Scientific
Research

# Hausdorff Dimension of Multi-Layer Neural Networks

**Jung-Chao Ban[1*], Chih-Hung Chang[2#]**
[1]Department of Applied Mathematics, National Dong Hwa University, Hualien, Taiwan
[2]Department of Applied Mathematics, Feng Chia University, Taichung, Taiwan
Email: jcban@mail.ndhu.edu.tw, chihhung@mail.fcu.edu.tw

## ABSTRACT

This elucidation investigates the Hausdorff dimension of the output space of multi-layer neural networks. When the factor map from the covering space of the output space to the output space has a synchronizing word, the Hausdorff dimension of the output space relates to its topological entropy. This clarifies the geometrical structure of the output space in more details.

**Keywords:** Multi-Layer Neural Networks; Hausdorff Dimension; Sofic Shift; Output Space

## 1. Introduction

The multi-layer neural networks (MNN, [1,2]) have received considerable attention and were successfully applied to many areas such as signal processing, pattern recognition ([3,4]) and combinatorial optimization ([5,6]) in the past few decades. The investigation of mosaic solution is the most essential in MNN models due to the learning algorithm and training processing. In [7-9], the authors proved that the output solutions space $\mathbf{Y}^{(2)}$ of a 2-layer MNN forms a so-called sofic shift space, which is a factor of a classical subshift of finite type. Thus, MNN model indeed produces abundant output patterns and makes learning algorithm more efficient. A useful quantity to classify the output solution space is the topological entropy $h(\mathbf{Y}^{(2)})$ ([10]). We call the output solution space *pattern formation* if $h(\mathbf{Y}^{(2)}) = 0$, and call it *spatial chaos* if $h(\mathbf{Y}^{(2)}) > 0$. The $h(\mathbf{Y}^{(2)}) = 0$ indicates that the output patterns grow subexponentially and exponentially for $h(\mathbf{Y}^{(2)}) > 0$. For positive entropy systems, the explicit value of $h(\mathbf{Y}^{(2)})$ presents how chaotic the system is. In [7], Ban and Chang provided a method to compute explicit values of $h(\mathbf{Y}^{(2)})$ for a 2-layer MNN. The method is quite general and it makes the computation of $h(\mathbf{Y}^{(N)})$ possible for arbitrary

$2 < N \in \mathbb{N}$, *i.e.*, $N$-layer MNN.

From the dynamical system (DS) point of view, the topological entropy reveals the complexity of the global patterns. However, it provides less information of the inner structure of a given DS, e.g., self-similarity or recurrent properties. The possible quantity reveals that such properties are the Hausdorff dimension (HD, [11]) since the Hausdorff dimension is an indicator of the geometrical structure. For most DS, the computation of Hausdorff dimension is not an easy task, and the box dimension (BD) is usually computed first to give the upper bound for HD. Due to the relationship of topological entropy and BD ([12]) of a symbolic DS[1], the previous work ([7]) for topological entropy gives the upper bound for HD of $N$-layer MNN. Nature question arises: Given a MNN, how to compute the explicit value for HD? The aim of this paper is to establish the HD formula for $N$-layer MNNs. Using the tool of symbolic DS, the HD formula will be established for $N$-layer MNNs which possesses a *synchronizing word* (Theorem 2.4). The result leads us to exploit the inner structure for a $N$-layer MNN. We believe that further interesting applications of the results presented (or of the generalizations) can be obtained.

[1]Pesin showed that the box dimension of a shift space $X$ is the quotient of the topological entropy and the metric on $X$. More precisely, if the metric is defined by $d(x,y) = \sum_{i \in \mathbb{Z}} \frac{|x_i - y_i|}{\beta^{|i|+1}}$ for $x, y \in X, \beta > 1$, then the box dimension is $\dim_B X = h(X)/\log \beta$.

This paper is organized as follows. Section 2 contains a brief disscussion for the computation of topological entropy in [7]. The main result is stated and proved therein. Section 3 presents an MNN model for which we can compute its HD.

## 2. Preliminaries and Main Results

A one-dimensional multi-layer neural network (MNN) is realized as

$$
\begin{cases}
\dfrac{\mathrm{d}}{\mathrm{d}t} x_i^{(k)}(t) = -x_i^{(k)}(t) + z^{(k)} \\
\qquad\qquad + a^{(k)} f\left(x_i^{(k)}(t)\right) + \sum_{\ell \in \mathcal{N}} b_\ell^{(k)} f\left(x_{i+\ell}^{(k-1)}(t)\right), \\
\dfrac{\mathrm{d}}{\mathrm{d}t} x_i^{(1)}(t) = -x_i^{(1)}(t) + z^{(1)} \\
\qquad\qquad + a^{(1)} f\left(x_i^{(1)}(t)\right) + \sum_{\ell \in \mathcal{N}} a_\ell^{(1)} f\left(x_{i+\ell}^{(1)}(t)\right),
\end{cases} \tag{1}
$$

for some $N \in \mathbb{N}$, $k = 2, \cdots, N$ and $i \in \mathbb{Z}$. The finite subset $\mathcal{N} \subset \mathbb{Z}$ indicates the *neighborhood*, and the piecewise linear map

$$
f(x) = \frac{1}{2}\left(|x+1| - |x-1|\right)
$$

is called the *output function*. The *template* $\mathbb{T} = [A, B, z]$ is composed of *feedback template* $A = (A_1, A_2)$ with

$$
A_1 = \left(a^{(1)}, \cdots, a^{(N)}\right), \quad A_2 = \left(a_\ell^{(1)}\right)_{\ell \in \mathcal{N}},
$$

*controlling template*

$$
B = \left(B_2, \cdots, B_N\right),
$$

and *threshold*

$$
z = \left(z^{(1)}, \cdots, z^{(N)}\right),
$$

where

$$
B_k = \left(b_\ell^{(k)}\right)_{\ell \in \mathcal{N}} \quad \text{for } k \geq 2.
$$

A solution

$$
x(t) = \left(x_i^{(1)}(t), \cdots, x_i^{(N)}(t)\right)_{i \in \mathbb{Z}} \in \mathbb{R}^{\mathbb{Z}_{\infty \times N}}
$$

of (1) is called *mosaic* if

$$
\left|x_i^{(k)}(t)\right| > 1
$$

for $1 \leq k \leq N, i \in \mathbb{Z}$, and $t \geq T_0$ for some $T_0$.

$$
y = \left(y_i^{(1)} \cdots y_i^{(N)}\right)_{i \in \mathbb{Z}} \in \{-1, 1\}^{\mathbb{Z}_{\infty \times N}}
$$

of a mosaic solution is called a *mosaic pattern*, where

$$
y_i^{(k)} = f\left(x_i^{(k)}\right).
$$

The *solution space* $\mathbf{Y}$ of (1) stores the patterns $y$, and the *output space* $\mathbf{Y}^{(N)}$ of (1) is the collection of the output patterns; more precisely,

$$
\mathbf{Y}^{(N)} = \left\{\left(y_i^{(N)}\right)_{i \in \mathbb{Z}} : \left(y_i^{(1)} \cdots y_i^{(N)}\right)_{i \in \mathbb{Z}} \in \mathbf{Y}\right\}.
$$

A neighborhood $\mathcal{N}$ is called the nearest neighborhood if $\mathcal{N} = \{-1, 1\}$. In [7], the authors showed that $2$-layer MNNs with nearest neighborhood are essential for the investigation of MNNs. In the rest of this manuscript, we refer MNNs to $2$-layer MNNs with nearest neighborhood unless otherwise stated.

### 2.1. Topological Entropy and Hausdorff Dimension

Since the neighborhood $\mathcal{N}$ is finite and is invariant for each $i$, the output space is determined by the so-called *basic set of admissible local patterns*. Replace the pattern $-1$ and $1$ by $-$ and $+$, respectively; the basic set of admissible local patterns of the first and second layer is a subset of

$$
\{---, --+, -+-, -++, +--, +-+, ++-, +++\}; \tag{2}
$$

and $\{p_1, \cdots, p_8\}$, respectively, where $\{p_1, \cdots, p_8\}$ denotes

$$
\left\{\begin{matrix} - \\ - \, - \end{matrix}, \begin{matrix} - \\ - \, + \end{matrix}, \begin{matrix} - \\ + \, - \end{matrix}, \begin{matrix} - \\ + \, + \end{matrix}, \begin{matrix} + \\ - \, - \end{matrix}, \begin{matrix} + \\ - \, + \end{matrix}, \begin{matrix} + \\ + \, - \end{matrix}, \begin{matrix} + \\ + \, + \end{matrix}\right\}. \tag{3}
$$

To ease the notation, we denote

$$
\begin{matrix} \alpha_1 \\ \alpha_2 \; \alpha_3 \end{matrix}
$$

by $\alpha_1 \Diamond \alpha_2 \alpha_3$. Given a template $\mathbb{T}$, the basic set of admissible local pattern

$$
\mathcal{B}(\mathbb{T}) = \left(\mathcal{B}^{(1)}, \mathcal{B}^{(2)}\right)
$$

is determined, where $\mathcal{B}^{(1)}$ and $\mathcal{B}^{(2)}$ are the basic set of admissible local patterns of the first and second layer, respectively. Let

$$
\mathcal{P}_8 = \left\{\left(a^{(2)}, b_{-1}^{(2)}, b_1^{(2)}, z^{(2)}, a^{(1)}, a_{-1}^{(1)}, a_1^{(1)}, z^{(1)}\right)\right\}
$$

denote the parameter space of (1). Theorem 2.1 asserts that $\mathcal{P}_8$ can be partitioned into finitely many subregions so that two templates in the same partition exhibit the same basic set of admissible local patterns.

**Theorem 2.1.** (See [7]) *There is a positive integer $K$ and unique set of open subregions $\{P_k\}_{k=1}^{K}$ satisfying* [(*i*)]

1) $\mathcal{P}_8 = \bigcup_{k=1}^{K} \bar{P}_k$.

2) $P_i \cap P_j = \varnothing$ if $i \neq j$.

3) Templates $\mathbb{T}, \mathbb{T}' \in P_k$ for some $k$ if and only if $\mathcal{B}(\mathbb{T}) = \mathcal{B}(\mathbb{T}')$.

Since the template of MNNs is spatially invariant, the so-called *transition matrix* is used to investigate the complexity of MNNs. The transition matrix $T$ is defined by

$$T(i,j) = \begin{cases} 1, & p_i, p_j \in \mathcal{B}^{(2)} \quad \text{and} \quad \alpha_{i-1}\alpha_{j-1}\alpha_{i+1}, \alpha_{j-1}\alpha_{i+1}\alpha_{j+1} \in \mathcal{B}^{(1)}; \\ 0, & \text{otherwise}; \end{cases} \tag{4}$$

herein $p_k$ is presented as $\alpha_k \Diamond \alpha_{k-1}\alpha_{k+1}$ for $k=1,\cdots,8$. Furthermore, the transition matrix of the second layer

$$T_2 \in \mathcal{M}_{8\times 8}\left(\{0,1\}\right)$$

is defined by

$$T_2(i,j) = 1 \quad \text{if and only if} \quad p_i, p_j \in \mathcal{B}^{(2)}; \tag{5}$$

the transition matrix of the first layer

$$T_1 \in \mathcal{M}_{4\times 4}\left(\{0,1\}\right)$$

is defined by

$$T_1(i,j) = 1 \quad \text{if and only if} \quad \mathbb{X}(i,j) \in \mathcal{B}^{(1)}, \tag{6}$$

where

$$\mathbb{X} = \begin{array}{c|cccc} & \boxed{--} & \boxed{-+} & \boxed{+-} & \boxed{++} \\ \hline \boxed{--} & \boxed{---} & \boxed{--+} & \varnothing & \varnothing \\ \boxed{-+} & \varnothing & \varnothing & \boxed{-+-} & \boxed{-++} \\ \boxed{+-} & \boxed{+--} & \boxed{+-+} & \varnothing & \varnothing \\ \boxed{++} & \varnothing & \varnothing & \boxed{++-} & \boxed{+++} \end{array}.$$

Write

$$T_1^2 = \left(T_{i,j}\right)_{i,j=1}^2$$

as four smaller $2\times 2$ matrices. Define $\bar{T}_1$ by

$$\bar{T}_1(p,q) = T_{i,j}(k,l), \quad \text{where} \quad p = 2i+j-2, q = 2k+l-2. \tag{7}$$

Ban and Chang [7] decomposed $T$ as the product of $T_1$ and $T_2$.

**Theorem 2.2.** (See [7]) *Suppose $T$ is the transition matrix of* (1), *and $T_1$ and $T_2$ are the transition matrices of the first and second layer, respectively. Let $\bar{T}_1$ be defined as in* (7). *Then*

$$T = T_2 \circ \left(E_2 \otimes \bar{T}_1\right), \tag{8}$$

where $E_k$ is a $k\times k$ matrix with all entries being 1's; $\circ$ and $\otimes$ are the Hadamard and Kronecker product, respectively.

As being demonstrated in [7-9,13], the solution space $\mathbf{Y}$ is a so-called *shift of finite type* (SFT, also known as a topological Markov shift) and the output space $\mathbf{Y}^{(2)}$ is a *sofic shift*. More specifically, a SFT can be represented as a directed graph $G=(\mathcal{V},\mathcal{E})$ and a sofic shift can be represented as a *labeled graph* $\mathcal{G}=(G,\mathcal{L})$ for some labeling $\mathcal{L}:\mathcal{E}\to\mathcal{A}$ and finite alphabet $\mathcal{A}$. A labeled graph $\mathcal{G}=(G,\mathcal{L})$ is called *right-resolving* if the restriction of $\mathcal{L}$ to $\mathcal{E}_I$ is one-to-one for all $I\in\mathcal{V}$, where $\mathcal{E}_I$ consists of those edges starting from $I$. If $\mathcal{G}$ is not right-solving, there exists a labeled graph $\mathcal{H}$, derived by applying *subset construction method* (SCM) to $\mathcal{G}$, such that the sofic shift represented by $\mathcal{H}$ is identical to the original space. A detailed instruction is referred to [14].

One of the most frequently used quantum for the measure of the spatial complexity is the *topological entropy*. Let $X$ be a symbolic space and let $\Sigma_n(X)$ denote the collection of the patterns of length $n$ in $X$. The topological entropy of $X$ is defined by

$$h(X) = \lim_{n\to\infty} \frac{\log \Gamma_n(X)}{n}, \quad \text{provided the limit exists.}$$

Herein

$$\Gamma_n(X) = \left|\Sigma_n(X)\right|$$

indicates the cardinality of $\Sigma_n(X)$.

**Theorem 2.3.** (See [7,9,13]) *Let $\mathcal{G}$ be the labeled graph obtained from the transition matrix $T$ of* (1). *The topological entropies of $\mathbf{Y}$ and $\mathbf{Y}^{(2)}$ are $h(\mathbf{Y}) = \log \rho_T$ and*

$$h\left(\mathbf{Y}^{(2)}\right) = \begin{cases} \log \rho_T, & \text{if } \mathcal{G} \text{ is right-resolving}; \\ \log \rho_H, & \text{otherwise}; \end{cases} \tag{9}$$

*respectively, where $H$ is the transition matrix of the labeled graph $\mathcal{H}$ which is obtained by applying SCM to $\mathcal{G}$.*

Aside from the topological entropy, the Hausdorff di-

mension 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. 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

$$\mathcal{A} = \{0, 1, \cdots, n-1\}.$$

The full $\mathcal{A}$-shift $\mathcal{A}^{\mathbb{Z}}$ is the collection of all biinfinite sequences with entries from $\mathcal{A}$. More precisely,

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

It is well-known that $\mathcal{A}^{\mathbb{Z}}$ is a compact metric space endowed with the metric

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

Suppose $X$ is a subspace of $\mathcal{A}^{\mathbb{Z}}$. Set

$$X^+ = \left\{ (x_i)_{i \geq 0} : (x_i)_{i \in \mathbb{Z}} \in X \right\}$$
$$\text{and } X^- = \left\{ (x_i)_{i \leq 0} : (x_i)_{i \in \mathbb{Z}} \in X \right\}.$$

It follows that $X^+$ and $X^-$ can be embedded in the close interval $[0, 1]$ separately. Moreover, $\mathcal{A}^{\mathbb{Z}^+}$ and $\mathcal{A}^{\mathbb{Z}^-}$ can be mapped onto the close interval $[0, 1]$, and $X$ is identified with the direct product $X^+ \times X^-$. This makes the elucidation of the Hausdorff dimension of the output space $\mathbf{Y}^{(2)}$ comprehensible. (Recall that the alphabet $\mathcal{A}$ of $\mathbf{Y}^{(2)}$ is $\mathcal{A} = \{-, +\} := \{0, 1\}$).

## 2.2. Main Result

Suppose $X, Y$ are shift spaces and $\phi : X \to Y$ is a factor map. We say that $\phi$ has a *synchronizing word* if there is a finite word $y_1 y_2 \cdots y_n \in \Sigma_n(Y)$ such that each element in $\phi^{-1}(y_1 y_2 \cdots y_n)$ admits the same terminal entry. More precisely, for any

$$x_1 x_2 \cdots x_m, x_1' x_2' \cdots x_m' \in \Sigma_m(X)$$

satisfying

$$\phi(x_1 x_2 \cdots x_m) = \phi(x_1' x_2' \cdots x_m') = y_1 y_2 \cdots y_n,$$

we have $x_m = x_m'$.

Suppose $\mathcal{G} = (G, \mathcal{L})$ is a labeled graph representation of the output space $\mathbf{Y}^{(2)}$ of (1). Denote by $W$ the SFT represented by the graph $G$ if $\mathcal{G}$ is right-resolving; otherwise, denote by $W$ the SFT represented by the graph $H$, where $\mathcal{H} = (H, \mathcal{L}')$ is obtained by applying SCM to $\mathcal{G}$. It follows that $W$ is a covering space of $\mathbf{Y}^{(2)}$ and there is a factor map $\phi : W \to \mathbf{Y}^{(2)}$ which is represented by the labeling $\mathcal{L}$ (or $\mathcal{L}'$). Theorem 2.4 asserts that the Hausdorff dimension of the output space $\dim \mathbf{Y}^{(2)}$ relates to the topological entropy of its cover-

ing space $h(W)$ if $\phi$ has a synchronizing word.

**Theorem 2.4.** *Along with the same assumption of Theorem* 2.3. *Let* $W$, *which is represented by* $G$ *if* $\mathcal{G}$ *is right-resolving and is represented by* $H$ *otherwise, be the covering space of* $\mathbf{Y}^{(2)}$. *Suppose the factor map* $\phi : W \to \mathbf{Y}^{(2)}$, *which is represented by the labeling* $\mathcal{L}$ (*or* $\mathcal{L}'$), *has a synchronizing word. Then*

$$\dim \mathbf{Y}^{(2)} = \begin{cases} \dfrac{\log \rho_T^2}{\log 2}, & \text{if } \mathcal{G} \text{ is right-resolving}; \\ \dfrac{\log \rho_H^2}{\log 2}, & \text{otherwise.} \end{cases} \tag{10}$$

*Restated*,

$$\dim \mathbf{Y}^{(2)} = \frac{2h(\mathbf{Y}^{(2)})}{\log 2}. \tag{11}$$

*Proof.* Suppose $X$ is a SFT and $\mu$ is an invariant probability measure on $X$. The Variational Principle indicates that the topological entropy of $X$ is the supremum of the measure-theoretic entropy of $X$; more precisely,

$$h(X) = \sup \{ h_\mu(X) : \mu \text{ is an invariant}$$
$$\text{probability measure on } X \}$$

A measure $\mu$ is called maximal if $h_\mu(X)$ attains the supremum. Let $\mu$ be a Markov measure which is derived from the transition matrix of $X$. Then $\mu$ is the unique measure that satisfies

$$h(X) = h_\mu(X)$$

if $X$ is topologically transitive (cf. [15]). Ban and Chang showed that, if $\phi : W \to \mathbf{Y}^{(2)}$ has a synchronizing word, then the Hausdorff dimension of the output space is

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

where $\mu^+ / \mu^-$ is a maximal measure of $W^+ / W^-$ (see [16], Theorem 2.6). Since $\mathcal{H}$ is right-resolving, the factor map $\phi : W \to \mathbf{Y}^{(2)}$ is finite-to-one. It follows that

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

Theorem 2.3 demonstrates that the topological entropy of the output space

$$h(\mathbf{Y}^{(2)}) = \log \rho_H$$

(respectively $h(\mathbf{Y}^{(2)}) = \log \rho_T$) if $\mathcal{G}$ is not right-resolving (respectively $\mathcal{G}$ is right-resolving). A straightforward examination infers that

$$h(W) = h(W^+) = h(W^-).$$

Hence we have

$$\dim \mathbf{Y}^{(2)} = \frac{2h\left(\mathbf{Y}^{(2)}\right)}{\log 2} = \begin{cases} \dfrac{\log \rho_T^2}{\log 2}, & \text{if } \mathcal{G} \text{ is right-resolving;} \\[3mm] \dfrac{\log \rho_H^2}{\log 2}, & \text{otherwise.} \end{cases}$$

this completes the proof.                                         □

## 3. Example

Suppose $\mathbb{T} = (A, B, z)$ with

$$A_1 = (2.2, 1.7), \quad A_2 = (-4, -2), \quad B = (-2.6, -1.4),$$

and $z = (-1.2, 0.3)$. The transition matrices for the first and second layer are

$$T_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and $T_2 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

respectively. Therefore, the transition matrix and the symbolic transition matrix of the MNN are

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

and $S = \begin{pmatrix} \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing \\ \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing \\ \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & a_{01} & \varnothing & \varnothing \\ \varnothing & \varnothing & \varnothing & \varnothing & a_{01} & a_{01} & \varnothing & \varnothing \\ \varnothing & \varnothing & a_{10} & \varnothing & \varnothing & a_{11} & \varnothing & \varnothing \\ \varnothing & \varnothing & a_{10} & \varnothing & \varnothing & a_{11} & \varnothing & \varnothing \\ \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing \\ \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing & \varnothing \end{pmatrix}$

respectively, where

$$a_{00} = --, \quad a_{01} = -+,$$
$$a_{10} = +-, \quad a_{11} = ++.$$

It is seen from the symbolic transition matrix $S$ that the labeled graph $\mathcal{G}$ is not right-resolving, and applying SMC to $\mathcal{G}$ derives a right-resolving labeled graph $\mathcal{H}$ (cf. **Figure 1**). The transition matrix of $\mathcal{H}$, indexed by $p_3, p_4, p_5, p_6, \{p_5, p_6\}$, is

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

Theorem 2.3 indicates that

$$h\left(\mathbf{Y}^{(2)}\right) = \log \rho_H = \log g,$$

where

$$g = \frac{1 + \sqrt{5}}{2}$$

is the golden mean.

The symbolic transition matrix of $\mathcal{H}$ is

$$S' = \begin{pmatrix} \varnothing & \varnothing & \varnothing & a_{01} & \varnothing \\ \varnothing & \varnothing & \varnothing & \varnothing & a_{01} \\ a_{10} & \varnothing & \varnothing & a_{11} & \varnothing \\ a_{10} & \varnothing & \varnothing & a_{11} & \varnothing \\ \varnothing & \varnothing & a_{10} & a_{11} & \varnothing \end{pmatrix}.$$

It is seen that both $a_{10} = +-$ and $a_{11} = ++$ are synchronizing words of $\phi$. Theorem 2.4 demonstrates that

$$\dim \mathbf{Y}^{(2)} = \frac{\log g^2}{\log 2} \approx 1.3885.$$

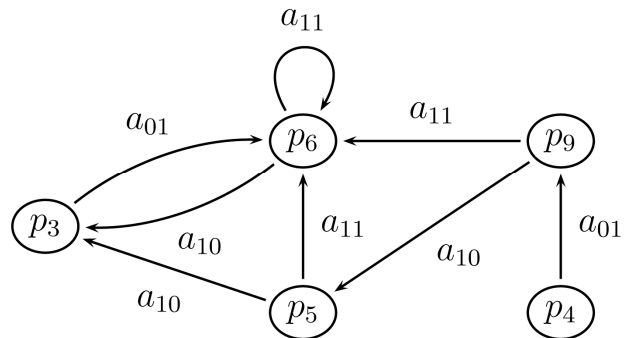The fractal set of the output space $\mathbf{Y}^{(2)}$ is seen in **Figure 2**.



**Figure 1. The right-resolving labeled graph $\mathcal{H}$ obtained by applying SCM to $\mathcal{G}$ in example. Here $p_9 = \{p_5, p_6\}$.**

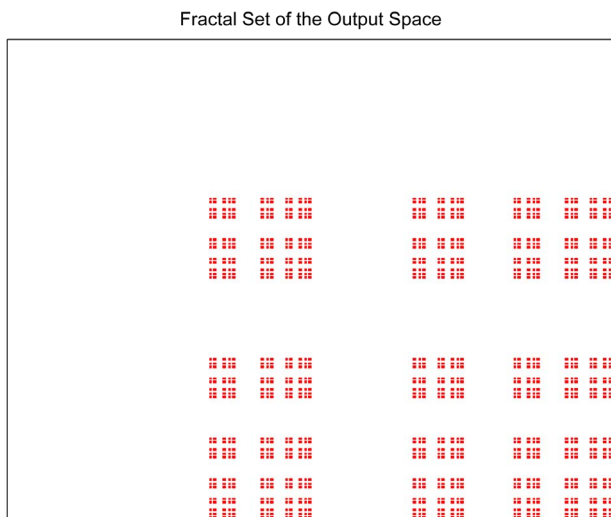Fractal Set of the Output Space



**Figure 2. The fractal set of the output space $Y^{(2)}$.**

## REFERENCES

[1] K. Hornik, M. Stinchcombe and H. White, "Multilayer Feedforward Networks Are Universal Approximators," *Neural Networks*, Vol. 2, No. 5, 1989, pp. 359-366. http://dx.doi.org/10.1016/0893-6080(89)90020-8

[2] B. Widrow and M. Lehr, "30 Years of Adaptive Neural Networks: Perceptron Madaline, and Backpropagation," *Proceedings of the IEEE*, Vol. 78, No. 9, 1990, pp. 1415-1442. http://dx.doi.org/10.1109/5.58323

[3] Y. A. Alsultanny and M. M. Aqul, "Pattern Recognition Using Multilayer Neural-Genetic Algorithm," *Neurocomputing*, Vol. 51, 2003, pp. 237-247. http://dx.doi.org/10.1016/S0925-2312(02)00619-7

[4] B. Widrow, "Layered Neural Nets for Pattern Recognition," *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, Vol. 36, No. 7, 1962, pp. 1109-1118. http://dx.doi.org/10.1109/29.1638

[5] J. J. Hopeld and D. W. Tank, "Neural Computation of Decisions in Optimization Problems," *Biological Cybernetics*, Vol. 52, No. 3, 1985, pp. 141-152.

[6] C. Peterson and B. Soderberg, "A New Method for Mapping Optimization Problems onto Neural Network," *International Journal of Neural Systems*, Vol. 1, 1989, pp. 3-22. http://dx.doi.org/10.1142/S0129065789000414

[7] J.-C. Ban and C.-H. Chang, "The Learning Problem of Multi-Layer Neural Networks," *Neural Networks*, Vol. 46, 2013, pp. 116-123. http://dx.doi.org/10.1016/j.neunet.2013.05.006

[8] J.-C. Ban, C.-H. Chang and S.-S. Lin, "The Structure of Multi-Layer Cellular Neural Networks," *Journal of Differential Equations*, Vol. 252, No. 8, 2012, pp. 4563-4597. http://dx.doi.org/10.1016/j.jde.2012.01.006

[9] 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*, Vol. 246, No. 2, 2009, pp. 552-580. http://dx.doi.org/10.1016/j.jde.2008.05.004

[10] 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*, Vol. 42, No. 10, 1995, pp. 746-756. http://dx.doi.org/10.1109/81.473583

[11] K. Falconer, "Fractal Geometry: Mathematical Foundations and Application," 2nd Edition, John Wilet & Sons, New York, London, Sydney, 2003. http://dx.doi.org/10.1002/0470013850

[12] Y. Pesin, "Dimension Theory in Dynamical Systems: Contemporary Views and Application," The University of Chicago Press, Chicago, 1997. http://dx.doi.org/10.7208/chicago/9780226662237.001.0001

[13] J. Juang and S.-S. Lin, "Cellular Neural Networks: Mosaic Pattern and Spatial Chaos," *SIAM Journal on Applied Mathematics*, Vol. 60, No. 3, 2000, pp. 891-915. http://dx.doi.org/10.1137/S0036139997323607

[14] D. Lind and B. Marcus, "An Introduction to Symbolic Dynamics and Coding," Cambridge University Press, Cambridge, 1995. http://dx.doi.org/10.1017/CBO9780511626302

[15] B. Kitchens, "Symbolic Dynamics. One-Sided, Two-Sided and Countable State Markov Shifts," Springer-Verlag, New York, 1998.

[16] J.-C. Ban and C.-H. Chang, "On the Structure of Multi-Layer Cellular Neural Networks. Part II: The Complexity between Two Layers," Submitted, 2012.