On the Monotonicity of Entropy for Multilayer Cellular Neural Networks.

Article *in* International Journal of Bifurcation and Chaos · November 2009

DOI: 10.1142/S0218127409025043 · Source: DBLP

International Journal of Bifurcation and Chaos, Vol. 19, No. 11 (2009) 3657–3670

ON THE MONOTONICITY OF ENTROPY FOR MULTILAYER CELLULAR NEURAL NETWORKS

JUNG-CHAO BAN
Department of Applied Mathematics,
National Dong Hwa University,
Hualian, 97063, Taiwan, R.O.C.
Taida Institute for Mathematical Sciences,
National Taiwan University,
Taipei, 10617, Taiwan, R.O.C.
jcban@mail.ndhu.edu.tw

CHIH-HUNG CHANG
National Center for Theoretical Sciences,
National Tsing Hua University,
Hsinchu, 30043, Taiwan, R.O.C.
chihhung@math.cts.nthu.edu.tw

Received March 20, 2008; Revised April 20, 2009

This work investigates the monotonicity of topological entropy for one-dimensional multilayer cellular neural networks. The interacting radius and number of layers are treated as parameters. Fix either one of them; the set of topological entropies grows as a strictly nested sequence with respect to one another. Apart from the comparison of the set of topological entropies, maximal and minimal templates are indicators of a dynamical system. Our results demonstrate that maximal and minimal templates of larger interacting radius (respectively number of layers) dominate those of smaller one. To be precise, the strict monotonicity of topological entropy is demonstrated through the comparison of the maximal and minimal templates as the parameters are varied.

Keywords: Cellular neural networks; sofic shift; topological entropy; forcing relation; maximal template; minimal template.

1. Introduction

Multilayer cellular neural networks (MCNNs) are known as locally connected networks given by

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

for $d \in \mathbb{N}, i \in \mathbb{Z}, 1 \le \ell \le n, z^{(\ell)}, u_i^{(1)} \in \mathbb{R}$ and

$$u_i^{(\ell+1)} = y_i^{(\ell)} = f(x_i^{(\ell)}) = \frac{1}{2}(|x_i^{(\ell)} + 1| - |x_i^{(\ell)} - 1|). \tag{2}$$

The piecewise linear function $f : \mathbb{R} \to \mathbb{R}$ is called the output function, $u_i^{(\ell)}$ is called the input term and d and n is the *interacting radius* and *number of layers*, respectively. Cellular neural networks (CNNs) have been widely applied in studying the signal propagation between neurons, image processing, information technology and VLSI since it was introduced [Chua, 1998; Chua & Yang, 1988a, 1988b; Crounse *et al.*, 1993; Lee & Pineda de Gyvez, 1996]. The study of equilibrium solution of (1) is essential for the understanding of CNNs (see [Chow *et al.*, 1996; Mallet-Paret & Chow, 1995] and references therein for more details). This gives rise to the elucidation of the following equation,

$$x_i^{(\ell)} = \sum_{|k|\le d} (a_k^{(\ell)} f(x_{i+k}^{(\ell)}) + b_k^{(\ell)} f(x_{i+k}^{(\ell-1)})) + z^{(\ell)}, \quad 2 \le \ell \le n \qquad (3)$$

$$x_i^{(1)} = \sum_{|k|\le d} (a_k^{(1)} f(x_{i+k}^{(1)}) + b_k^{(1)} u_{i+k}) + z^{(1)}.$$

A solution $\mathbf{x} = \{x_i^{(\ell)}\}$ of (3) is called mosaic provided $|x_i^{(\ell)}| > 1$ for all $i \in \mathbb{Z}, 1 \le \ell \le n$. A so-called mosaic pattern is an image of a mosaic solution under $f(x)$. The output space, denoted by

$$\mathscr{O} = \{y = \{f(x_i^{(n)})\} : \{x_i^{(n)}\} \in \mathbf{x} \text{ for some } \mathbf{x}\}, \qquad (4)$$

has been intensively investigated and many phenomena are observed [Ban *et al.*, 2009; Hsu *et al.*, 2000; Juang & Lin, 2000]. One of the indicators for the study of $\mathscr{O}$ is topological entropy which measures the capacity of output space. This quantity is important because of its invariance, and positive topological entropy means such a system is complex enough and possesses exponential growth rate of number of output patterns.

Set $\mathcal{P} = \{(A^{(1)}, \ldots, A^{(n)}, B^{(1)}, \ldots, B^{(n)}, z^{(1)}, \ldots, z^{(n)})\} \subseteq \mathbb{R}^{(4d+3)n}$ the parameter space of (3), where $A^{(\ell)} = (a_{-d}^{(\ell)}, \ldots, a_d^{(\ell)})$ and $B^{(\ell)} = (b_{-d}^{(\ell)}, \ldots, b_d^{(\ell)})$ for $1 \le \ell \le n$. Ban *et al.* [2009] demonstrated that $\mathcal{P}$ can be partitioned into finite regions such that there is a one-to-one correspondence between the partitions of the parameter space and the family of basic sets of feasible local patterns. Since (3) is locally coupled and is spatially homogeneous, $\mathscr{O}$ is completely determined by feasible local patterns [Ban *et al.*, 2009]. For $P_1, P_2 \in \mathcal{P}$ such that $P_1 \ne P_2$, denote by $P_1 \sim P_2$ if they admit the same set of feasible local patterns, i.e. P_1 and P_2 are in the same partition. It can be easily checked that $\sim$ is an equivalence relation. In the forthcoming, a presentation parameter $P(\mathbf{A}, \mathbf{B}, \mathbf{z}) = (A^{(1)}, \ldots, A^{(n)}, B^{(1)}, \ldots, B^{(n)}, z^{(1)}, \ldots, z^{(n)}) \in \mathcal{P}_\sim$ is used to indicate its corresponding mosaic patterns for clarification. $P = P(\mathbf{A}, \mathbf{B}, \mathbf{z}) \in \mathcal{P}_\sim$ is called a template and the set of feasible local patterns admitted by P is written by $\mathcal{B} = \mathcal{B}(P)$.

Let $h(P)$ denote the topological entropy of output space induced by template P, it is obvious that $0 \le h(P) \le \log 2$ for all $P \in \mathcal{P}_\sim$. (The reader may refer to [Ban *et al.*, 2009; Juang & Lin, 2000] for more details.) Set $\mathcal{E} = \{P \in \mathcal{P}_\sim : h(P) \ne 0, \log 2\}$, a template $P \in \mathcal{E}$ is called maximal (respectively minimal) if $h(P) \ge h(Q)$ (respectively $h(P) \le h(Q)$) for all $Q \in \mathcal{E}$. Two natural questions then arise.

Problem 1. Fix the number of layers as n. Let $\mathcal{P}_d$ denote the parameter space of (3) for $d \in \mathbb{N}$.

(i) Let $H(d) = \{h(P) : P \in \mathcal{P}_d\}$ be the collection of all possible topological entropies of $\mathscr{O}$ possessed by $\mathcal{P}_d$. Is $\{H(d)\}$ a strictly nested sequence of sets, i.e. $H(1) \subsetneq H(2) \subsetneq \cdots \subsetneq H(d) \subsetneq \cdots$?
(ii) Does $H(d_2)$ dominate $H(d_1)$ for $d_1 < d_2$? More precisely, let P_{d_1} and P_{d_2} (respectively p_{d_1} and p_{d_2}) be maximal (respectively minimal) templates of $\mathcal{P}_{d_1}$ and $\mathcal{P}_{d_2}$, respectively, is $h(P_{d_1}) \lneqq h(P_{d_2})$ (respectively $h(p_{d_1}) \gneqq h(p_{d_2})$)?

The establishment of strict monotonicity indicates that the increase in coupled cells does enhance the capacity of MCNNs. This motivates the elucidation of this work. It is easy to see that $\{H(d)\}$ is monotone, but the strict monotonicity fails in general. A well-known fact is the Feigenbaum constant, which is a universal constant for functions approaching chaos via period doubling [Feigenbaum, 1979]. Another interesting phenomenon is the discovery of *window* in unimodal map $1 - \mu x^2$ and logistic map $\nu x(1 - x)$ [May, 1976]. When the parameter μ (or ν) is varied in the window, the entropy function is a constant. For example, $f_\nu(x) = \nu x(1 - x)$ has only periodic orbits with period 2^k for some $k \in \mathbb{N}$ as $0 \le \nu \le 3.6$, thus $h(f_\nu) = 0$ for $0 \le \nu \le 3.6$ [Bowen & Frank, 1976; Misiurewicz & Szlenk, 1980].

Similarly, if interacting radius d is fixed, it is then interesting to investigate the following problem.

Problem 2. Let $\mathcal{P}_n$ denote the parameter space of (3) for $n \in \mathbb{N}$.

(i) Is $\{H(n)\}$ a strictly nested sequence of sets?
(ii) Does $H(n_2)$ dominate $H(n_1)$ for $n_1 < n_2$?

The main difficulty of Problems 1 and 2 comes from the constraint of basic set of feasible local patterns. Hsu *et al.* [2000] showed that the basic set of feasible local patterns exhibited by CNN without input is constrained by separation property (see Appendix B for more details). This leads to the study of geometrical structure of $\mathscr{O}$.

Problem 3. Is the set of feasible local patterns of $\mathscr{O}$, denoted by $\mathcal{B}_{\mathscr{O}}$, constrained by separation property whenever MCNNs with input are considered in general? Moreover, is there any correspondence between the output patterns of MCNNs with input and CNNs without input?

In [Ban *et al.*, 2009], the authors indicate that the asymmetry of topological entropy of $\mathscr{O}$ is caused by the input terms. The study of Problem 3 in some sense gives it the most likely explanation.

In the rest of this paper, CNNs are used to mention single layer cellular neural networks without input while MCNNs is used for multilayer cellular neural networks with input for clarification.

This elucidation is organized as follows. Section 2 states the main theorems and some examples are also given here. The proofs are demonstrated in Sec. 3. Section 4 concludes the discussions of this investigation and proposes two open problems. Appendices A and B give supplementary materials so that this paper is self-contained.

2. Main Results

First fix the number of layers n. Let $\mathcal{P}_d$ denote the parameter space of (3) for $d \in \mathbb{N}$ and let $H(d) = \{h(P) : P \in \mathcal{P}_d\}$ be the collection of all possible topological entropies of output spaces $\mathscr{O}$ possessed by $\mathcal{P}_d$. The following theorem gives an affirmative answer for Problem 1.

Theorem 2.1. *Let $d_1, d_2 \in \mathbb{N}$ with $d_1 < d_2$ and let P_{d_1}, p_{d_1} (respectively P_{d_2}, p_{d_2}) be maximal and minimal templates of $\mathcal{P}_{d_1}$ (respectively $\mathcal{P}_{d_2}$), then*

(i) $H(d_1) \subsetneqq H(d_2)$;
(ii) $h(P_{d_1}) < h(P_{d_2})$;
(iii) $h(p_{d_1}) > h(p_{d_2})$.

The following example asserts an intuitive explanation for Theorem 2.1.

Example 2.1. Consider CNNs

$$\frac{dx_i}{dt} = -x_i + ay_i + a_r y_{i+1} + z, \tag{D1}$$

and

$$\frac{dx_i}{dt} = -x_i + ay_i + a_1 y_{i+1} + a_2 y_{i+2} + z. \tag{D2}$$

For clarification, denote the output patterns 1 and -1 by $+$ and $-$ respectively. It can be verified that $\mathcal{B}_1 = \{--, -+, +-\}$ is the set of feasible local patterns of output space possessed by $P_1 = (1.23, -0.40, -0.45)$ which is a maximal template of (D1), and $\mathcal{B}_2 = \{- - -, - - +, - + +, + - -, + - +, + + -, + + +\}$ is the set of feasible local patterns of output space possessed by $P_2 = (3.30, 2.13, -1.12, 1.02)$ which is a maximal template of (D2). Moreover, the topological entropies of P_1 and P_2 are $h(P_1) = g \doteq 1.61803$, which is the golden mean, and $h(P_2) = \lambda \doteq 1.83928$, which is the maximal root of $t^3 - t^2 - t - 1 = 0$, respectively. For the details of correspondence between sets of feasible local patterns and partitions, readers are referred to [Juang & Lin, 2000].

Similarly, it can be seen that $\mathcal{B}_1 = \{--, -+, +-\}$ is the set of feasible local patterns of output space possessed by $p_1 = (1.23, -0.40, -0.45)$, a minimal template of (D1), and $\mathcal{B}_2 = \{- - +, - + -, - + +, + - -, + - +\}$ is the set of feasible local patterns of output space possessed by $p_2 = (2.14, -1.86, -0.51, -0.30)$, a minimal template of (D2). The topological entropy of p_2, $h(p_2) = \tau \doteq 1.32472$, is the maximal root of $t^3 - t - 1 = 0$.

Consider the number of layers being varied and leaving the interacting radius as fixed. Denote by $\mathcal{P}_d$ the parameter space of (3) for $n \in \mathbb{N}$ and let $H(n) = \{h(P) : P \in \mathcal{P}_n\}$ be the collection of all possible topological entropies of output spaces $\mathscr{O}$ possessed by $\mathcal{P}_n$.

Theorem 2.2. *Let $n_1, n_2 \in \mathbb{N}$ with $n_1 < n_2$ and let P_{n_1}, p_{n_1} (respectively P_{n_2}, p_{n_2}) be maximal and minimal templates of $\mathcal{P}_{n_1}$ (respectively $\mathcal{P}_{n_2}$), then*

(i) $\{h(P) : P \in \mathcal{P}_{n_1}\} \subsetneqq \{h(P) : P \in \mathcal{P}_{n_2}\}$;
(ii) $h(P_{n_1}) < h(P_{n_2})$;
(iii) $h(p_{n_1}) > h(p_{n_2})$.

Example 2.2. Consider the following two systems

$$\frac{dx_i}{dt} = -x_i + ay_i + a_r y_{i+1} + z, \tag{N1}$$

and

$$\begin{aligned} \frac{dx_i^{(2)}}{dt} &= -x_i^{(2)} + a^{(2)} y_i^{(2)} + a_1^{(2)} y_{i+1}^{(2)} \\ &\quad + b y_i^{(1)} + b_1 y_{i+1}^{(1)} + z^{(2)}, \end{aligned} \tag{N2}$$

$$\frac{dx_i^{(1)}}{dt} = -x_i^{(1)} + a^{(1)} y_i^{(1)} + a_1^{(1)} y_{i+1}^{(1)} + z^{(1)}.$$

As discussed in Example 2.1, $\mathcal{B}_1 = \{--, -+, +-\}$ is both the set of feasible local patterns of minimal and maximal templates of (N1), denoted by P_1, with topological entropy $h(P_1) = g$.

To find a maximal template of (N2) out, consider $P_2 = P(\mathbf{A}, \mathbf{B}, \mathbf{z})$, where

$$(A^{(1)}, z^{(1)}) = (0.64, -1.20, 0.64),$$

$$(A^{(2)}, B, z^{(2)}) = (5.17, -2.56, 1.44, 1.21, -1.52).$$

It is easily seen that $\mathcal{B}_2 = \{--, -+, +-, ++\}$ with $\mathcal{B}_{2;1} = \{-+, +-\}$ and $\mathcal{B}_{2;2} = \{y_1y_2 \diamond u_1u_2 : yi, u_i \in \{-,+\}\} \backslash \{++\diamond--, ++\diamond-+\}$, herein $y_1y_2 \diamond u_1u_2$ means output pattern y_1y_2 coupled with input pattern u_1u_2. Although the feasible local output patterns consist of all possible choices, the topological entropy of maximal template of (N2) is $h(P_2) \doteq 1.83928$ which is the maximal root of $t^3 - t^2 - t - 1 = 0$. An intuitive observation for the computation of topological entropy of P_2 is that the permanent local pattern generated by P_2 is $\{---, --+, -++, +--, +-+, ++-, +++\}$. For the theoretical algorithm of topological entropy and correspondence between sets of feasible local patterns and partitions of parameter space, the reader may refer to [Ban *et al.*, 2009].

Let p_2 be obtained from P_2 by replacing $(A^{(2)}, B, z^{(2)})$ with $(A'^{(2)}, B', z'^{(2)}) = (1.02, -0.36, 1.64, 1.06, 0.18)$, then $\mathcal{B}'_2 = \{--, -+, +-, ++\}$ with

$$\mathcal{B}'_{2;2} = \left\{\begin{matrix} --\diamond--, --\diamond-+, -+\diamond--, \\ -+\diamond-+, -+\diamond+-, +-\diamond++, \\ +-\diamond+-, ++\diamond++, ++\diamond+- \end{matrix}\right\}.$$

It can be easily verified that the topological entropy of minimal template is $h(p_2) \doteq 1.32472$ which is the maximal root of $t^3 - t - 1 = 0$ since the permanent local pattern generated by p_2 is $\{--+, -+-, -++, +--, +-+\}$.

In [Hsu *et al.*, 2000], the authors demonstrate the necessary and sufficient condition whether a given set of local patterns $\mathcal{S}$ is associated with a CNN, i.e. $\mathcal{S} = \mathcal{B}(P)$ for some template $P = P(A, z)$. More precisely, if such a template P exists, then $\mathcal{S}$ must satisfy separation property (see [Hsu *et al.*, 2000] for more details). Furthermore, Ban *et al.* [2009] indicated that the symmetry of a family of topological entropies of MCNNs does not hold as that of CNNs and the asymmetry is caused by the effect of input terms. The following theorem demonstrates that the set of feasible local patterns of output space induced by a MCNN is actually the one possessed by a CNN. This supports the viewpoint of [Ban *et al.*, 2009].

Theorem 2.3. *Given $P = P(\mathbf{A}, \mathbf{B}, \mathbf{z})$ a template of (3), then the set of feasible local patterns of output space coincides with the set of feasible local patterns of output space which is obtained from a CNN. That is, there exists $A' \in \mathbb{R}^{2d+1}$, $z' \in \mathbb{R}$ such that $\mathcal{B}(P) = \mathcal{B}(A', z')$. Moreover, A' has only one element different from $A^{(n)}$, i.e. $a'_i = a_i^{(n)}$ for $|i| \le d, i \ne 0$.*

Example 2.3. Consider

$$\frac{dx_i^{(2)}}{dt} = -x_i^{(2)} + 1.02y_i^{(2)} - 0.36y_{i+1}^{(2)} + 1.64y_i^{(1)} + 1.06y_{i+1}^{(1)} + 0.18, \tag{P2}$$

$$\frac{dx_i^{(1)}}{dt} = -x_i^{(1)} + 0.64y_i^{(1)} - 1.2y_{i+1}^{(1)} + 0.64.$$

The discussion in Example 2.2 indicates that the feasible local pattern of (P2) is $\mathcal{B} = \{--, -+, +-, ++\}$ which coincides with the following system:

$$\frac{dx_i}{dt} = -x_i + 2.56y_i - 0.36y_{i+1} + 1.06. \tag{P1}$$

3. Proof of Theorems

Once a specified template $P = P(\mathbf{A}, \mathbf{B}, \mathbf{z})$ is given, a set of feasible local patterns, denoted by $\mathcal{B} = \mathcal{B}(P)$, is determined such that if $\mathbf{x}$ is a mosaic solution of (3) with its corresponding pattern $\mathbf{y}$, then $\mathbf{y} \in \mathcal{B}$ locally.

Denote the parameter space $\mathcal{P}$ of (3) by $\mathcal{P}_{n,d}$ for clarification. Two templates $P_1 = P(\mathbf{A_1}, \mathbf{B_1}, \mathbf{z_1}), P_2 = P(\mathbf{A_2}, \mathbf{B_2}, \mathbf{z_2})$ are said to be $P_1 \sim P_2$ if and only if $\mathcal{B}(P_1) = \mathcal{B}(P_2)$. Then $\sim$ is an equivalence relation [Ban *et al.*, 2009]. For simplicity, denote $\mathcal{P}_{n,d;\sim}$ by $\mathcal{P}_{n,d}$. For each $P \in \mathcal{P}_{n,d}$, there exists a unique P-graph, say G, up to the labeling of edges. The study of output space exhibited by $\mathcal{B}(P)$ is equivalent to the study of the symbolic dynamical system induced by G. We refer readers to [Ban *et al.*, 2009; Lind & Marcus, 1995] for details.

3.1. *Proof of Theorem 2.1*

First, elucidate the effect of the parameter d in (3) and leave the number of layers n as fixed. For simplicity, the proof for $n = 1$ and $B = 0$ is given. The general case can be processed via the same method, thus is skipped.

Rewrite (3) in the following form,

$$x_i = z + \sum_{|k| \le d} a_k f(x_{i+k}), \tag{5}$$

where $i \in \mathbb{Z}$ and $d \in \mathbb{N}$. Denote by

$$\mathcal{P}_d = \{(A_d, z) : A_d \in \mathbb{R}^{2d+1}, z \in \mathbb{R}\}, \tag{6}$$

the parameter space with respect to $d \in \mathbb{N}$, where $A_d = (a_{-d}, \ldots, a_d)$. For a given template $P = P(A_d, z)$, the set of feasible local patterns $\mathcal{B} = \mathcal{B}(P)$ consists of those patterns with size $n_d \times 1$, where $n_d = 2d + 1$.

Let $X = \{1, -1\}$ and let $\sigma : X^{\mathbb{Z}} \to X^{\mathbb{Z}}$ be the shift map, the projection map $\pi_r : X^{\mathbb{Z}} \to X^{\mathbb{Z}_{r\times 1}}$ is defined by $\pi_r(x) = x_0 x_1 \cdots x_{r-1}$, where $x = (x_i)_{i\in\mathbb{Z}}$. For simplicity, a projection map from $X^{\mathbb{Z}_{m\times 1}}$ to $X^{\mathbb{Z}_{r\times 1}}$ is also denoted by π_r and is defined by $\pi_r(y) = y_1 y_2 \cdots y_r$, where $y = (y_i)_{i=1}^m$ and $m \geq r$. The proof of Theorem 2.1 is divided into two parts. Before giving the proof of the first statement, some definitions should be stated first.

Definition 3.1. For $d \in \mathbb{N}$, $\mathcal{U} \subseteq X^{\mathbb{Z}_{n_d\times 1}}$, the shift space induced by $\mathcal{U}$ is defined by

$$\mathbf{X}_{\mathcal{U}} = \{x \in X^{\mathbb{Z}} : \pi_{n_d}(\sigma^i(x)) \in \mathcal{U} \text{ for all } i \in \mathbb{Z}\}. \tag{7}$$

Furthermore, denoting $\Sigma_m(\mathbf{X}_{\mathcal{U}})$ the set of $(m \times 1)$-blocks in $\mathbf{X}_{\mathcal{U}}$ by

$$\Sigma_m(\mathbf{X}_{\mathcal{U}}) = \{v \in X^{\mathbb{Z}_{m\times 1}} : \text{ there exists } x \in \mathbf{X}_{\mathcal{U}} \text{ such that } \pi_m(x) = v\}. \tag{8}$$

Definition 3.2. For $d_1, d_2 \in \mathbb{N}$ such that $d_1 < d_2$, let $\mathcal{U}_{d_i} \subseteq X^{\mathbb{Z}_{n_{d_i}\times 1}}$ for $i = 1, 2$.

(i) Define $\phi : \mathcal{U}_{d_2} \to \mathcal{U}_{d_1}$ by obtaining $\phi(u)$ from $u \in \mathcal{U}_{d_2}$ by deleting the rightmost and the leftmost $(d_2 - d_1)$ elements. That is, $\phi(u)$ deletes the boundary of u with size $(d_2 - d_1)$.

(ii) For $m \in \mathbb{N}$, defining $\phi_m : \Sigma_m(\mathbf{X}_{\mathcal{U}_{d_2}}) \to \Sigma_{m-2(d_2-d_1)}(\mathbf{X}_{\mathcal{U}_{d_1}})$ by obtaining $\phi_m(v)$ from $v \in \Sigma_m(\mathbf{X}_{\mathcal{U}_{d_2}})$ by deleting the boundary of v with size $(d_2 - d_1)$.

For simplicity, denote ϕ_m by ϕ. In other words, ϕ presents the map deleting the boundary with size $d_2 - d_1$ once d_1, d_2 are given. A lemma then follows.

Lemma 3.3. *Consider $d_1 < d_2$, $\mathcal{U}_{d_1} \subseteq X^{\mathbb{Z}_{n_{d_1}\times 1}}$. Let $\mathcal{U}_{d_2} = \phi^{-1}(\mathcal{U}_{d_1}) \subseteq X^{\mathbb{Z}_{n_{d_2}\times 1}}$, then $h(\mathbf{X}_{\mathcal{U}_{d_1}}) = h(\mathbf{X}_{\mathcal{U}_{d_2}})$.*

Proof. It is easy to see that G_{d_2} forces G_{d_1}, where G_{d_i} is the graph representation of $\mathcal{U}_{d_i}$ for $i = 1, 2$, hence $h(\mathbf{X}_{\mathcal{U}_{d_1}}) \leq h(\mathbf{X}_{\mathcal{U}_{d_2}})$ by Theorem A.8. It remains to show that $h(\mathbf{X}_{\mathcal{U}_{d_1}}) \geq h(\mathbf{X}_{\mathcal{U}_{d_2}})$.

First claim that $\phi^{-1}(\Sigma_{n_{d_1}+k}(\mathbf{X}_{\mathcal{U}_{d_1}})) = \Sigma_{n_{d_2}+k}(\mathbf{X}_{\mathcal{U}_{d_2}})$ for $k \in \mathbb{N}$. For $\mu \in \Sigma_{n_{d_2}+k}(\mathbf{X}_{\mathcal{U}_{d_2}})$, $\pi_{n_{d_2}}(\sigma^r\mu) \in \mathcal{U}_{d_2}$ for $0 \leq r \leq k$. Hence, $\phi(\pi_{n_{d_2}}(\sigma^r\mu)) \in \mathcal{U}_{d_1}$ for $0 \leq r \leq k$. Moreover, it is easy to check that $\phi \circ \pi_{n_{d_2}} = \pi_{n_{d_1}} \circ \phi$ and $\phi \circ \sigma^r = \sigma^r \circ \phi$. This implies

$$\pi_{n_{d_1}}(\sigma^r(\phi(\mu))) \in \mathcal{U}_{d_1}, \quad \text{for } 1 \leq r \leq k. \tag{9}$$

Thus, $\phi(\mu) \in \Sigma_{n_{d_1}+k}(\mathbf{X}_{\mathcal{U}_{d_1}})$.

Conversely, if $\mu \in \phi^{-1}(\Sigma_{n_{d_1}+k}(\mathbf{X}_{\mathcal{U}_{d_1}}))$, then $\phi(\mu) \in \Sigma_{n_{d_1}+k}(\mathbf{X}_{\mathcal{U}_{d_1}})$. Moreover, $\pi_{n_{d_1}}(\sigma^r(\phi(\mu))) \in \mathcal{U}_{n_{d_1}}$ for $1 \leq r \leq k$. This deduces $\pi_{n_{d_2}}(\sigma^r(\mu)) \in \mathcal{U}_{n_{d_2}}$ for $1 \leq r \leq k$. Thus $\mu \in \Sigma_{n_{d_2}+k}(\mathbf{X}_{\mathcal{U}_{d_2}})$. This completes the proof of the claim.

Since $\phi^{-1}(\Sigma_{n_{d_1}+k}(\mathbf{X}_{\mathcal{U}_{d_1}})) = \Sigma_{n_{d_2}+k}(\mathbf{X}_{\mathcal{U}_{d_2}})$ for $k \in \mathbb{N}$,

$$\Gamma_{n_{d_2}+k}(\mathbf{X}_{\mathcal{U}_{d_2}}) = 2^{n_{d_2}-n_{d_1}}\Gamma_{n_{d_1}+k}(\mathbf{X}_{\mathcal{U}_{d_1}}), \tag{10}$$

where $\Gamma_r(\mathbf{X})$ is the cardinality of $\Sigma_r(\mathbf{X})$. It can be easily checked that $h(\mathbf{X}_{\mathcal{U}_{d_1}}) \geq h(\mathbf{X}_{\mathcal{U}_{d_2}})$, this completes the proof. ■

Next, demonstrate the strict monotonicity of entropy on d. The following lemma comes immediately from [Hsu *et al.*, 2000], thus the proof is omitted.

Lemma 3.4. *Given $d \in \mathbb{N}$, consider $\mathcal{U}_d \subseteq X^{\mathbb{Z}_{n_d\times 1}}$ with cardinality $|\mathcal{U}_d| = 2^{n_d-1}$, then there exists $(A_d, z) \in \mathcal{P}_d$ such that $\mathcal{B}(A_d, z) = \mathcal{U}_d$.*

Lemma 3.5. *Consider $d_1 < d_2$, $\mathcal{U}_{d_i} \subseteq \mathbf{X}^{\mathbb{Z}_{n_{d_i}\times 1}}$ with $\mathcal{U}_{d_i} = X^{\mathbb{Z}_{n_{d_i}\times 1}} \setminus \{1\}^{\mathbb{Z}_{n_{d_i}\times 1}}$ for $i = 1, 2$, then $h(\mathbf{X}_{\mathcal{U}_{d_1}}) < h(\mathbf{X}_{\mathcal{U}_{d_2}})$.*

Proof. It can be checked without difficulty that G_{d_2} forces G_{d_1}, this implies $h(\mathbf{X}_{\mathcal{U}_{d_1}}) \leq h(\mathbf{X}_{\mathcal{U}_{d_2}})$ by Theorem A.8.

Moreover, G_{d_2} is irreducible since $\mathcal{U}_{d_2} = X^{\mathbb{Z}_{n_{d_2}\times 1}} \setminus \{1\}^{\mathbb{Z}_{n_{d_2}\times 1}}$. The fact that card $V(G_{d_2}) \gneqq$ card $V(G_{d_1})$ demonstrates that G_{d_1} is isomorphic to a proper subgraph of G_{d_2}. In other words, $\mathbf{X}_{\mathcal{U}_{d_1}}$ is a proper subshift of an irreducible shift space $\mathbf{X}_{\mathcal{U}_{d_2}}$. Thus, $h(\mathbf{X}_{\mathcal{U}_{d_1}}) < h(\mathbf{X}_{\mathcal{U}_{d_2}})$ by Theorem A.10 and the proof is completed. ■

Proof of Theorem 2.1. Consider $\mathcal{B}(A_{d_1}, z_1) \subseteq X^{\mathbb{Z}_{n_{d_1}\times 1}}$ the set of feasible local patterns induced by some $(A_{d_1}, z_1) \in \mathcal{P}_{d_1}$, it is easily seen that there exists $(A_{d_2}, z_2) \in \mathcal{P}_{d_2}$ such that $\mathcal{B}(A_{d_2}, z_2) = \phi^{-1}(\mathcal{B}(A_{d_1}, z_1)) \subseteq X^{\mathbb{Z}_{n_{d_2}\times 1}}$. By Lemma 3.3,

$h(\mathbf{X}_{\mathcal{B}(A_{d_1},z_1)}) = h(\mathbf{X}_{\mathcal{B}(A_{d_2},z_2)})$. Thus $\{h(P) : P \in \mathcal{P}_{d_1}\} \subseteq \{h(P) : P \in \mathcal{P}_{d_2}\}$.

Lemma 3.4 shows that there exists $P_i = P(A_{d_i}, z_i) \in \mathcal{P}_{d_i}$ such that the cardinality of $\mathcal{B}(P_i)$ is $2^{n_{d_i}-1}$, for $i = 1, 2$. It can be easily checked that $h(\mathbf{X}_{\mathcal{B}(P_i)}) < \log 2$, thus is a maximal template of $\mathcal{P}_{d_i}$, for $i = 1, 2$. Lemma 3.5 demonstrates that $h(\mathbf{X}_{\mathcal{B}(A_{d_1},z_1)}) < h(\mathbf{X}_{\mathcal{B}(A_{d_2},z_2)})$. This completes the proof of (ii).

The proof of (iii) can be done analogously, thus is omitted. ■

3.2. *Proof of Theorem 2.2*

Denote by $\mathcal{P}_n$ the parameter space of (3) with $d \in \mathbb{N}$ being fixed. For each $P = P(\mathbf{A}, \mathbf{B}, \mathbf{z}) \in \mathcal{P}_n$, the output space $\mathbf{X}_\mathcal{B}$ is a sofic shift rather than a subshift of finite type, where $\mathcal{B} = \mathcal{B}(\mathbf{A}, \mathbf{B}, \mathbf{z})$ is the set of feasible local patterns induced by P. Moreover, the topological entropy $h(\mathbf{X}_\mathcal{B})$ can be exactly formulated [Ban *et al.*, 2009].

Proof of Theorem 2.2. For simplicity, we consider the case where $n_1 = 1, n_2 = 2$ and $d = 1$ is proved. The general case can be done via analogous method, thus is omitted.

Given a set of feasible local patterns $\mathcal{B}(A, B, z)$ produced by (3) with $n = 1$, let $(A^{(2)}, B^{(2)}, z^{(2)}) = (A, B, z)$ be the template of the second layer. The template $(A^{(1)}, B^{(1)}, z^{(1)})$ is chosen so that $\mathcal{B}(A^{(1)}, B^{(1)}, z^{(1)})$ consists of all possible patterns and the set of input patterns is assigned by $\mathcal{U} = X^{\mathbb{Z}_{3\times 1}}$. It can be easily seen that $h(\mathbf{X}_{\mathcal{B}(\mathbf{A},\mathbf{B},\mathbf{z})}) = h(\mathbf{X}_{\mathcal{B}(A,B,z)})$, where $\mathbf{E} = (E^{(1)}, E^{(2)})$ for $E \in \{A, B, z\}$. This shows that $H(n_1) \subseteq H(n_2)$.

To elucidate the inequality, it suffices to show that (iii) is true. (ii) can be done in an analogous method, thus is omitted. Consider $p_1 = P(A, B, z)$ is a minimal template for $n = 1$. Hence, $h(P(A', B', z')) \geq h(p_1)$ for all $P(A', B', z')$ so that $h(P(A', B', z')) \neq 0$. Decompose $\mathbf{X}_\mathcal{B}$ into irreducible components $\mathbf{X}_{\mathcal{B}_1}, \mathbf{X}_{\mathcal{B}_2}, \ldots, \mathbf{X}_{\mathcal{B}_k}$, where $\mathcal{B} = \mathcal{B}(p_1)$. Without loss of generality, assume that $h(\mathbf{X}_{\mathcal{B}_1}) = h(p_1)$. Similarly as above, the template for second layer is set up as $(A^{(2)}, B^{(2)}, z^{(2)}) = (A, B, z)$. It remains to decide the parameters for the first layer.

Denote by τ_l the *back projection* defined by

$$\tau_l(x) = (x_{n-l+1}, x_{n-l+2}, \ldots, x_n) \in \mathbb{R}^l, \quad \text{where } x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n. \tag{11}$$

Let $\mathcal{B}' = \tau_3(\mathcal{B})$, without loss of generality, assume that $\mathcal{B}'$ is exhibited by $\tilde{A} = (\tilde{a}_l^{(1)}, \tilde{a}^{(1)}, \tilde{a}_r^{(1)})$ and $\tilde{z}^{(1)}$ for some template $(\tilde{A}, \tilde{z}^{(1)})$. Consider $B^{(1)} = (b_l^{(1)}, b^{(1)}, b_r^{(1)})$ satisfying the following conditions:

(i) $b^{(1)} > b_l^{(1)} > b_r^{(1)} > 0$ and $b^{(1)} > b_l^{(1)} + b_r^{(1)}$;
(ii) $b_l^{(1)} + b^{(1)} + b_r^{(1)} < \min\{|\tilde{a}_l^{(1)}|, |\tilde{a}^{(1)}|, |\tilde{a}_r^{(1)}|\}$.

Let $y^+ = (y_1^+, y_2^+), y^- = (y_1^-, y_2^-) \in X^2$ satisfy

$$(y_1^+, 1, y_2^+) \in \mathcal{B}',\ y^+ \cdot \alpha < y \cdot \alpha \text{ for all } y = (y_1, y_2) \in X^2 \text{ so that } (y_1, 1, y_2) \in \mathcal{B}',$$

and

$$(y_1^-, -1, y_2^-) \in \mathcal{B}',\ y^- \cdot \alpha > y \cdot \alpha \text{ for all } y = (y_1, y_2) \in X^2 \text{ so that } (y_1, -1, y_2) \in \mathcal{B}',$$

respectively, where $\alpha = (\tilde{a}_l^{(1)}, \tilde{a}_r^{(1)})$. Denote $\ell = b^{(1)} - b_l^{(1)}$, let

$$a' = 1 + \ell - \frac{1}{2}(y^+ - y^-) \cdot \alpha,$$

$$z' = -\frac{1}{2}(y^+ + y^-) \cdot \alpha,$$

$$\text{and} \quad A^{(1)} = (\tilde{a}_l^{(1)}, a', \tilde{a}_r^{(1)}).$$

It could be foreseen from the above data that

$$\pi_3(\mathcal{B}(A^{(1)}, B^{(1)}, z')) = \mathcal{B}' \text{ and } \tau_3(\mathcal{B}(A^{(1)}, B^{(1)}, z')) = X^{\mathbb{Z}_{3\times 1}}.$$

If, furthermore, the set of input patterns is assigned by $\mathcal{U} = \{(-1, 1, -1), (1, -1, 1)\}$, then it is easy to check that, in such an event, $h(\mathbf{X}_{\mathcal{B}(\mathbf{A},\mathbf{B},\mathbf{z};\mathcal{U})}) = h(\mathbf{X}_\mathcal{B})$ and the decomposition of $\mathcal{B}(\mathbf{A}, \mathbf{B}, \mathbf{z}; \mathcal{U})$, denoted by $\hat{\mathcal{B}}_1, \hat{\mathcal{B}}_2, \ldots, \hat{\mathcal{B}}_k$, are also irreducible. More precisely, $\mathcal{B}(A^{(1)}, B^{(1)}, z'; \mathcal{U}) = \{(y, u) \in X^3 \times X^3 : y \in \mathcal{B}, u \in \mathcal{U}\}$ and $h(\mathbf{X}_{\hat{\mathcal{B}}_1}) = h(\mathbf{X}_{\mathcal{B}(\mathbf{A},\mathbf{B},\mathbf{z};\mathcal{U})}) > 0$.

Consider $A^{(1)'} = (\tilde{a}_l^{(1)}, \hat{a}, \tilde{a}_r^{(1)})$, where $\hat{a} = 1 + 1/4(3\ell - b_r^{(1)}) - 1/2(y^+ - y^-) \cdot \alpha$. Let $\hat{z} = 1/4(\ell + b_r^{(1)}) - 1/2(y^+ + y^-) \cdot \alpha$, then

$$\mathcal{B}(A^{(1)'}, B^{(1)}, \hat{z}) = \mathcal{B}(A^{(1)}, B^{(1)}, z') - \{((y_1^-, -1, y_2^-; -1, 1, -1))\}.$$

Theorem A.10 shows that $h(\mathbf{X}_{\mathcal{B}(\mathbf{A}',\mathbf{B},\mathbf{z}';\mathcal{U})}) < h(\mathbf{X}_{\mathcal{B}(\mathbf{A},\mathbf{B},\mathbf{z};\mathcal{U})})$ since $\mathbf{X}_{\mathcal{B}(\mathbf{A}',\mathbf{B},\mathbf{z}';\mathcal{U})}$ is a proper subshift of $\mathbf{X}_{\mathcal{B}(\mathbf{A},\mathbf{B},\mathbf{z};\mathcal{U})}$, where $\mathbf{A}' = (A^{(1)'}, A^{(2)}), \mathbf{z}' = (\hat{z}, z^{(2)})$. The discussion of $h(\mathbf{X}_{\mathcal{B}(\mathbf{A}',\mathbf{B},\mathbf{z}';\mathcal{U})}) > 0$ is essentially the same as above.

This completes the proof. ■

3.3. *Proof of Theorem 2.3*

First consider the effect of the input term for $n = 1$. For $n = 1$ and $B = 0$, when a template (A, z) is given, the set of feasible local patterns with "1" in the center is defined by

$$\mathcal{B}(+, A, z) = \{y \in X^{2d} : \alpha \cdot y + a + z - 1 > 0\}, \quad (12)$$

where $\alpha = (a_{-d}, \ldots, a_{-1}, a_1, \ldots, a_d)$ represents the parameters in A without the center. Moreover, the basic set of admissible output patterns with "-1" in the center is defined by

$$\mathcal{B}(-, A, z) = \{y \in X^{2d} : \alpha \cdot y - a + z + 1 < 0\}. \quad (13)$$

The set of all feasible output patterns exhibited by (A, z) is then defined by $\mathcal{B}(A, z) = (\mathcal{B}(+, A, z), \mathcal{B}(-, A, z))$.

Furthermore, considering $B \neq 0$, when (A, B, z) is given, let $\tilde{\mathcal{B}}(+, A, B, z)$ and $\tilde{\mathcal{B}}(-, A, B, z)$ be defined in (B.16) and (B.17), $\mathcal{B}(*, A, B, z) = \pi_n(\tilde{\mathcal{B}}(*, A, B, z))$ for $* \in \{+, -\}$. The set of all feasible output patterns induced by (A, B, z) is defined by $\mathcal{B}(A, B, z) = (\mathcal{B}(+, A, B, z), \mathcal{B}(-, A, B, z))$.

Theorem 3.6. *Given templates A, B and threshold z, there exists threshold z' such that*

$$\mathcal{B}(A, B, z) = \mathcal{B}(A, z'). \quad (14)$$

Proof. Let $u_+, u_- \in X^{2d+1}$ satisfy

$$\beta \cdot u_+ > \beta \cdot u, \quad \beta \cdot u_- < \beta \cdot u, \quad \text{for all } u \in X^{2d+1}, \quad (15)$$

respectively, where $\beta = (b_{-d}, \ldots, b_{-1}, b, b_1, \ldots, b_d)$ represents the parameters in B. It can be easily checked that $u_+ + u_- = 0$. Denote by $k = \beta \cdot u_+ = -\beta \cdot u_-$ and $z' = z + k$.

If $y \in \mathcal{B}(+, A, B, z)$, then $(y, u_+) \in \tilde{\mathcal{B}}(+, A, B, z)$. Moreover,

$$a - 1 + z > -(\alpha \cdot y + \beta \cdot u_+), \quad \text{for all } y \in \mathcal{B}(+, A, B, z), \quad (16)$$

and

$$a - 1 + z < -(\alpha \cdot y + \beta \cdot u_+), \quad \text{for all } y \notin \mathcal{B}(+, A, B, z). \quad (17)$$

In other words,

$$a - 1 + z' > -\alpha \cdot y \Leftrightarrow y \in \mathcal{B}(+, A, z'). \quad (18)$$

Thus,

$$\mathcal{B}(+, A, B, z) = \mathcal{B}(+, A, z'). \quad (19)$$

Similarly, $y \in \tilde{\mathcal{B}}(-, A, B, z)$ if and only if

$$a - 1 - z > \alpha \cdot y + \beta \cdot u_-, \quad \text{for all } y \in \mathcal{B}(-, A, B, z), \quad (20)$$

and

$$a - 1 - z < -\alpha \cdot y + \beta \cdot u_-, \quad \text{for all } y \notin \mathcal{B}(-, A, B, z). \quad (21)$$

This is equivalent to

$$a - 1 - z' > \alpha \cdot y \Leftrightarrow y \in \mathcal{B}(-, A, z'), \quad (22)$$

and deducing that

$$\mathcal{B}(-, A, B, z) = \mathcal{B}(-, A, z'). \quad (23)$$

This completes the proof. ■

The proof of Theorem 2.3 can be done via the same idea.

Proof of Theorem 2.3. For simplicity, the case for $n = 2$ is proved. The general case for $n \in \mathbb{N}$ can be done by mathematical induction.

Theorem 3.6 shows the existence of $\tilde{z}$ such that $\mathcal{B}(A^{(1)}, \tilde{z}) = \mathcal{B}(A^{(1)}, B^{(1)}, z^{(1)})$. Therefore, for $y \in \mathcal{B}(A, B, z)$, there exists $u \in \mathcal{B}(A^{(1)}, \tilde{z})$ such that $(y, u) \in \tilde{\mathcal{B}}(A^{(2)}, B^{(2)}, z^{(2)})$. Without loss of generality, assume that $|\mathcal{B}(A^{(1)}, \tilde{z})| = k$ and

$$\beta^{(2)} \cdot u_i > \beta^{(2)} \cdot u_j, \quad \text{for } 1 \leq i < j \leq k. \quad (24)$$

Then $y \in \mathcal{B}(*, A, B, z)$ if and only if $(y, u_i) \in \tilde{\mathcal{B}}(*, A^{(2)}, B^{(2)}, z^{(2)})$ for some $1 \leq i \leq k$, $* \in \{+, -\}$. Let $\{y_i^+\}_{i=1}^{2^{2d}}$, $\{y_j^-\}_{j=1}^{2^{2d}}$ be the set of all possible output patterns with 1 and -1 in the center, respectively, and

$$\alpha^{(2)} \cdot y_i^+ > \alpha^{(2)} \cdot y_j^+, \quad \text{for } 1 \leq i < j \leq 2^{2d}, \quad (25)$$

$$\alpha^{(2)} \cdot y_k^- < \alpha^{(2)} \cdot y_\ell^-, \quad \text{for } 1 \leq k < \ell \leq 2^{2d}. \quad (26)$$

Hence, $y_i^+ \in \mathcal{B}(+, A, B, z)$ for $1 \leq i \leq \ell$, and $y_j^- \in \mathcal{B}(-, A, B, z)$ for $1 \leq j \leq m$, where $\ell = |\mathcal{B}(+, A, B, z)|$, $m = |\mathcal{B}(-, A, B, z)|$.

Moreover, consider $u_+, u_- \in X^{2d+1}$ satisfying

$$\beta^{(2)} \cdot u_+ > \beta^{(2)} \cdot u, \; \beta^{(2)} \cdot u_- < \beta^{(2)} \cdot u, \quad \text{for all } u \in X^{2d+1}, \quad (27)$$

respectively. Again, $u_+ + u_- = 0$. Let

$$a' = \frac{1}{4}\alpha^{(2)} \cdot (y_m^- + y_{m+1}^- - y_\ell^+ - y_{\ell+1}^+) + \beta^{(2)} \cdot u_- + 1, \quad (28)$$

$$z' = -\frac{1}{4}\alpha^{(2)} \cdot (y_m^- + y_{m+1}^- + y_\ell^+ + y_{\ell+1}^+), \quad (29)$$

then

$$a' - 1 + z' > -(\alpha^{(2)} \cdot y + \beta^{(2)} \cdot u_+), \quad \text{for all } y \in \mathcal{B}(+, A, B, z), \quad (30)$$

and

$$a' - 1 + z' < -(\alpha^{(2)} \cdot y + \beta^{(2)} \cdot u_+), \quad \text{for all } y \notin \mathcal{B}(+, A, B, z). \quad (31)$$

Similarly,

$$a' - 1 - z' > \alpha^{(2)} \cdot y + \beta^{(2)} \cdot u_-, \quad \text{for all } y \in \mathcal{B}(-, A, B, z), \quad (32)$$

and

$$a' - 1 - z' < \alpha^{(2)} \cdot y + \beta^{(2)} \cdot u_-, \quad \text{for all } y \notin \mathcal{B}(-, A, B, z). \quad (33)$$

Let $A^{(2)'}$ be the template obtained from $A^{(2)}$ by replacing the center element by a', then $\mathcal{B}(*, A, B, z) = \mathcal{B}(*, A^{(2)'}, z')$ for $* \in \{+, -\}$. This completes the proof. ■

4. Discussion and Conclusion

This work demonstrates the following facts:

(a) The entropy set for multilayer cellular neural networks is strictly nested with respect to interacting radius (respectively number of layers) when the number of layers (respectively interacting radius) is fixed. More precisely, denote by $H(d, n)$ the collection of those topological entropies derived from (3), then $H(d, n)$ is a strictly nested sequence of sets whenever either d or n is fixed.

(b) The maximal and minimal templates of $\mathcal{P}_{d,n}$ satisfy a forcing relation. If $n \in \mathbb{N}$ is fixed, let P_d and p_d be maximal and minimal templates of $\mathcal{P}_{d,n}$, respectively. Then $h(P_{d_1}) \lneqq h(P_{d_2})$ (respectively $h(p_{d_1}) \gneqq h(p_{d_2})$) for $d_1 < d_2$. The same result holds for n as $d \in \mathbb{N}$ is fixed. We emphasize that it gets much more complex when n varies since this relates to the convolution of sofic shifts.

(c) The set of feasible local patterns exhibited by a multilayer cellular neural network can be realized via a single layer cellular neural network without input. Ban *et al.* [2009] indicated that these two systems do not possess the same set of topological entropies and such a difference is caused by input terms. Our result gives it a positive support.

Due to the constraint of separation property, the set of feasible local patterns cannot be arbitrary. The strict monotonicity of topological entropy on interacting radius and the number of layers inspire the elucidation of the following two open problems.

Problem 4. Fix $n \in \mathbb{N}$ in (3) and $\epsilon > 0$. For any $0 \le \kappa \le \log 2$, does there exist $d \in \mathbb{N}$ and $P_{\epsilon,\kappa} \in \mathcal{P}_d$ such that $|h(P_{\epsilon,\kappa}) - \kappa| < \epsilon$?

Problem 5. Fix $d \in \mathbb{N}$ in (3) and $\epsilon > 0$. For any $0 \le \kappa \le \log 2$, does there exist $n \in \mathbb{N}$ and $P_{\epsilon,\kappa} \in \mathcal{P}_n$ such that $|h(P_{\epsilon,\kappa}) - \kappa| < \epsilon$?

Our conjecture is: The answers for Problems 4 and 5 are both positive, i.e. the topological entropy set of (3) is dense in the interval $[0, \log 2]$.

References

Alsedà, L., Llibre, J. & Misiurewicz, M. [2000] *Combinatorial Dynamics and Entropy in Dimension One*, Advanced Series in Nonlinear Dynamics, Vol. 5 (World Scientific, Singapore).

Ban, J.-C., Chang, C.-H., Lin, S.-S. & Lin, Y.-H. [2009] "Spatial complicity in multi-layer cellular neural networks," *J. Diff. Eqs.* **246**, 552–580.

Bowen, R. & Frank, J. [1976] "The periodic points of maps of the disk and the interval," *Topology* **15**, 337–342.

Chow, S.-N., Mallet-Paret, J. & Van Vleck, E.S. [1996] "Dynamics of lattice differential equations," *Int. J. Bifurcation and Chaos* **6**, 1605–1621.

Chua, L. O. & Yang, L. [1988a] "Cellular neural networks: Theory," *IEEE Trans. Circuits Syst.* **35**, 1257–1272.

Chua, L. O. & Yang, L. [1988b] "Cellular neural networks: Applications," *IEEE Trans. Circuits Syst.* **35**, 1273–1290.

Chua, L. O. [1998] *CNN: A Paradigm for Complexity*, World Scientific Series on Nonlinear Science, Series A, Vol. 31 (World Scientific, Singapore).

Crounse, K. R., Roska, T. & Chua, L. O. [1993] "Image halftoning with cellular neural networks," *IEEE Trans. Circuits Syst.* **40**, 267–283.

Feigenbaum, M. J. [1979] "The universal metric properties of nonlinear transformations," *J. Stat. Phys.* **21**, 669–706.

Hsu, C.-H., Juang, J., Lin, S.-S. & Lin, W.-W. [2000] "Cellular neural networks: Local patterns for general template," *Int. J. Bifurcation and Chaos* **10**, 1645–1659.

Juang, J. & Lin, S.-S. [2000] "Cellular neural networks: Mosaic pattern and spatial chaos," *SIAM J. Appl. Math.* **60**, 891–915.

Lee, C.-C. & Pineda de Gyvez, J. [1996] "Color image processing in a cellular neural network environment," *IEEE Trans. Neural Networks* **7**, 1086–1098.

Lind, D. & Marcus, B. [1995] *An Introduction to Symbolic Dynamics and Coding* (Cambridge University Press, Cambridge).

Mallet-Paret, J. & Chow, S.-N. [1995] "Pattern and spatial chaos in lattice dynamical systems — Part II," *IEEE Trans. Circuits Syst.* **42**, 752–756.

May, R. B. [1976] "Simple mathematical models with very complicated dynamics," *Nature* **261**, 459–467.

Misiurewicz, M. & Szlenk, W. [1980] "Entropy of piecewise monotone mappings," *Studia Math.* **67**, 45–63.

Appendix A

Entropy of Cycles, Patterns and Forcing Relation

In this section, some useful lemmas and theorems for the topological entropy of one-dimensional interval maps are recalled. A suggested reference is [Alsedà *et al.*, 2000].

Let X be a topological space and let $\mathcal{X}$ be the collection of all continuous maps from X to X. A cycle (C, ϕ) consists of a finite subset $C \subseteq X$ and a permutation $\phi : C \to C$. For abbreviation, the cycle (C, ϕ) is identified with C itself. For $g \in \mathcal{X}$ and C a cycle, C is said to be a cycle of g or, in other words, g has a cycle C if $g|_C$ is a permutation. Two cycles (C_1, ϕ_1) and (C_2, ϕ_2) are denoted by $(C_1, \phi_1) \sim (C_2, \phi_2)$ if and only if there exists a homeomorphism $\tau : \mathrm{conv}(C_1) \to \mathrm{conv}(C_2)$ such that $\tau \circ \phi_1 = \phi_2 \circ \tau$, where $\mathrm{conv}(S)$ is the convex hull of S. The following property is easy to check, thus the proof is omitted.

Proposition A.1. $\sim$ *is an equivalence relation.*

Let $X = I$ or $\mathbb{R}$ and let $C \subseteq X$ be a finite subset, where I is a closed finite interval in $\mathbb{R}$. The quasipartition of X by C is the set C of all maximal proper closed intervals J such that $\mathring{J} \bigcap C = \varnothing$, where $\mathring{J}$ is the interior of J. It is remarkable that if $K \neq J \in \mathrm{C}$, then $\mathring{K} \bigcap \mathring{J} = \varnothing$. The elements of $\mathrm{C} \bigcap \mathrm{conv}(C)$ are called C-basic intervals. $g \in \mathcal{X}$ is named by C-monotone if g is constant for each connected component of $X \backslash \mathrm{conv}(C)$ and g is monotone on each C-basic interval.

Let $g \in \mathcal{X}$ and let $K, J \subseteq X$ be two intervals. Then K g-covers J whenever $g(K) \supseteq J$. Given $C \subseteq X$ a finite subset, a g-graph of C, denoted by $G = (V, E)$, is a directed graph consisting of the following:

(a) The vertices set $V = \{\text{all } C\text{-basic intervals}\}$.
(b) For $K, J \in V$, if K g-covers J m times, then there are m edges from K to J.

Remark A.2. The g-graph of C is unique up to labeling of E.

Denote by $\mathcal{C}$ the set of cycles. $P \in \mathcal{C}_\sim$ is called a pattern. For any pattern $P \in \mathcal{C}_\sim$, a cycle $(C, \phi) \in P$ is said to have a pattern P. If a map g has a cycle C with pattern P, then it is remarked that g exhibits P.

Let $P_1, P_2 \in \mathcal{C}_\sim$, P_1 is said to force P_2 and denoted by $P_1 \Rightarrow P_2$ if and only if for every $g \in \mathcal{X}$ which has a cycle from P_1 has a cycle from P_2.

Let X be a compact topological space, $g \in \mathcal{X}$ and let $\mathcal{A}$ be an open cover of $\mathcal{X}$. Define $\mathcal{A}^n = \{\bigcap_{i=0}^{n-1} g^{-i} A_{k_i} : A_{k_i} \in \mathcal{A}\}$, then the limit

$$h(g, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} \log \mathcal{N}(\mathcal{A}^n) \tag{A.1}$$

exists and is called the topological entropy of g on the cover $\mathcal{A}$, herein $\mathcal{N}(\mathcal{U})$ is the minimal possible cardinality of a subcover chosen from $\mathcal{U}$. The topological entropy of g is defined by

$$h(g) = \sup h(g, \mathcal{A}), \tag{A.2}$$

where supremum is taken over all open cover of $\mathcal{X}$.

The entropy of a cycle $C \in \mathcal{C}$ is defined by

$$h(C) = \inf\{h(g) : g \in \mathcal{I} \text{ has the cycle } C\}.$$

Since topological entropy is an invariant, any two cycles having the same pattern admit exactly one entropy. This makes the discussion of entropy of a pattern P comprehensible and is denoted by $h(P)$. Note that instead of the cycles in $\mathcal{C}$ the same idea can be extended to a general notion, namely finite invariant sets.

Considering $g \in \mathcal{I}$, C is a finite g-invariant set with g-graph G. Define its corresponding matrix

$M = (m_{ij})$ in the following way.

$$m_{ij} = \begin{cases} 1, & \text{if there is an edge from the } i\text{th} \\ & \text{vertex to the } j\text{th one;} \\ 0, & \text{otherwise.} \end{cases}$$

Theorem A.3 [Alsedà *et al.*, 2000]. *Let C be a finite invariant set for g. If M is the matrix of g-graph and g is C-monotone, then $h(g) = \max\{0, \log \rho(M)\}$, where $\rho(M)$ is the spectral radius of M.*

A corollary follows immediately.

Corollary A.4. *Under the assumption of the above theorem, $h(g) = \log \rho(M)$ if and only if M is not nilpotent.*

Remark A.5. If $\log(x)$ is extensively defined by $\log : \mathbb{R}^+ \bigcup \{0\} \to \mathbb{R}$ with $\log 0 = 0$, then Theorem A.3 can be reformulated as $h(g) = \log \rho(M)$ by Perron–Frobenius theorem.

Lemma A.6 [Alsedà *et al.*, 2000]. *If P_1, P_2 are two distinct patterns such that P_1 forces P_2, then $h(P_1) \geq h(P_2)$.*

Theorem A.7 [Alsedà *et al.*, 2000]. *For any pattern P,*

(a) *$h(P) = \sup\{h(Q) : P \Rightarrow Q$ and $P \neq Q\}$.*
(b) *If g is C-monotone and M is the matrix representation of g-graph, then $h(P) = \log \rho(M)$.*

Given two graphs G and H, G is said to force H (denoted by $G \Rightarrow H$) if and only if $\mathbf{X}_H$ is topological conjugate to a subshift of $\mathbf{X}_G$. From Lemma A.6 and Theorems A.3 and A.7, the following theorem comes immediately. The proof is omitted.

Theorem A.8. *Given two distinct graphs G_1 and G_2. If $G_1 \Rightarrow G_2$, then $h(G_1) \geq h(G_2)$, where $h(G) = \log \rho(M)$ and M is the matrix representation corresponding to G. Moreover, if there exist two distinct patterns P_1, P_2 where P_1 forces P_2 such that G_1, G_2 are g_1- and g_2-graphs, respectively, where g_i has P_i for $i = 1, 2$. Then the following statements are satisfied*:

(i) *$h(G) = \sup\{h(H) : G \Rightarrow H\}$.*
(ii) *If g_i are P_i-monotone for $i = 1, 2$, then $\rho(M_1) \geq \rho(M_2)$. Herein M_i is the matrix representation of g_i-graph of P_i for $i = 1, 2$.*

It is natural to study under what kind of condition the equality does not hold anymore. If $h(P_1) \gneqq h(P_2)$, then P_1 is said to strongly force P_2.

A graph G is said to be irreducible if there exists a path for any two vertices of G. A shift space $\mathbf{X}_G$ is irreducible if and only if its underlying graph is irreducible.

Lemma A.9 [Lind & Marcus, 1995]. *Given a sofic shift $\mathbf{X}$ with transition matrix $\mathbf{T}$, there exists $\mathbf{X}_1, \ldots, \mathbf{X}_k$ so that $\mathbf{X}_i$ is an irreducible sofic subshift of $\mathbf{X}$ for each i and*

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 & 0 & 0 & 0 \\ * & \mathbf{T}_2 & 0 & 0 \\ * & * & \ddots & 0 \\ * & * & * & \mathbf{T}_k \end{bmatrix},$$

where $\mathbf{T}_i$ is the transition matrix corresponding to $\mathbf{X}_i$ for $1 \leq i \leq k$. Moreover, $h(\mathbf{X}) = \max_{1 \leq i \leq k} h(\mathbf{X}_i)$.

Theorem A.10 [Lind & Marcus, 1995]. *If $\mathbf{X}$ is an irreducible sofic shift and $\mathbf{Y}$ is a proper subshift of $\mathbf{X}$, then $h(\mathbf{Y}) < h(\mathbf{X})$.*

Remark A.11. It is also shown in [Lind & Marcus, 1995] that Lemma A.9 and Theorem A.10 remain true for subshift of finite type.

Appendix B

Separation Property

This section studies the separation property of n-dimensional convex hull in $\mathbb{R}^n$, where $\mathbb{R}$ is the set of real numbers. One of its applications is the constraint of local patterns produced by n-dimensional MCNNs.

B.1. *Separation property of convex hull in* $\mathbb{R}^n$

In this subsection, the separation property of n-dimensional convex hull will be introduced. The discussion below will focus mainly on those convex hulls with lattice vertex sets for simplicity.

Definition B.1. Denote $X = \{1, -1\}$. $\mathcal{U} \subseteq X^n$ is called separable if there is a hyperplane H in $\mathbb{R}^n$ such that $\text{conv}(\mathcal{U})$ and $\text{conv}(\mathcal{U}^c)$ can be separated by H, where $\mathcal{U}^c = X^n \backslash \mathcal{U}$ and $\text{conv}(E)$ is the convex hull of E for $E \subseteq \mathbb{Z}^n$.

A necessary condition for separability of $\mathcal{U}$ is then easy to check.

Remark B.2. If $\mathcal{U} \subseteq X^n$ is separable and $\mathcal{U}, \mathcal{U}^c$ both contain at least two points, then for each $u \in \mathcal{U}$

(respectively $\mathcal{U}^c$), there exists $u' \in \mathcal{U}$ (respectively $\mathcal{U}^c$) such that $\|u - u'\| = 2$, where $\|\cdot\|$ is the sup-norm. In other words, there is no isolated point in a separable set.

Definition B.3. Denote $S_{n,l} \subseteq X^n$ by

$$S_{n,l} = \{x = (x_1, \ldots, x_n) \in X^n : x_k = -1 \text{ for all } l+1 \le k \le n\} \tag{B.1}$$

for $1 \le l \le n-1$, and $S_{n,n} = X^n$. $S_{n,l}$ is called the l-dimensional subcube (i-subcube, for abbreviation) in $\mathbb{R}^n$.

Lemma B.4. *$S_{n,l}$ is separable in $\mathbb{R}^n$ for all $l \le n$. Furthermore, $\tilde{S}_{n,l} \equiv (S_{n,l}, S_{n,l}) \subseteq X^{2n}$ is also separable.*

Proof. To show that $S_{n,i}$ is separable, defining the linear functional $g : \mathbb{R}^n \to \mathbb{R}$ to be

$$g(x) = \sum_{i=l+1}^{n} x_i \quad \text{for all } x = (x_i)_{i=l}^n \in \mathbb{R}^n. \tag{B.2}$$

Let $h(x) = g(x) + (n - l - 1)$. Then it can be easily checked that $h(x) < 0$ for all $x \in S_{n,l}$, and $h(x) > 0$ for all $x \in S^c_{n,l}$. That is, $S_{n,l}$ and $S^c_{n,l}$ can be separated by the hyperplane

$$H = \{x \in \mathbb{R}^n : g(x) = l - n + 1\}. \tag{B.3}$$

Similarly, consider the linear functional $\hat{g} : \mathbb{R}^{2n} \to \mathbb{R}$ defined by $\hat{g}(z) = g(z_1) + g(z_2)$, where $z = (z_1, z_2) \in \mathbb{R}^{2n}$ and $z_i \in \mathbb{R}^n$ for $i = 1, 2$. Then $\tilde{H} = \{z \in \mathbb{R}^{2n} : \hat{g}(z) = 2l - 2n + 1\}$ is the desired separating hyperplane in $\mathbb{R}^{2n}$ for $\tilde{S}_{n,l}$ and $\tilde{S}^c_{n,l}$. This completes the proof. ∎

Remark B.5. In general, the separability of $\mathcal{U} \subseteq X^n$ does not imply the separability of $\tilde{\mathcal{U}} \equiv (\mathcal{U}, \mathcal{U}) \subseteq X^{2n}$. For instance, consider $\mathcal{U} = \{(-1,-1), (-1,1), (1,-1)\} \subseteq X^2$. It is easy to check that $\mathcal{U}$ is separable; however, $\tilde{\mathcal{U}}$ is not separable.

Conversely, the projection of a separable set is still separable.

Definition B.6. Given $n, k, l \in \mathbb{N}$, $k, l < n$. Define the front projection $\pi_k : \mathbb{R}^n \to \mathbb{R}^k$ as

$$\pi_k(x) = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k, \quad \text{where } x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n. \tag{B.4}$$

Moreover, the back projection $\tau_l : \mathbb{R}^n \to \mathbb{R}^l$ is defined by

$$\tau_l(x) = (x_{n-l+1}, x_{n-l+2}, \ldots, x_n) \in \mathbb{R}^l, \quad \text{where } x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n. \tag{B.5}$$

Lemma B.7. *Given $\mathcal{U} \subseteq X^n$, let $\mathcal{U}_k \equiv \pi_k(\mathcal{U})$ and $\mathcal{U}_l \equiv \tau_l(\mathcal{U})$, $1 \le k, l \le n$. Then $\mathcal{U}_k$ and $\mathcal{U}_l$ are both separable if $\mathcal{U}$ is separable.*

Proof. For simplicity, it suffices to show that $\mathcal{U}_k$ is separable. The separability of $\mathcal{U}_l$ can be done analogously.

If $\mathcal{U}_k$ is not separable, then $\mathrm{conv}(\mathcal{U}_k) \cap \mathrm{conv}(\mathcal{U}^c_k) \neq \varnothing$. That is, there exist $x_1, x_2, \ldots, x_j \in \mathcal{U}_k$ and $0 < t_1, t_2, \ldots, t_j < 1, t_1 + t_2 + \cdots + t_j = 1$ such that $t_1x_1 + t_2x_2 + \cdots + t_jx_j \in \mathrm{conv}(\mathcal{U}^c_k)$. Moreover, there exist $w_1, w_2, \ldots, w_j \in X^{n-k}$ such that $(x_i, w_i) \in \mathcal{U}$ for $1 \le i \le j$, and $(y, w_i) \in \mathcal{U}^c$ for all $y \in \mathcal{U}^c_k, 1 \le i \le j$. Since $\mathrm{conv}(\mathcal{U}^c)$ is the convex hull generated by $\mathcal{U}^c$,

$$(y, s_1w_1 + s_2w_2 + \cdots + s_jw_j) \in \mathrm{conv}(\mathcal{U}^c), \quad \text{for all } y \in \mathrm{conv}(\mathcal{U}^c_k), \tag{B.6}$$

$0 \le s_1, s_2, \ldots, s_j \le 1$, and $s_1 + s_2 + \cdots + s_j = 1$. In other words,

$$(t_1x_1 + t_2x_2 + \cdots + t_jx_j, t_1w_1 + t_2w_2 + \cdots + t_jw_j) \in \mathrm{conv}(\mathcal{U}^c). \tag{B.7}$$

However, $\mathrm{conv}(\mathcal{U})$ is the convex hull generated by $\mathcal{U}$, that means

$$(t_1x_1 + t_2x_2 + \cdots + t_jx_j, t_1w_1 + t_2w_2 + \cdots + t_jw_j) \in \mathrm{conv}(\mathcal{U}). \tag{B.8}$$

Thus, $\mathrm{conv}(\mathcal{U}) \cap \mathrm{conv}(\mathcal{U}^c) \neq \varnothing$. This is a contradiction, and the proof is completed. ∎

Instead of projection, Lemma B.7 can be generalized to an abstract form.

Theorem B.8. *Let $\mathcal{U} \subseteq X^{n+m}$ be separable, $\mathcal{V} \subseteq X^m$. Define $\tilde{\mathcal{U}}$ as*

$$\tilde{\mathcal{U}} \equiv \{u \in X^n : \text{There exists } v \in \mathcal{V} \text{ such that } (u, v) \in \mathcal{U}\}. \tag{B.9}$$

Then $\tilde{\mathcal{U}}$ is separable.

Proof. Using Lemma B.7, it suffices to show that $\mathcal{V} \subsetneq \mathcal{U}_m$, where $\mathcal{U}_m = \tau_m(\mathcal{U})$.

If $\tilde{\mathcal{U}}$ is not separable, then $\mathrm{conv}(\tilde{\mathcal{U}}) \cap \mathrm{conv}(\tilde{\mathcal{U}}^c) \neq \varnothing$. That is, there exist $x_1, x_2, \ldots, x_k \in \tilde{\mathcal{U}}$ and $0 < t_1, t_2, \ldots, t_k < 1, t_1 + t_2 + \cdots + t_k = 1$ such that $t_1x_1 + t_2x_2 + \cdots + t_kx_k \in \mathrm{conv}(\tilde{\mathcal{U}}^c)$. Moreover, there exist $w_1, w_2, \ldots, w_k \in \mathcal{V}$ such that $x_iw_i \in \mathcal{U}$ for $1 \le i \le k$, and $yw_i \in \mathcal{U}^c$ for all $y \in \tilde{\mathcal{U}}^c, 1 \le i \le k$.

Since $\mathrm{conv}(\mathcal{U}^c)$ is the convex hull generated by $\mathcal{U}^c$,

$$(y, s_1w_1 + s_2w_2 + \cdots + s_kw_k) \in \mathrm{conv}(\mathcal{U}^c),\quad \text{for all } y \in \mathrm{conv}(\tilde{\mathcal{U}}^c), \tag{B.10}$$

$0 \leq s_1, s_2, \ldots, s_k \leq 1$, and $s_1 + s_2 + \cdots + s_k = 1$. In other words,

$$(t_1x_1 + t_2x_2 + \cdots + t_kx_k, t_1w_1 + t_2w_2 + \cdots + t_kw_k) \in \mathrm{conv}(\mathcal{U}^c). \tag{B.11}$$

However, $\mathrm{conv}(\mathcal{U})$ is the convex hull generated by $\mathcal{U}$, that means

$$(t_1x_1 + t_2x_2 + \cdots + t_kx_k, t_1w_1 + t_2w_2 + \cdots + t_kw_k) \in \mathrm{conv}(\mathcal{U}). \tag{B.12}$$

Thus, $\mathrm{conv}(\mathcal{U}) \cap \mathrm{conv}(\mathcal{U}^c) \neq \varnothing$. This is a contradiction, and the proof is completed. ■

Next subsection introduces an application of separation property of n-dimensional convex hull to n-dimensional MCNNs, where $n \in \mathbb{N}$. For simplicity, the theorems will be given for the one-dimensional case. The general case can be done analogously.

B.2. *An application to multilayer cellular neural networks*

B.2.1. *Cellular neural networks with input*

One-dimensional CNN with input is of the form,

$$\frac{dx_i}{dt} = -x_i + z + \sum_{|k|\leq d} a_k f(x_{i+k}) + \sum_{|k|\leq d} b_k u_{i+k}, \tag{B.13}$$

for $i \in \mathbb{Z}$, $d \in \mathbb{N}$, $f(x)$ is a piecewise-linear output function defined by

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

Herein, $A = [a_{-d}, \ldots, a_d]$ and $B = [b_{-d}, \ldots, b_d]$ are called feedback and controlling templates, respectively; z is called a biased term or threshold. The quantity x_i denotes the state at cell C_i, and y_i denotes the output at C_i.

As generally known, stationary solutions $\bar{x} = (\bar{x}_i)$ are essential for understanding CNN, in which their output is called patterns. $\bar{x}$ is called a mosaic solution if $|\bar{x}_i| > 1$ for all $i \in \mathbb{Z}$. For a given mosaic solution $\bar{x}$, the output pattern of cell C_i is $+$, i.e. $\bar{x}_i > 1$, if and only if

$$\sum_{|k|\leq d, k\neq 0} a_k \bar{y}_{i+k} + \sum_{|k|\leq d} b_k \bar{u}_{i+k} + a + z - 1 > 0. \tag{B.14}$$

Similarly, the output pattern of cell C_i is $-$, i.e. $\bar{x}_i < -1$, if and only if

$$\sum_{|k|\leq d, k\neq 0} a_k \bar{y}_{i+k} + \sum_{|k|\leq d} b_k \bar{u}_{i+k} - a + z + 1 < 0. \tag{B.15}$$

Denote by $n = 2d$. For a given pair of template A and threshold z, the basic set of admissible local patterns with "+" state and "−" state in the center are defined by

$$\tilde{\mathcal{B}}(+, A, B, z) = \{(y, u) \in X^n \times X^{n+1} : \alpha \cdot y + \beta \cdot u + a + z - 1 > 0\}, \tag{B.16}$$

and

$$\tilde{\mathcal{B}}(-, A, B, z) = \{(y, u) \in X^n \times X^{n+1} : \alpha \cdot y + \beta \cdot u - a + z + 1 < 0\}, \tag{B.17}$$

respectively, where $\alpha = (a_{-d}, \ldots, a_d)$, $\beta = (b_{-d}, \ldots, b_d)$, $y = (y_1, \ldots, y_n)$ and $u = (u_1, \ldots, u_{n+1})$ is obtained from

$$[y_1, \ldots, y_d, y_i, y_{d+1}, \ldots, y_n] = [y_{i-d}, \ldots, y_i, \ldots, y_{i+d}]$$

and

$$[u_1, \ldots, u_d, \ldots, u_{n+1}] = [u_{i-d}, \ldots, u_i, \ldots, u_{i+d}],$$

respectively. In other words, $\tilde{\mathcal{B}}(+, A, B, z)$ and $\tilde{\mathcal{B}}(-, A, B, z)$ represent the sets of output patterns coupled with input patterns with "+" and "−" in the center of output patterns, respectively. Let $\mathcal{B}(+, A, B, z) = \pi_n(\tilde{\mathcal{B}}(+, A, B, z))$ be the set of output patterns with "+" in the center and let $\mathcal{B}(-, A, B, z) = \pi_n(\tilde{\mathcal{B}}(-, A, B, z))$ be the set of output patterns with "−" in the center. The separation property of one-layer CNN with input is indicated in [Ban *et al.*, 2009].

Theorem B.9 [Ban *et al.*, 2009]. *Given* $\mathcal{U} \subseteq X^n \times X^{n+1}$, *there exists* (A, B, z) *such that* $\mathcal{U} = \tilde{\mathcal{B}}(*, A, B, z)$ *for some* $* \in \{+, -\}$ *if and only if* $\mathcal{U}$ *is separable.*

The following theorem shows the separation property of the feasible *output* local patterns produced by one-layer CNN with input.

Theorem B.10. *Given* $\mathcal{U} \subseteq X^n$, *there exists* (A, B, z) *such that* $\mathcal{U} = \mathcal{B}(*, A, B, z)$ *for some* $* \in \{+, -\}$ *if and only if* $\mathcal{U}$ *is separable.*

Proof. Without loss of generality, it suffices to show that the theorem holds for the case $*$ is $+$. Another case can be done analogously.

If there exists (A, B, z) such that $\mathcal{U} = \mathcal{B}(+, A, B, z)$, then there exists $\tilde{\mathcal{U}} \subseteq X^n \times X^{n+1}$ such that $\tilde{\mathcal{B}}(+, A, B, z) = \tilde{\mathcal{U}}$. Thus $\tilde{\mathcal{B}}(+, A, B, z)$ is separable by Theorem B.9, the separability of $\mathcal{U}$ then follows from Lemma B.7.

Conversely, construct $\tilde{\mathcal{U}}$ from $\mathcal{U}$ by

$$\tilde{\mathcal{U}} = \{(y, u) : y \in \mathcal{U},\ u = \{1\}^{\mathbb{Z}_{(n+1)\times(n+1)}}\}.$$

Then $\tilde{\mathcal{U}}$ is separable. By Theorem B.9, there exists (A, B, z) such that

$$\tilde{\mathcal{B}}(+, A, B, z) = \tilde{\mathcal{U}}.$$

This implies $\mathcal{B}(+, A, B, z) = \mathcal{U}$, and the proof is completed. ■

B.2.2. *Multilayer cellular neural networks*

In this subsection, the separation property of MCNNs is investigated. MCNN is of the form,

$$\frac{dx_i^{(\ell)}}{dt} = -x_i^{(\ell)} + z^{(\ell)} + \sum_{|k|\le d} a_k^{(\ell)} f(x_{i+k}^{(\ell)}) + \sum_{|k|\le d} b_k^{(\ell)} u_{i+k}^{(\ell)}, \tag{B.18}$$

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

$$u_i^{(\ell)} = y_i^{(\ell-1)}, \quad 2 \le \ell \le N, \tag{B.19}$$

$$u_i^{(1)} = u_i. \tag{B.20}$$

For simplicity, restrict the discussion for the case $N = 2$. The general case $N \ge 2$ can be done via analogous method and mathematical induction.

Two-layer CNN with input can be realized as the following.

$$\frac{dx_i^{(2)}}{dt} = -x_i^{(2)} + \sum_{|k|\le d} a_k^{(2)} y_{i+k}^{(2)} + \sum_{|k|\le d} b_k^{(2)} u_{i+k}^{(2)} + z^{(2)},$$

$$\frac{dx_i^{(1)}}{dt} = -x_i^{(1)} + \sum_{|k|\le d} a_k^{(1)} y_{i+k}^{(1)} + \sum_{|k|\le d} b_k^{(1)} u_{i+k}^{(1)} + z^{(1)}, \tag{B.21}$$

for some $d \in \mathbb{N}, i \in \mathbb{Z}$, where

$$u_i^{(2)} = y_i^{(1)}, \quad u_i^{(1)} = u_i. \tag{B.22}$$

The feedback templates and controlling templates of each layer are $A^{(\ell)}$ and $B^{(\ell)}$ for $\ell \in \{1, 2\}$, and the threshold of each layer is $z^{(\ell)}$, $\ell \in \{1, 2\}$. Denote by $\mathbf{A} = (A^{(1)}, A^{(2)}), \mathbf{B} = (B^{(1)}, B^{(2)})$ and $\mathbf{z} = (z^{(1)}, z^{(2)})$. For $\ell \in \{1, 2\}$, if $(A^{(\ell)}, B^{(\ell)}, z^{(\ell)})$ is given, then the basic set of admissible local patterns,

$$\tilde{\mathcal{B}}^{(\ell)}(+, A^{(\ell)}, B^{(\ell)}, z^{(\ell)}) = \left\{\begin{matrix}(y^{(\ell)}, u^{(\ell)}) \in X^n \times X^{n+1} : \\ \alpha^{(\ell)} \cdot y^{(\ell)} + \beta^{(\ell)} \cdot u^{(\ell)} + a^{(\ell)} + z^{(\ell)} - 1 > 0\end{matrix}\right\},$$

and

$$\tilde{\mathcal{B}}^{(\ell)}(-, A^{(\ell)}, B^{(\ell)}, z^{(\ell)}) = \left\{\begin{matrix}(y^{(\ell)}, u^{(\ell)}) \in X^n \times X^{n+1} : \\ \alpha^{(\ell)} \cdot y^{(\ell)} + \beta^{(\ell)} \cdot u^{(\ell)} - a^{(\ell)} + z^{(\ell)} + 1 < 0\end{matrix}\right\},$$

are determined. As defined in the last section, $\alpha^{(\ell)}$ represents the surrounding template of $A^{(\ell)}$ without center, $\beta^{(\ell)}$ represents template $B^{(\ell)}$, and $y^{(\ell)}, u^{(\ell)}$ represent the output and input patterns of cell $C_i^{(\ell)}$ in ℓth layer.

Furthermore, let $\mathcal{B}(+, \mathbf{A}, \mathbf{B}, \mathbf{z}) \subseteq X^n$ be defined by $y \in \mathcal{B}(+, \mathbf{A}, \mathbf{B}, \mathbf{z})$ if and only if

(a) There exists $u \in X^{n+1}$ such that $(y, u) \in \tilde{\mathcal{B}}^{(2)}(+, A^{(2)}, B^{(2)}, z^{(2)})$.
(b) There exists $v \in X^{n+1}$ such that $(u', v) \in \tilde{\mathcal{B}}^{(1)}(*, A^{(1)}, B^{(1)}, z^{(1)})$ for some $* \in \{+, -\}$, where $u' \in X^n$ is obtained from u by deleting the entry in the center.

In other words, $\mathcal{B}(+, \mathbf{A}, \mathbf{B}, \mathbf{z})$ consists of the feasible output patterns of the 2-layer CNN system with "+" in the center. Similarly, the basic set of admissible output patterns with "−" in the center, denoted by $\mathcal{B}(-, \mathbf{A}, \mathbf{B}, \mathbf{z}) \subseteq X^n$, is $y \in \mathcal{B}(-, \mathbf{A}, \mathbf{B}, \mathbf{z})$ if and only if

(a) There exists $u \in X^{n+1}$ such that $(y, u) \in \tilde{\mathcal{B}}^{(2)}(-, A^{(2)}, B^{(2)}, z^{(2)})$.

(b) There exists $v \in X^{n+1}$ such that $(u', v) \in \tilde{\mathcal{B}}^{(1)}(*, A^{(1)}, B^{(1)}, z^{(1)})$ for some $* \in \{+, -\}$.

The separation property theorem for 2-layer CNN then follows.

Theorem B.11. *Given $\mathcal{U} \subseteq X^n$, there exists $(\mathbf{A}, \mathbf{B}, \mathbf{z})$ such that $\mathcal{U} = \mathcal{B}(*, \mathbf{A}, \mathbf{B}, \mathbf{z})$ for some $* \in \{+, -\}$ if and only if $\mathcal{U}$ is separable.*

Proof. The proof will be given for the case that $*$ is $+$, another case can be considered in an analogous way. For simplicity, denote $\mathcal{B}(+, \mathbf{A}, \mathbf{B}, \mathbf{z}) = \mathcal{B}$, and $\tilde{\mathcal{B}}^{(\ell)}(+, A^{(\ell)}, B^{(\ell)}, z^{(\ell)}) = \tilde{\mathcal{B}}^{(\ell)}$ for $\ell = 1, 2$.

If there exists $(\mathbf{A}, \mathbf{B}, \mathbf{z})$ such that $\mathcal{U} = \mathcal{B}$, then $\tilde{\mathcal{B}}^{(1)}$, $\tilde{\mathcal{B}}^{(2)}$ are both separable. Let $\hat{\mathcal{B}} \subseteq X^{n+1}$ be defined by

$$\hat{\mathcal{B}} = \{u : u' \in \pi_n(\tilde{\mathcal{B}}^{(1)})\}, \tag{B.23}$$

where u' is obtained from u by deleting the entry in the center. Then $\mathcal{B} = \{y : \exists\ u \in \hat{\mathcal{B}}$ such that $(y, u) \in \tilde{\mathcal{B}}^{(2)}\}$. The separability of $\mathcal{B}$ follows by Theorem B.8.

Conversely, if $\mathcal{U}$ is separable. Let $v = \{1\}^{\mathbb{Z}_{(n+1)\times 1}}$, then $\mathcal{U}_2 = \{(u, v) : u \in \mathcal{U}\}$ is separable. Thus, there exists $(A^{(2)}, B^{(2)}, z^{(2)})$ such that $\tilde{\mathcal{B}}^{(2)} = \mathcal{U}_2$. Similarly, $\mathcal{U}_1 = \{(v', v)\}$ is also separable, there exists $(A^{(1)}, B^{(1)}, z^{(1)})$ such that $\tilde{\mathcal{B}}^{(1)} = \mathcal{U}_1$. Let $A = (A^{(1)}, A^{(2)}), B = (B^{(1)}, B^{(2)}), z = (z^{(1)}, z^{(2)})$, then $\mathcal{B} = \mathcal{U}$. This completes the proof. ■