exists?

For the case where $h(\mathbf{Y}^{(i)}) \neq h(\mathbf{Y}^{(j)})$, a uniform factor provides the one-to-one correspondence between the maximal measures of two spaces. When the Markov condition is satisfied, Theorem 10 indicates an if-and-only-if criterion. Notably, we can use Theorem 10 only if the explicit form of the factor map is found. Therefore, the most difficult part is the determination of a uniform factor.

Problem 6. How to find, in general, a uniform factor?

Acknowledgments

References

[1] J.-C. Ban, C.-H. Chang, and S.-S. Lin, "On the Structure of Multilayer Cellular Neural Networks," *Journal of Differential Equations*, **252**(8), 2012 pp. 4563–4597. doi:10.1016/j.jde.2012.01.006.

[2] P. Arena, S. Baglio, L. Fortuna, and G. Manganaro, "Self-Organization in a Two-Layer CNN," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, **45**(2), 1998 pp. 157–162. doi:10.1109/81.661681.

[3] L. O. Chua and T. Roska, *Cellular Neural Networks and Visual Computing*, New York: Cambridge University Press, 2002.

[4] L. O. Chua and L. Yang, "Cellular Neural Networks: Applications," *IEEE Transactions on Circuits and Systems*, **35**(10), 1988 pp. 1273–1290. doi:10.1109/31.7601.

[5] K. R. Crounse and L. O. Chua, "Methods for Image Processing and Pattern Formation in Cellular Neural Networks: A Tutorial," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, **42**(10), 1995 pp. 583–601. doi:10.1109/81.473566.

[6] K. R. Crounse, T. Roska, and L. O. Chua, "Image Halftoning with Cellular Neural Networks," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, **40**(4), 1993 pp. 267–283. doi:10.1109/82.224318.

[7] X. Li, "Analysis of Complete Stability for Discrete-Time Cellular Neural Networks with Piecewise Linear Output Functions," *Neural Computation*, **21**(5), 2009 pp. 1434–1458. doi:10.1162/neco.2008.04-08-744.

[8] V. Murugesh, "Image Processing Applications via Time-Multiplexing Cellular Neural Network Simulator with Numerical Integration Algorithms," *International Journal of Computer Mathematics*, **87**(4), 2010 pp. 840–848. doi:10.1080/00207160802217219.

[9] J. Peng, D. Zhang, and X. Liao, "A Digital Image Encryption Algorithm Based on Hyper-Chaotic Cellular Neural Network," *Fundamenta Informaticae*, **90**(3), 2009 pp. 269–282. doi:10.3233/FI-2009-0018.

[10] S. Xavier-de Souza, M. E. Yalcin, J. A. K. Suykens, and J. Vandewalle, "Toward CNN Chip-Specific Robustness," *IEEE Transactions on Circuits and Systems I: Regular Papers*, **51**(5), 2004 pp. 892–902. doi:10.1109/TCSI.2004.827618.

[11] Z. Yang, Y. Nishio, and A. Ushida, "Image Processing of Two-Layer CNNs: Applications and Their Stability," *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences*, **E85-A**(9), 2002 pp. 2052–2060.

[12] S.-N. Chow and J. Mallet-Paret, "Pattern Formation and Spatial Chaos in Lattice Dynamical Systems: I and II," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, **42**(10), 1995 pp. 746–756.

[13] S.-N. Chow, J. Mallet-Paret, and E. S. van Vleck, "Dynamics of Lattice Differential Equations," *International Journal of Bifurcation and Chaos*, **6**(9), 1996 pp. 1605–1621. doi:10.1142/S0218127496000977.

[14] S.-N. Chow, J. Mallet-Paret, and E. S. van Vleck, "Pattern Formation and Spatial Chaos in Spatially Discrete Evolution Equations," *Random and Computational Dynamics*, **4**(2), 1996 pp. 109–178.

[15] S.-N. Chow and W. X. Shen, "Dynamics in a Discrete Nagumo Equation: Spatial Topological Chaos," *SIAM Journal on Applied Mathematics*, **55**(6), 1995 pp. 1764–1781. doi:10.1137/S0036139994261757.

[16] K. R. Crounse, L. O. Chua, P. Thiran, and G. Setti, "Characterization and Dynamics of Pattern Formation in Cellular Neural Networks," *International Journal of Bifurcation and Chaos*, **6**(9), 1996 pp. 1703–1724. doi:10.1142/S0218127496001053.

[17] M. Gilli, M. Biey, and P. Checco, "Equilibrium Analysis of Cellular Neural Networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, **51**(5), 2004 pp. 903–912. doi:10.1109/TCSI.2004.827624.

[18] M. Itoh and L. O. Chua, "Structurally Stable Two-Cell Cellular Neural Networks," *International Journal of Bifurcation and Chaos*, **14**(8), 2004 pp. 2579–2653. doi:10.1142/S0218127404010977.

[19] J. Juang and S.-S. Lin, "Cellular Neural Networks: Mosaic Pattern and Spatial Chaos," *SIAM Journal on Applied Mathematics*, **60**(3), 2000 pp. 891–915. doi:10.1137/S0036139997323607.

[20] Y.-Q. Ke and C.-F. Miao, "Existence Analysis of Stationary Solutions for RTD-Based Cellular Neural Networks," *International Journal of Bifurcation and Chaos*, **20**(7), 2010 pp. 2123–2136. doi:10.1142/S0218127410026964.

[21] X. M. Li and L. H. Huang, "On the Complete Stability of Cellular Neural Networks with External Inputs and Bias," *Acta Mathematicae Applicatae Sinica*, **26**(3), 2003 pp. 475–486.

[22] S.-S. Lin and C.-W. Shih, "Complete Stability for Standard Cellular Neural Networks," *International Journal of Bifurcation and Chaos*, **9**(5), 1999 pp. 909–918. doi:10.1142/S021812749900064X.

[23] C.-W. Shih, "Influence of Boundary Conditions on Pattern Formation and Spatial Chaos in Lattice Systems," *SIAM Journal on Applied Mathematics*, **61**(1), 2000 pp. 335–368. doi:10.1137/S0036139998340650.

[24] J.-C. Ban and C.-H. Chang, "On the Monotonicity of Entropy for Multilayer Cellular Neural Networks," *International Journal of Bifurcation and Chaos*, **19**(11), 2009 pp. 3657–3670. doi:10.1142/S0218127409025043.

[25] J.-C. Ban and C.-H. Chang, "The Spatial Complexity of Inhomogeneous Multilayer Neural Networks," *Neural Processing Letters*, 2014 pp. 1–17. doi:10.1007/s11063-014-9400-7.

[26] S. Wolfram, *A New Kind of Science*, Champaign IL: Wolfram Media, Inc., 2002.

[27] J.-C. Ban, C.-H. Chang, S.-S. Lin, and Y.-H Lin, "Spatial Complexity in Multi-layer Cellular Neural Networks," *Journal of Differential Equations*, **246**(2), 2009 pp. 552–580. doi:10.1016/j.jde.2008.05.004.

[28] M. Boyle and K. Petersen, "Hidden Markov Processes in the Context of Symbolic Dynamics," in *Entropy of Hidden Markov Processes and Connections to Dynamical Systems* (B. Marcus, K. Petersen, and T. Weissman, eds.), New York: Cambridge University Press, 2011 pp. 5–71.

[29] Y. B. Pesin, *Dimension Theory in Dynamical Systems: Contemporary Views and Application*, Chicago: The University of Chicago Press, 1997.

[30] J.-C. Ban, C.-H. Chang, and T.-J. Chen, "Measures of the Full Hausdorff Dimension for a General Sierpinski Carpet." arxiv.org/abs/1206.1190v1.

[31] B. Kitchens and S. Tuncel, "Finitary Measures for Subshifts of Finite Type and Sofic Systems," *Memoirs of the American Mathematical Society*, **58**(338), 1985 pp. 1–68. doi:10.1090/memo/0338.

[32] D. A. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, New York: Cambridge University Press, 1995.

[33] R. Bowen, "Hausdorff Dimension of Quasi-Circles," *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, **50**(1), 1979 pp. 11–25. doi:10.1007/BF02684767.

[34] B. P. Kitchens, *Symbolic Dynamics. One-Sided, Two-Sided, and Countable State Markov Shifts*, New York: Springer-Verlag, 1998.

[35] J.-C. Ban and C.-H. Chang, "Diamond in Multi-layer Cellular Neural Networks," *Applied Mathematics and Computation*, **222**, 2013 pp. 1–12. doi:10.1016/j.amc.2013.07.010.

[36] M. Boyle and S. Tuncel, "Infinite-to-One Codes and Markov Measures," *Transactions of the American Mathematical Society*, **285**(2), 1984 pp. 657–684. doi:10.1090/S0002-9947-1984-0752497-0.

[37] D. Blackwell, "The Entropy of Functions of Finite State Markov Chains," in *Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Liblice, Czechoslovakia 1956, Prague: Ceskoslovenska akademie ved. Sekce technicka, 1957 pp. 13–20.

[38] R. Meester and J. E. Steif, "Higher-Dimensional Subshifts of Finite Type, Factor Maps and Measures of Maximal Entropy," *Pacific Journal of Mathematics*, **200**(2), 2001 pp. 497–510.
       doi:10.2140/pjm.2001.200.497.