Theoretically Provable Spiking Neural Networks

Shao-Qun Zhang **Zhi-Hua Zhou***

National Key Laboratory for Novel Software Technology
Nanjing University, Nanjing 210093, China
{zhangsq,zhouzh}@lamda.nju.edu.cn

Abstract

Spiking neural networks have attracted increasing attention in recent years due to their potential of handling time-dependent data. Many algorithms and techniques have been developed; however, theoretical understandings of many aspects of spiking neural networks are far from clear. A recent work [44] disclosed that typical spiking neural networks could hardly work on spatio-temporal data due to their bifurcation dynamics and suggested that the *self-connection* structure has to be added. In this paper, we theoretically investigate the approximation ability and computational efficiency of spiking neural networks with self connections, and show that the self-connection structure enables spiking neural networks to approximate discrete dynamical systems using a polynomial number of parameters within polynomial time complexities. Our theoretical results may shed some insight for the future studies of spiking neural networks.

1 Introduction

The past decades have witnessed an increasing interest in spiking neural networks (SNNs) due to their great potential of modeling time-dependent data [28, 32, 37]. The fundamental units of SNNs are usually formulated as the combination of an integrated process (e.g., some first-order parabolic equations) and a step firing function. There has been significant progress on computational and implementation techniques for SNNs in computer vision [13, 33], speech recognition [30, 39], reinforcement learning [11, 38], few-short learning [16, 22, 25], etc. However, theoretical understandings of many aspects of SNNs, such as the approximation ability and computational efficiency on spatio-temporal systems, are far from clear.

Some researchers [19, 20, 21, 31] focused on the approximation universality of SNNs, in which some typical SNNs can simulate the standard computational models such as Turing machines, random access machines, threshold circuits, sequence-to-sequence mapping, etc. There are also efforts on the computational efficiency of SNNs for some specific issues, such as the convergence in the limit results and computational complexity of SNNs for the sparse coding problem [34, 35] and temporal quadratic programming [5, 8], respectively.

Amazingly, a recent study [44] theoretically proved that, contrary to previous beliefs, typical SNNs can hardly work well on spatio-temporal data, because they in nature are bifurcation dynamical systems with fixed eigenvalues in which many patterns inherently cannot be learned. They also suggested that adding *self-connection* structure can enhance the representation ability of SNNs on spatio-temporal systems that fully connect the spiking neurons in the same layer and solves adaptive eigenvalues of discrete dynamical systems.

*Zhi-Hua Zhou is the corresponding author.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

In this paper, we theoretically investigate the approximation ability and computational efficiency of the self-connection spike neural networks (scSNNs). Our theoretical results show that equipped with self connections, scSNNs can approximate discrete dynamical systems using polynomial number of parameters within polynomial time complexities. Our main contributions are summarized as follows:

- We prove that the proposed scSNNs are universal approximators in Theorem 1.
- As for spatial approximation, we prove that a broad range of radial functions can be well approximated by scSNNs with polynomial spiking neurons in Theorem 2.
- As for temporal approximation, we prove that multivariate spike flows can be approximated by scSNNs within polynomial time in Theorem 3 and verify this conclusion in simulation experiments.

The rest of this paper is organized as follows. Section 2 introduces notations. Section 3 presents the scSNNs and some related concepts. Section 4 provides three key theorems to show the universal approximation ability and (parameters and time) complexity of scSNNs for approximating the discrete dynamical system. Section 5 develops in-depth discussions of the effects led by self-connection. Section 6 concludes this work.

2 Notations

Here, we provide some useful notations, which are detailed in Appendix. Let $[N] = \{1, 2, \ldots, N\}$ be an integer set for $N \in \mathbb{N}^+$, and $|\cdot|_\#$ denotes the number of elements in a collection, e.g., $|[N]|_\# = N$. Let $\mathrm{i} = \sqrt{-1}$ be the imaginary unit, and $\boldsymbol{x} \preccurlyeq 0$ means that every element $x_i \leq 0$ for any i. Let the sphere $\mathcal{S}(r)$ and globe $\mathcal{B}(r)$ be $\mathcal{S}(r) = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| = r\}$ and $\mathcal{B}(r) = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| \leq r\}$ for any $r \in \mathbb{R}$, respectively. Given a function $g(n)$, we denote by $h_1(n) = \Theta(g(n))$ if there exist positive constants c_1, c_2 and n_0 such that $c_1 g(n) \leq h_1(n) \leq c_2 g(n)$ for every $n \geq n_0$; $h_2(n) = \mathcal{O}(g(n))$ if there exist positive constants c and n_0 such that $h_2(n) \leq cg(n)$ for every $n \geq n_0$; $h_3(n) = \Omega(g(n))$ if there exist positive constants c and n_0 such that $h_3(n) \geq cg(n)$ for every $n \geq n_0$; $h_4(n) = o(g(n))$ if there exists positive constant n_0 such that $h_4(n) < cg(n)$ for every $c > 0$ and $n \geq n_0$.

Let $\mathcal{C}(K, \mathbb{R})$ be the set of all scalar functions $f : K \to \mathbb{R}$ continuous on $K \subset \mathbb{R}^n$. Given $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)^\top \in \mathbb{N}^m$, we define

$$D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) = \frac{\partial^{\alpha_1}}{\partial x^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x^{\alpha_2}} \cdots \frac{\partial^{\alpha_m}}{\partial x^{\alpha_m}} f(\boldsymbol{x}) \,,$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in K$. Further, we define

$$\mathcal{C}^l(K, \mathbb{R}) = \{f \mid f \in \mathcal{C}(K, \mathbb{R}) \text{ and } D^r f \in \mathcal{C}(K, \mathbb{R}), \ \text{ for } \ r \in [l]\} \ .$$

For $1 \leq p < \infty$, we define

$$\mathcal{L}^p(K, \mathbb{R}) = \left\{ f \;\middle|\; f \in \mathcal{C}(K, \mathbb{R}) \text{ and } \|f\|_{p,K} \triangleq \left(\int_K |f(\boldsymbol{x})|^p \, \mathrm{d}\boldsymbol{x} \right)^{1/p} < \infty \right\} \ .$$

This work considers the Sobolev space $\mathcal{W}_\mu^{l,p}(K, \mathbb{R})$, defined as the collection of all functions $f \in \mathcal{C}^l(K, \mathbb{R})$ and $D^r f \in \mathcal{L}^p(K, \mathbb{R})$ for all $|\boldsymbol{\alpha}| \in [l]$, that is,

$$\|D^{\boldsymbol{\alpha}} f\|_{p,K} = \left(\int_K |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|^p \, \mathrm{d}\boldsymbol{x} \right)^{1/p} < \infty \ .$$

This paper employs $\mathbf{E}_n$ to denote the $n \times n$ unit matrix and $\det(\cdot)$ to indicate the determinant operation on the matrix. Two n-by-n matrices $\mathbf{A}$ and $\mathbf{B}$ are called *similar*, denoted as $\mathbf{A} \sim \mathbf{B}$, if there exists an invertible $n \times n$ matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. The general linear group over field $\mathbb{F}$, denoted as $\mathbf{GL}(n, \mathbb{F})$, is the set of $n \times n$ invertible matrices with entries in $\mathbb{F}$. Especially, we define that a special linear group $\mathbf{SL}(n, \mathbb{F})$ is the subgroup of $\mathbf{GL}(n, \mathbb{F})$ and consists of matrices with determinant 1. For any field $\mathbb{F}$, the $n \times n$ orthogonal matrices form the following subgroup

$$\mathbf{O}(n, \mathbb{F}) = \{\mathbf{P} \in \mathbf{GL}(n, \mathbb{F}) \mid \mathbf{P}^\top \mathbf{P} = \mathbf{P}\mathbf{P}^\top = \mathbf{E}_n\}$$

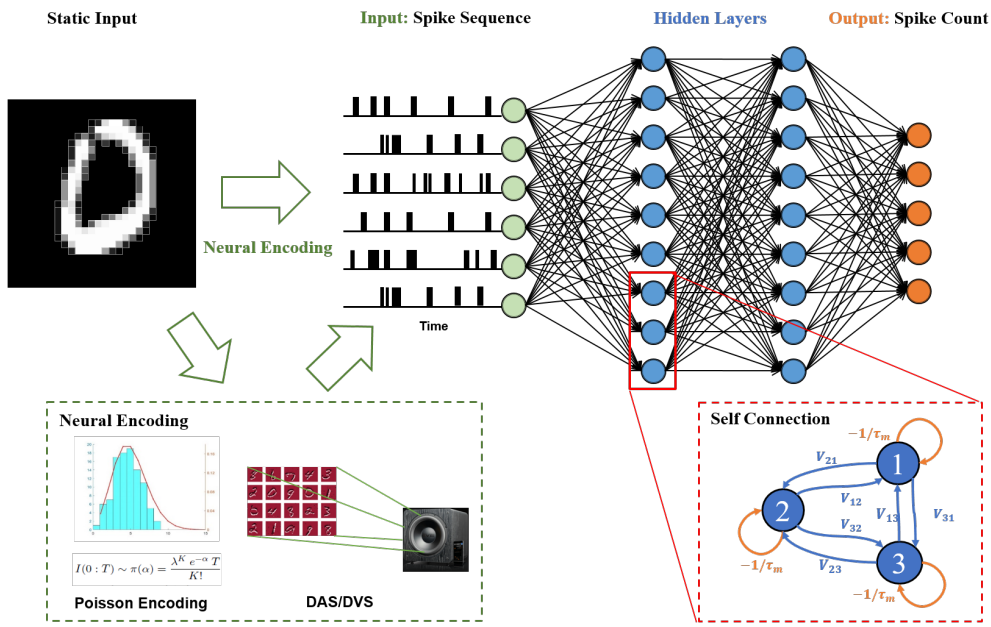


Figure 1: Illustrations of scSNNs and Self Connections.

of the general linear group $\mathbf{GL}(n, \mathbb{F})$. Similarly, we have the special orthogonal group, denoted as $\mathbf{SO}(n, \mathbb{F})$, which consists of all orthogonal matrices of determinant 1 and is a normal subgroup of $\mathbf{O}(n, \mathbb{F})$. This group is also called the rotation group, generalizing that linearly transforms geometries while holding the surface orientation.

Let $\phi^2(\boldsymbol{x})$ be the density function of some probability measure μ, which satisfies

$$\int_{\boldsymbol{x}\in\mathbb{R}^m} \phi^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{x}\in\mathcal{B}(1)} 1\,\mathrm{d}\boldsymbol{x} = 1\,.$$

For continuous functions f and g, we have the following equalities under Fourier transform

$$\|\widehat{f\phi} - \widehat{g\phi}\|_{L_2(\mu)} = \|f\phi - g\phi\|_{L_2(\mu)} \quad \text{and} \quad \widehat{f\phi} = \hat{f} * \hat{\phi}\,,$$

for the convolution operator $*$.

3 Self-connection Spiking Neural Networks

Let n be the number of spiking neurons, $\boldsymbol{u}(t) \in \mathbb{R}^n$ and $\boldsymbol{s}(t) \in \mathbb{R}^n$ denote the vectors of membrane potentials and spikes in which $\boldsymbol{u}_i(t)$ and $\boldsymbol{s}_i(t)$ are the membrane potential and spikes of neuron $i \in [n]$ at time $t \geq 0$, respectively. Inheriting the recognition in [44], we here consider the self-connection SNN (scSNN) as follows:

$$\frac{\mathrm{d}\boldsymbol{u}_i(t)}{\mathrm{d}t} = -\frac{1}{\tau_m}\boldsymbol{u}_i(t) + \sum_{j\in[n]} \mathbf{V}_{ij}\boldsymbol{s}_j(t) + \sum_{k\in[m]} \mathbf{W}_{ik}\mathbf{I}_k(t)\,, \tag{1}$$

where τ_m is a positive-valued hyper-parameter with respect to membrane time, $\mathbf{I}_k(t)$ indicates the signal from input channel $k \in [m]$ at time t, $\mathbf{V} \in \mathbb{R}^{n\times n}$ and $\mathbf{W} \in \mathbb{R}^{n\times m}$ denote the self-connection and connection weights matrices, respectively. Hence, $\sum_{k\in[m]} \mathbf{W}_{ik}\mathbf{I}_k(t)$ indicates the signal received by neuron i at time t, and $\sum_{j\in[n]} \mathbf{V}_{ij}\boldsymbol{s}_j(t)$ denotes the effect on the membrane potential of neuron i when neuron j fires a spike, as illustrated in Figure 1. In general, we force the self-connection matrix $\mathbf{V}$ to be symmetric.

Based on the *spike response model* scheme [12], Eq. (1) has the following solution with the boundary condition $u_{rest} = 0$

$$\boldsymbol{u}_i(t) = \int_{t'}^{t} \exp\left(\frac{s-t'}{\tau_m}\right)\left(\sum_{j\in[n]} \mathbf{V}_{ij}\boldsymbol{s}_j(t) + \sum_{k\in[m]} \mathbf{W}_{ik}\mathbf{I}_k(t)\right)\mathrm{d}s\,, \tag{2}$$

where t' denotes the last firing time $t' = \max\{s \mid \boldsymbol{u}_i(s) = u_{firing},\ s < t\}$ for a pre-given firing threshold $u_{firing} > 0$. Spiking neuron model employs the typical threshold rule, that is, neuron i fires spikes $\boldsymbol{s}_i(t)$ at time t if and only if $\boldsymbol{u}_i(t) \geq u_{firing}$. After firing, the membrane potential is instantaneously reset to a lower value u_{rest} (rest voltage). Note that this work does not consider using absolute refractory periods [14] or refractory kernels [9].

Let the timing set $\mathbb{T}_i = \{t \mid \boldsymbol{u}_i(t) = u_{firing}, t \in [0, T]\}$ record the firing times of neuron i and $N_i = |\mathbb{T}_i|_{\#}$. In general, we consider using the *firing rate* $f_i^{\text{ave}}(T) = N_i/T$ to characterize the spike dynamics, where $f_i^{\text{ave}}(T)$ indicates the average number of spikes of neuron i at time interval $[0, T]$. It is clearly observed that $f_i^{\text{ave}}(t)$ is discontinuous since N_i is a step function regarding $\boldsymbol{u}_i(t)$ and time t. To ensure the well-posed characteristics of the firing rate functions, we here employ the spike excitation function $f_e : u \to s$ to smooth the firing procedure, for example,

$$\begin{cases} \text{Linear:} \quad f_e(\boldsymbol{u}_i(t)) \triangleq \dfrac{\boldsymbol{u}_i(t)}{u_{firing}}, \\ \text{Cos-based:} \quad f_e(\boldsymbol{u}_i(t)) \triangleq \left[1 - \cos\left(\dfrac{\pi}{2}\dfrac{t - t^{(k)}}{t^{(k+1)} - t^{(k)}}\right)\right] \left\lfloor \dfrac{\boldsymbol{u}_i(t)}{u_{firing}} \right\rfloor, \end{cases} \tag{3}$$

where $t^{(k)}$ and $t^{(k+1)}$ are two adjacent timings from $\mathbb{T}_i$ in which $k \in \mathbb{N}^+$ and $t^{(k)} < t^{(k+1)}$. Similar smooth treatments on spike excitation functions can refer to [13, 33]. Thus, $f_i^{\text{ave}}(T)$ can be approximated by an **Instantaneous Firing Rate** (IFR) function

$$f_i(\boldsymbol{x}, t) = f_e(\boldsymbol{u}_i(t))/(t - t') \ , \tag{4}$$

where $\boldsymbol{x} \in \mathbb{R}^m$, $t' \in \mathbb{T}_i$, and $t \in [T]$. It is observed that the proposed IFR function $f_i(\boldsymbol{x}, t)$ induces a discrete dynamical system, in which $f(\cdot, t) : \mathbb{R}^m \to \mathbb{R}$ is a Spatial function and $f(\boldsymbol{x}, \cdot) : \mathbb{R} \to \mathbb{R}$ is a temporal flow.

About Neural Encoding. The actual input data (e.g., image or video) should usually be pre-converted into a spiking version before fed up to SNNs. The conversion procedure is called *neural encoding*, as shown in Figure 1. There are two main categories of neural encoding approaches: *temporal encoding* and *rate-based encoding*; the former encodes input data by exploiting the distance between time instances that fire spikes [40], and the latter encodes input data as a count sequence of the fired spikes within temporal windows [18]. The rate-based encoding is the simplest and most popular scheme in SNNs. The representative techniques are usually encoded by a Poisson distribution or recorded by a dynamic vision sensor [3, 27]. Recent years have witnessed a lot of efforts on the information capacity of neural encoding, specially rate-based encoding, from empirical [6, 17] and theoretical [24, 35, 36] sides. Throughout this paper, we adopt rate-based encoding as the default and focus on the firing rates of SNNs, generalizing the computational powers concerning the spike count.

About Firing Rates. When we investigate the dynamics and neural computation of SNNs, the *firing rates* or equally the number of firing spikes are the key measure of network activities for investigating neural computation and model dynamics because of the close relation between firing rates and network function (including neural input, connectivity, spiking function, and firing process) [1, 2]. There are great efforts to use firing rates in SNNs for some real-world tasks, such as vision [13, 33, 44] and speech recognition [30, 39]. Besides, Barrett et al. [5] and Chou et al. [8] showed that the averaged firing rate can approximate the optimal solutions of some quadratic programs within polynomial complexity. This work employs an "instantaneous" firing rate rather than the averaged firing rate or the total number of firing spikes used in previous studies. This manner provides a feasible way to construct the discrete dynamical systems using the IFR functions, based on which we can develop in-depth understandings of SNNs from spatial and temporal aspects.

4 Approximation to Discrete Dynamical Systems

In this section, we first present the universal approximation for scSNNs, and then show the parameters and time complexities of the IFR functions led by scSNNs for approximating discrete dynamical systems from spatial and temporal aspects, respectively.

4.1 Universal Approximation

Now, we present our first theorem about the universal approximation of scSNNs as follows:

Definition 1 *Let $l \in \mathbb{N}^+$. The function $f(x)$ is said to be **l-finite** if f is an l-times differentiable scalar function that satisfies*

$$0 < \left| \int_{\mathbb{R}} D^l f(x) \, \mathrm{d}x \right| < \infty \, .$$

Theorem 1 *Let $K \subset \mathbb{R}^m$ be a compact set. If the spike excitation function f_e is l-finite and $w_i \in \mathbb{R}$ where $i \in [n]$, then for all $r \in [l]$, there exists some time t such that the set of IFR functions $f(\cdot, t) : K \to \mathbb{R}$ of the form $f(\boldsymbol{x}, t) = \sum_{i \in [n]} w_i f_i(\boldsymbol{x}, t)$ is dense in $\mathcal{C}^r(K, \mathbb{R})$.*

Theorem 1 shows that the scSNN is a universal approximator, which provides a solid cornerstone for developing SNNs' theory.

The proof idea of this theorem can be summarized as follows. We utilize the invertibility of the Fourier transform on Sobolev space $\mathcal{W}_{\mu}^{l,p}(K, \mathbb{R})$ $(p > 1)$, to project the concerned functional space $\mathcal{C}^r(K, \mathbb{R})$ into a characteristic space, and the corresponding objective function is transformed as a single integral over the characteristic space. According to Fubini's theorem, the approximation problem on $\mathcal{C}^r(K, \mathbb{R})$ can be converted into another that uses multiple integrals to a single integral on the characteristic space. The subsequent proof can then be completed along the thought lines of the technical exposition given by Carslaw and Rogosinski [7]. The full proof of Theorem 1 can be obtained in Appendix A.

4.2 ϵ-Approximation with Parameters Complexity

This subsection presents our second theorem about the parameters complexity of scSNNs as follows:

Definition 2 *We say that g is a radial function if $g(\boldsymbol{x}') = g(\boldsymbol{x})$ for any $\|\boldsymbol{x}'\| = \|\boldsymbol{x}\|$.*

Theorem 2 *Given a compact set $K^m \subset \mathbb{R}^m$, a probability measure μ, and a radial function $g : K^m \to \mathbb{R}$. For some apposite spike excitation function f_e and any $\epsilon > 0$, there exists some time t such that the radial function g can be well approximated by a one-hidden-layer scSNN of $\mathcal{O}(Cm^{15/4})$ spiking neurons, that is,*

$$\|f(\boldsymbol{x}, t) - g(\boldsymbol{x})\|_{L^2(\mu)} < \epsilon \, .$$

Theorem 2 shows that the scSNNs with polynomial spiking neurons can well approximate a broad scope of radial functions, which may shed some insights on that scSNNs admit input rotations [36], such as the rotation-based data argument techniques [29] and invariant models [23].

This paper provides two ways for proving this theorem. Here, we only introduce the proof idea of the interesting one (full proof is shown in Appendix B), and another proof way can be accessed in Appendix C. This proof idea can be summarized as follows. The radial function is invariant to rotations and dependent on the input norm, corresponding to the phase and norm (i.e., radial) of inputs, respectively. It is observed that $\boldsymbol{s}(t)$ in Eq. (1) induces a local recurrent function concerning the input $\boldsymbol{x}(t)$. Thus, there exists a collection of parameters such that the IFR function is dependent on the norm of inputs. Besides, the IFR function admits the phase since $\mathbf{V}$ is a symmetric matrix. Thus, there exist some linear connections (including rotation transformations) such that the weighted aggregation of these spiking neurons is invariant to rotations. Summing up the approximation recognition, there is a family of radial functions that can be well approximated by one-hidden-layer scSNNs.

We formally begin our proof of Theorem 2 with some useful lemmas.

Lemma 1 *Let $g : [r, R] \to \mathbb{R}$ be an L-Lipschitz function for $r \leq R$. For any $\delta > 0$, $C_s \geq 1$, and $n \leq C_s R^2 L m / (\sqrt{r}\delta)$, there exist some time t and an IFR function $f(\boldsymbol{x}, t)$ led by a one-hidden-layer scSNN of n spiking neurons such that*

$$\sup_{\boldsymbol{x} \in \mathbb{R}^m} |g(\|\boldsymbol{x}\|) - f(\boldsymbol{x}, t)| \leq \delta \, .$$

Lemma 2 *For $m > C_2 > 0$, $g : \mathbb{R}^m \to \mathbb{R}$ is an L-Lipschitz radial function supported on the set*

$$\mathcal{S}_{\Delta} = \{\boldsymbol{x} : 0 < C_2\sqrt{m} \leq \|\boldsymbol{x}\| \leq 2C_2\sqrt{m}\} \, .$$

For any $\delta > 0$, there exist some time t and an IFR function $f(\boldsymbol{x}, t)$ led by scSNNs of one-hidden layer with width at most $C_s(C_2)^{3/2}L(m)^{7/4}/\delta$ such that

$$\sup_{\boldsymbol{x}\in\mathbb{R}^m} |g(\boldsymbol{x}) - f(\boldsymbol{x}, t)| < \delta \ .$$

Lemma 2 shows that the L-Lipschitz radial functions can be approximated by the IFR function $f(\boldsymbol{x}, t)$ led by scSNNs of one-hidden layer with polynomial parameters.

Lemma 3 *Define*

$$g(\boldsymbol{x}) = \sum_{i=1}^{N} \epsilon_i g_i(\boldsymbol{x}) \quad \textit{with} \quad g_i(\boldsymbol{x}) = \mathbb{I}\{\|\boldsymbol{x}\| \in \Omega_i\} \ , \tag{5}$$

where $\epsilon_i \in \{-1, +1\}$, N is a polynomial function of m, and Ω_i's are disjoint intervals of width $\mathcal{O}(1/N)$ on values in the range $\Theta(\sqrt{m})$. For any $\epsilon_i \in \{-1, +1\}$ $(i \in [N])$, there exists a Lipschitz function $h\colon \mathcal{S}_\Delta \to [-1, +1]$ such that

$$\int_{\mathbb{R}^m} (g(\boldsymbol{x}) - h(\boldsymbol{x}))^2 \, \phi^2(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \leq \frac{3}{(C_2)^2\sqrt{m}} \ .$$

Lemma 3 shows that any non-Lipschitz function $h(\boldsymbol{x})$ can be approximated and bounded by a Lipschitz function with density ϕ^2.

Proof of Theorem 2. Let $g(\boldsymbol{x}) = \sum_{i=1}^{N} \epsilon_i g_i(\boldsymbol{x})$ be defined by Eq. (5) and $N \geq 4C_2^{5/2}m^2$. According to Lemma 3, there exists a Lipschitz function h with range $[-1, +1]$ such that

$$\|h(\boldsymbol{x}) - g(\boldsymbol{x})\|_{L_2(\mu)} \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} \ .$$

Based on Lemmas 1 and 2, any Lipschitz radial function supported on $\mathcal{S}_\Delta$ can be approximated by an IFR function $f(\boldsymbol{x}, t)$ led by scSNNs of one-hidden layer with width at most $C_3C_s(m)^{15/4}$, where C_3 is a constant relative to C_2 and δ. This means that there exists some time t such that

$$\sup_{\boldsymbol{x}\in\mathbb{R}^m} |h(\boldsymbol{x}) - f(\boldsymbol{x}, t)| \leq \delta \ .$$

Thus, we have

$$\|h(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \delta \ .$$

Hence, the range of $f(\cdot, t)$ is in $[-1 - \delta, +1 + \delta] \subseteq [-2, +2]$. Provided the radial function, defined by Eq. (5), we have

$$\|g(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \|g(\boldsymbol{x}) - h(\boldsymbol{x})\|_{L_2(\mu)} + \|h(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} + \delta \ .$$

This implies that given constants $m > C_2 > 0$ and $C_3 > 0$, for any $\delta > 0$ and $\epsilon_i \in \{-1, +1\}$ $(i \in [N])$, there exists some time t, such that the target radial function g can be approximated by an IFR function $f(\boldsymbol{x}, t)$ led by scSNNs of one-hidden layer with range in $[-2, +2]$ and width at most $C_3C_s(m)^{15/4}$, that is,

$$\|g(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} + \delta < \delta_1 \ .$$

This completes the proof. □

4.3 ϵ-Approximation with Time Complexity

The IFR function generates a discrete dynamical system, comprising a Spatial function and a temporal flow. Both Theorems 1 and 2 focus on the Spatial (e.g., spatial) approximation characteristics of the IFR function. This subsection investigates its temporal characteristics, especially how long it takes for a self-connected SNN to achieve a specified task or target function. Now, we present our third theorem about the time complexity of scSNNs as follows.

Definition 3 *A matrix $\mathbf{V} \in \mathbb{R}^{m\times n}$ (if $m \leq n$) is said to be* ***non-degenerate*** *if any $m \times m$ sub-matrix of $\mathbf{V}$ has full rank.*

Theorem 3 *Let $n \geq m$. Let $\boldsymbol{x}_k(t) \in \pi(\lambda_k)$ and $\mathbb{E}(\boldsymbol{x}_k) = \lambda_k$ for $\lambda_k > 0$, $k \in [m]$, and $t \in [T]$. If $\mathbf{V}$ is a non-degenerate matrix, then for any $\epsilon > 0$ and matrix $\mathbf{G} \in \mathbb{R}^{m\times n}$ with $\|\mathbf{G}\|_2 < \infty$, when the time complexity satisfies*

$$T \geq \Omega\left(\frac{\sqrt{n}\,\|\mathbf{G}\|_2}{\epsilon\sqrt{\|\mathbf{V}\|_2}}\right),$$

there exists some one-hidden-layer scSNNs with the IFR vector $\boldsymbol{f} = (f_1, \ldots, f_n)^\top$ such that

$$\|\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{V}\boldsymbol{f}(\boldsymbol{x}, T)\right] - \mathbf{G}\boldsymbol{\lambda}\|_2 < \epsilon\,. \tag{6}$$

Theorem 3 shows that without training, the scSNNs can well approximate the multivariate spike flows (MSF, i.e., linear functions of $\boldsymbol{\lambda}$) within an explicit polynomial time complexity. In contrast to the conventional ANN theory and our proposed theorems above that show the spatial approximation ability, this theorem portrays the computational efficiency of a flow function led by scSNNs. Besides, there are few studies on the computational efficiency of SNNs, and we summarize the comparative results in Table 1. Notice that approximating specific functions, such as MSF in this work, is more challenging than solving some problems with locally competitive algorithms [34]. Thus, it is observed that Theorem 3 provides a rigorous guarantee for SNNs to solve some algorithmic problems. Besides, Theorem 3 relaxes the dimension of the self-connection matrix $\mathbf{V}$ from $n \times n$ to $n \times m$, which provides a more general result for approximating MSF with any dimension.

Table 1: Comparative Results on Computational Efficiency of SNNs.

Works	**Models**	**Objects**	**Computational Complexity**
Tang et al. [35]	configured SNNs	solving Sparse Coding Problem	Convergence
Chou et al. [8]	simple SNNs	solving Quadratic Programming	Polynomial Complexity
Our Work	scSNNs with Poisson inputs	approximating MSF	Polynomial Complexity

The proof idea of this theorem originates from the attractor theory in dynamical systems. One key lemma is as follows.

Lemma 4 *Given $\mathbf{U} = \sqrt{\boldsymbol{\Lambda}_U}\ \widetilde{\mathbf{U}}^\top, \tilde{\boldsymbol{\lambda}} \in \mathbb{R}^n$, and $\epsilon > 0$, when $t \geq \Omega\left(\frac{\sqrt{n}\,\|\Lambda_G\|_2}{\epsilon\,\|\mathbf{U}\|_2}\right)$, it holds $\|\mathbf{U}\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\lambda}}, t) - \tilde{\boldsymbol{\lambda}}\| \leq \epsilon$, where the modified IFR function $\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\lambda}}, t)$ is led by a scSNN without connection matrix $\mathbf{W}$ and fed up to the constant spike sequence $\tilde{\boldsymbol{\lambda}}$ at every timestamp.*

Lemma 4 shows that when fed up to a constant spike sequence $\tilde{\boldsymbol{\lambda}}$ at every timestamp, a weighted sum of IFR functions can converge to an attractor around the inputs within a polynomial temporal computation. Based on the results of Lemma 4, it suffices to prove that the expectation of the concerned IFR function $\boldsymbol{f}(\boldsymbol{x}, t)$ can approximate $\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\lambda}}, t^*)$ within time interval $[0, T]$ where $t^* \leq T \leq \Omega(\frac{\sqrt{n}\,\|\mathbf{G}\|_2}{\epsilon\sqrt{\|\mathbf{V}\|_2}})$. Full proof of Theorem 3 can be obtained in Appendix D.

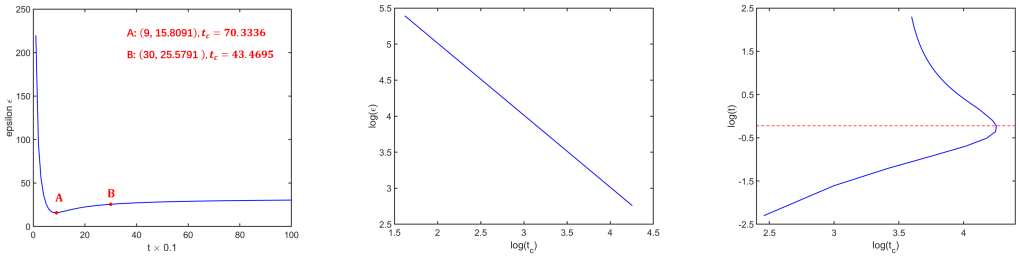
The aforementioned results, time complexity bound in detail, can be verified by a simulated experiment. We here simulate a $4\times 10,000$ spike sequence from the Iris data sets with a timestamp of 0.001 using Poisson encoding. We employ the one-hidden-layer scSNN [44] that contains self-connection structure as the conducted SNN model. The number of input channels and hidden neurons are 4 and 10, respectively. The self-connection matrix $\mathbf{V}$ is randomly sampled from $[0, 1]$ with bias $= 1/3$. The above configurations meet the conditions of Theorem 3. Table 2 lists the hyper-parameter values in the conducted scSNN. For any linear matrix $\mathbf{G} \in \mathbb{R}^{4\times 10}$ with $\|\mathbf{G}\|_2 < \infty$, we define an indicator $t_c = \frac{\sqrt{n}\,\|\mathbf{G}\|_2}{\epsilon\sqrt{\|\mathbf{V}\|_2}}$. Thus, by exploiting the relation among ϵ, t, and t_c, we can verify the explicit polynomial bound, especially the order of magnitude function $\Omega(\cdot)$ in Theorem 3.

Figure 2 plots the experimental results. From Figure 2(a), the conducted scSNN first approximates the objective function at a faster rate and then maintains a lower approximation error. Figure 2(b) signifies that $\log(\epsilon)$ is inversely proportional to $\log(t_c)$, i.e., $a\log(t_c) + b\log(\epsilon) = c$ where $a, b, c > 0$. Figure 2(c) shows the relation plots between $\log(t_c)$ and $\log(t)$. Notice that the conducted ScNN

Table 2: Hyper-parameter Setting of scSNNs.

Parameters	Value	Parameters	Value
Time Step	0.001	Firing Threshold	1
Expect Spike Count (True)	100	Membrane Time τ_m	0.2
Expect Spike Count (False)	10	Time Constant of Synapse τ_s	0.008
Encoding Length T	10, 000	Maximum Firing	10
Refractory Period	0.016	Maximum Time	10

can achieve the same approximation error at different timestamps, we only care about the shortest one, that is, the blue solid curve below the red dashed line. It is observed that $\log(t)$ is proportional to $\log(t_c)$, i.e., $\log(t) = a\log(t_c) + b$ where $a > 0$ and $b \in \mathbb{R}$. In summary, we have demonstrated the effectiveness of our theoretical results.



(a) Curve: t-ϵ. (b) Curve: $\log(t_c)$-$\log(\epsilon)$. (c) Curve: $\log(t_c)$-$\log(t)$.

Figure 2: Magnitude Curves of ϵ, t, and t_c.

5 Discussions

Despite an increasing focus on the potential of handling spatio-temporal data, the theoretical understandings of many aspects (e.g., approximation ability and computational efficiency) of SNNs are still far from clear. Some seminal studies focused on the universality of SNNs. Maass et al. [19, 20, 21] showed that some typical SNNs can simulate the standard computational models such as Turing machines, random access machines, threshold circuits, etc. She et al. [31] showed some universal approximation properties of SNNs by exploiting the spike propagation paths. However, to the best of our knowledge, few theoretical guarantees on the approximation complexity and computational efficiency of SNNs have been provided. There are only some academic studies on the convergence in the limit results for SNNs solving the sparse coding problem [34, 35] and the convergence rates for SNNs solving temporal quadratic programming [8]. Recently, two studies [41] and [44] provided calculable ways for approaching the lower and upper bounds of adaptive SNN systems, respectively.

Our starting point is a recent advance proposed by Zhang and Zhou [44], in which they showed that adding self connections enables SNNs to achieve adaptive dynamical systems with spatio-temporal representation. This work developed an in-depth analysis of approximation ability and computational efficiency of self connections in SNNs. It is observed that self connections facilitate our results; re-using firing spike variables $\boldsymbol{s}(t)$ contributes to approximating the norm of input sequences $\boldsymbol{x}(t)$, and the symmetric self-connection matrix coincides with rotation and dual transformations, which are crucial in the proofs of Theorems 2 and 3. Our theoretical derivatives not only show the universal approximation properties of scSNNs but also disclose the parameters and time complexities for scSNNs approximating discrete dynamical systems, which prospectively provide solid support for the theory development of SNNs. Unfortunately, this paper has not yet shown the theoretical advantages of scSNNs over non-self-connection SNNs (it is valuable to be studied in the future). But we believe that the current results have disclosed the importance and power of self connections for enhancing the approximation and computational ability of SNNs, which may shed some insights on developing provable and sound SNNs.

In light of the preceding merits, many issues are worthy of being studied in the future. One important future issue is to explore some practical techniques for scSNNs. For example, Theorem 2 shows the power of scSNNs on representing radial or equally rotation-invariant functions. So we conjecture that adding self connections may be more compatible with some invariant models and data augmentation techniques, such as image rotation [23, 29]. Besides, the current implementation roughly follows the ideas of Zhang and Zhou [44], i.e., the effect from the k-th neuron to the i-th neuron equals to the last spike of neuron k multiplied by a weighted factor as shown in Eq. (1). Such a self-connection graph enable SNNs to maintain adaptive characteristics for representing discrete dynamical systems. However, it inevitably leads to a larger memory consumption when the input spike sequences are high-dimensional and high-frequency. Therefore, it is prospective to explore some more practical techniques or modules for scSNNs, which may open up possibilities for achieving sound SNNs.

Another important future issue is to develop in-depth theoretical understandings of SNNs, from aspects of approximation complexity, computational efficiency, representation ability [15, 26, 43], and over-parameterized architectures [4, 45]. Furthermore, it is interesting to explore the theoretical advantages of SNNs with self connections over SNNs without ones, especially from the perspectives of approximation, optimization, and generalization.

6 Conclusions

In this paper, we present the theoretical understandings of the approximation ability and computational efficiency of the self-connection SNNs, i.e., scSNNs. We provided three main theorems to show the universal approximation properties, parameters complexity for spatial approximation, and time complexity for temporal approximation of the scSNNs. Our theoretical results disclose the effects of self connections of scSNNs for approximating discrete dynamical systems using polynomial number of parameters within time complexities.

Acknowledgments and Disclosure of Funding

This research was supported by the National Natural Science Foundation of China (NSFC 61921006) and the Collaborative Innovation Center of Novel Software Technology and Industrialization.

References

[1] E. D. Adrian. The impulses produced by sensory nerve endings. *The Journal of Physiology*, 61(49):156–193, 1926.

[2] A. Aertsen, P. I.M. Johannesma, and D. J. Hermes. Spectro-temporal receptive fields of auditory neurons in the grassfrog. *Biological Cybernetics*, 38(4):235–248, 1980.

[3] J. Anumula, D. Neil, T. Delbruck, and S.-C. Liu. Feature representations for neuromorphic audio spike streams. *Frontiers in Neuroscience*, 12:23, 2018.

[4] S. Arora, S. Du, W. Hu, Z. Li, and R.Wang. Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks. In *Proceedings of the 36th International Conference on Machine Learning*, pages 322–332, 2019.

[5] D. G. Barrett, S. Denève, and C. K. Machens. Firing rate predictions in optimal balanced networks. In *Advances in Neural Information Processing Systems 26*, pages 1538–1546, 2013.

[6] W. Bialek, F. Rieke, R. van Steveninck, and D. Warland. Reading a neural code. *Science*, 252(5014):1854–1857, 1991.

[7] H. S. Carslaw and W. Rogosinski. Introduction to the theory of Fourier's series and integrals. *Bulletin of the American Mathematical Society*, 37:510–511, 1931.

[8] C.-N. Chou, K.-M. Chung, and C.-J. Lu. On the algorithmic power of spiking neural networks. In *Proceedings of the 10th Innovations in Theoretical Computer Science Conference*, pages 26:1–26:20, 2019.

[9] G. Dumont, A. Payeur, and A. Longtin. A stochastic-field description of finite-size spiking neural networks. *PLoS Computational Biology*, 13(8):1005691, 2017.

[10] R. Eldan and O. Shamir. The power of depth for feedforward neural networks. In *Proceedings of the 29th Annual Conference on Learning Theory*, pages 907–940, 2016.

[11] R. Florian. Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural Computation*, 19(6):1468–1502, 2007.

[12] W. Gerstner. Time structure of the activity in neural network models. *Physical Review E*, 51(1):738, 1995.

[13] D. Huh and T. J. Sejnowski. Gradient descent for spiking neural networks. In *Advances in Neural Information Processing Systems 32*, pages 1440–1450, 2018.

[14] E. Hunsberger and C. Eliasmith. Spiking deep networks with LIF neurons. *arXiv:1510.08829*, 2015.

[15] A. Jacot, F. Gabriel, and C. Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems 32*, pages 8571 – 8580, 2018.

[16] S. R. Kheradpisheh, M. Ganjtabesh, S. J. Thorpe, and T. Masquelier. STDP-based spiking deep convolutional neural networks for object recognition. *Neural Networks*, 99:56–67, 2018.

[17] N. Kuwabara and N. Suga. Delay lines and amplitude selectivity are created in subthalamic auditory nuclei: The brachium of the inferior colliculus of the mustached bat. *Journal of Neurophysiology*, 69(5):1713–1724, 1993.

[18] J. L. Lobo, J. Del Ser, A. Bifet, and N. Kasabov. Spiking neural networks and online learning: An overview and perspectives. *Neural Networks*, 121:88–100, 2020.

[19] W. Maass. Lower bounds for the computational power of networks of spiking neurons. *Neural Computation*, 8(1):1–40, 1996.

[20] W. Maass. Networks of spiking neurons: The third generation of neural network models. *Neural Networks*, 10(9):1659–1671, 1997.

[21] W. Maass and C. M. Bishop. *Pulsed Neural Networks*. MIT press, 2001.

[22] T. Masquelier and S. J. Thorpe. Unsupervised learning of visual features through spike timing dependent plasticity. *PLoS Computational Biology*, 3(2):e31, 2007.

[23] S. Mei, T. Misiakiewicz, and A. Montanari. Learning with invariances in random features and kernel models. In *Proceedings of the 34th Annual Conference on Learning Theory*, pages 3351–3418, 2021.

[24] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[25] P. Panda and K. Roy. Unsupervised regenerative learning of hierarchical features in spiking deep networks for object recognition. In *Proceedings of the 2016 International Joint Conference on Neural Networks*, pages 299–306, 2016.

[26] T. Poggio, A. Banburski, and Q. Liao. Theoretical issues in deep networks. *Proceedings of the National Academy of Sciences*, 117(48):30039–30045, 2020.

[27] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried. Invariant visual representation by single neurons in the human brain. *Nature*, 435(7045):1102, 2005.

[28] K. Roy, A. Jaiswal, and P. Panda. Towards spike-based machine intelligence with neuromorphic computing. *Nature*, 575(7784):607–617, 2019.

[29] R. Roy and T. Kailath. Esprit-estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal processing*, 37(7):984–995, 1989.

[30] B. Schrauwen, M. DHaene, D. Verstraeten, and J. Van Campenhout. Compact hardware liquid state machines on FPGA for real-time speech recognition. *Neural Networks*, 21(2-3):511–523, 2008.

[31] X. She, S. Dash, and S. Mukhopadhyay. Sequence approximation using feedforward spiking neural network for spatiotemporal learning: Theory and optimization methods. In *Proceedings of the 9th International Conference on Learning Representations*, 2021.

[32] T. Shimokawa, K. Pakdaman, and S. Sato. Time-scale matching in the response of a leaky integrate-and-fire neuron model to periodic stimulus with additive noise. *Physical Review E*, 59(3):3427, 1999.

[33] S. B. Shrestha and G. Orchard. SLAYER: Spike layer error reassignment in time. In *Advances in Neural Information Processing Systems 32*, pages 1419–1428, 2018.

[34] P. T. P. Tang. Convergence of LCA flows to (c) LASSO solutions. *arXiv:1603.01644*, 2016.

[35] P. T. P. Tang, T.-H. Lin, and M. Davies. Sparse coding by spiking neural networks: Convergence theory and computational results. *arXiv:1705.05475*, 2017.

[36] S. Thorpe, A. Delorme, and R. Van Rullen. Spike-based strategies for rapid processing. *Neural Networks*, 14(6-7):715–725, 2001.

[37] R. VanRullen, R. Guyonneau, and S. J. Thorpe. Spike times make sense. *Trends in Neurosciences*, 28(1):1–4, 2005.

[38] E. Vasilaki, N. Frémaux, R. Urbanczik, W. Senn, and W. Gerstner. Spike-based reinforcement learning in continuous state and action space: When policy gradient methods fail. *PLoS Computational Biology*, 5(12):e1000586, 2009.

[39] D. Verstraeten, B. Schrauwen, D. Stroobandt, and J. Van Campenhout. Isolated word recognition with the liquid state machine: A case study. *Information Processing Letters*, 95(6):521–528, 2005.

[40] Q. Yu, H. Tang, K. C. Tan, and H. Yu. A brain-inspired spiking neural network model with temporal encoding and learning. *Neurocomputing*, 138:3–13, 2014.

[41] G. Zhang and S.-Q. Zhang. Structural stability of spiking neural networks. *arXiv:2207.04876*, 2022.

[42] S.-Q. Zhang, W. Gao, and Z.-H. Zhou. Towards understanding theoretical advantages of complex-reaction networks. *Neural Networks*, 151:80–93, 2022.

[43] S.-Q. Zhang, F. Wang, and F.-L. Fan. Neural network gaussian processes by increasing depth. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–6, 2022.

[44] S.-Q. Zhang, Z.-Y. Zhang, and Z.-H. Zhou. Bifurcation spiking neural network. *Journal of Machine Learning Research*, 22(253):1–21, 2021.

[45] Z.-H. Zhou. Why over-parameterization of deep neural networks does not overfit? *Science China Information Sciences*, 64(1):1–3, 2021.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes]
   (b) Did you include complete proofs of all theoretical results? [Yes]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] The simulation experiments employ the codes of Zhang and Zhou [44], and detail the data, settings, and configurations in Pages 7-8. So we believe it is easily to reproduce our experimental results. We will publicash our codes as soon as possible if this paper is accepted.
   (b) Did you specify all the training details (e.g., data splits, hyper-parameters, how they were chosen)? [Yes] See Table 2.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes]
   (b) Did you mention the license of the assets? [Yes]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

Supplementary Materials of Theoretically Provable Spiking Neural Networks (Appendix)

In this Appendix, we provide the supplementary materials for our work "Theoretically Provable Spiking Neural Networks", constructed according to the corresponding sections therein. Before that, we review the useful notations as follows.

Let $[N] = \{1, 2, \ldots, N\}$ be an integer set for $N \in \mathbb{N}^+$, and $|\cdot|_\#$ denotes the number of elements in a collection, e.g., $|[N]|_\# = N$. Let $\mathrm{i} = \sqrt{-1}$ be the imaginary unit, and $\boldsymbol{x} \preccurlyeq 0$ means that every element $x_i \leq 0$ for any i. Let the sphere $\mathcal{S}(r)$ and globe $\mathcal{B}(r)$ be $\mathcal{S}(r) = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| = r\}$ and $\mathcal{B}(r) = \{\boldsymbol{x} \mid \|\boldsymbol{x}\| \leq r\}$ for any $r \in \mathbb{R}$, respectively. Given a function $g(n)$, we denote by $h_1(n) = \Theta(g(n))$ if there exist positive constants c_1, c_2 and n_0 such that $c_1 g(n) \leq h_1(n) \leq c_2 g(n)$ for every $n \geq n_0$; $h_2(n) = \mathcal{O}(g(n))$ if there exist positive constants c and n_0 such that $h_2(n) \leq cg(n)$ for every $n \geq n_0$; $h_3(n) = \Omega(g(n))$ if there exist positive constants c and n_0 such that $h_3(n) \geq cg(n)$ for every $n \geq n_0$; $h_4(n) = o(g(n))$ if there exists positive constant n_0 such that $h_4(n) < cg(n)$ for every $c > 0$ and $n \geq n_0$.

Let $\mathcal{C}(K, \mathbb{R})$ be the set of all scalar functions $f : K \to \mathbb{R}$ continuous on $K \subset \mathbb{R}^n$. Given $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_m)^\top \in \mathbb{N}^m$, we define

$$D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) = \frac{\partial^{\alpha_1}}{\partial x^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x^{\alpha_2}} \cdots \frac{\partial^{\alpha_m}}{\partial x^{\alpha_m}} f(\boldsymbol{x}) \ ,$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in K$. Further, we define

$$\mathcal{C}^l(K, \mathbb{R}) = \{f \mid f \in \mathcal{C}(K, \mathbb{R}) \text{ and } D^r f \in \mathcal{C}(K, \mathbb{R}), \ \text{ for } \ r \in [l]\} \ .$$

For $1 \leq p < \infty$, we define

$$\mathcal{L}^p(K, \mathbb{R}) = \left\{ f \ \middle| \ f \in \mathcal{C}(K, \mathbb{R}) \text{ and } \|f\|_{p,K} \triangleq \left(\int_K |f(\boldsymbol{x})|^p \, \mathrm{d}\boldsymbol{x} \right)^{1/p} < \infty \right\} \ .$$

This work considers the Sobolev space $\mathcal{W}_\mu^{l,p}(K, \mathbb{R})$, defined as the collection of all functions $f \in \mathcal{C}^l(K, \mathbb{R})$ and $D^r f \in \mathcal{L}^p(K, \mathbb{R})$ for all $|\boldsymbol{\alpha}| \in [l]$, that is,

$$\|D^{\boldsymbol{\alpha}} f\|_{p,K} = \left(\int_K |D^{\boldsymbol{\alpha}} f(\boldsymbol{x})|^p \, \mathrm{d}\boldsymbol{x} \right)^{1/p} < \infty \ .$$

This paper employs $\mathbf{E}_n$ to denote the $n \times n$ unit matrix and $\det(\cdot)$ to indicate the determinant operation on the matrix. Two n-by-n matrices $\mathbf{A}$ and $\mathbf{B}$ are called *similar*, denoted as $\mathbf{A} \sim \mathbf{B}$, if there exists an invertible $n \times n$ matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P}$. The general linear group over field $\mathbb{F}$, denoted as $\mathbf{GL}(n, \mathbb{F})$, is the set of $n \times n$ invertible matrices with entries in $\mathbb{F}$. Especially, we define that a special linear group $\mathbf{SL}(n, \mathbb{F})$ is the subgroup of $\mathbf{GL}(n, \mathbb{F})$ and consists of matrices with determinant 1. For any field $\mathbb{F}$, the $n \times n$ orthogonal matrices form the following subgroup

$$\mathbf{O}(n, \mathbb{F}) = \{\mathbf{P} \in \mathbf{GL}(n, \mathbb{F}) \mid \mathbf{P}^\top \mathbf{P} = \mathbf{P}\mathbf{P}^\top = \mathbf{E}_n\}$$

of the general linear group $\mathbf{GL}(n, \mathbb{F})$. Similarly, we have the special orthogonal group, denoted as $\mathbf{SO}(n, \mathbb{F})$, which consists of all orthogonal matrices of determinant 1 and is a normal subgroup of $\mathbf{O}(n, \mathbb{F})$. This group is also called the rotation group, generalizing that linearly transforms geometries while holding the surface orientation.

Let $\phi^2(\boldsymbol{x})$ be the density function of some probability measure μ, which satisfies

$$\int_{\boldsymbol{x} \in \mathbb{R}^m} \phi^2(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{x} \in \mathcal{B}} 1 \, \mathrm{d}\boldsymbol{x} = 1 \ ,$$

where $\mathcal{B}$ is an abbreviation of the unit ball $\mathcal{B}(1)$. For continuous functions f, g, we have the following equalities under Fourier transform

$$\|\widehat{f\phi} - \widehat{g\phi}\|_{L_2(\mu)} = \|f\phi - g\phi\|_{L_2(\mu)} \ ,$$

and

$$\widehat{f\phi} = \hat{f} * \hat{\phi}, \quad \text{for} \quad \text{the convolution operator} * \ .$$

A Full Proof for Theorem 1

For convenience, we abbreviate $f(\cdot, t)$ as $f(\cdot)$ in this proof. For $r \in [l]$, we have

$$
\begin{aligned}
D^r f(\boldsymbol{x}) &= \int_{\mathbb{R}^m} \widehat{D^r f}(\boldsymbol{y}) \exp\left(2\pi\,\mathrm{i}\boldsymbol{y}^\top \boldsymbol{x}\right) \mathrm{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^m} \widehat{D^r f}(\beta\boldsymbol{y}) \exp\left(2\pi\beta\,\mathrm{i}\boldsymbol{y}^\top \boldsymbol{x}\right) \mathrm{d}(\beta\boldsymbol{y}) \\
&= \int_{\mathbb{R}^m} (2\pi\beta\,\mathrm{i}\boldsymbol{y})^r \widehat{f}(\beta\boldsymbol{y}) \exp\left(2\pi\beta\,\mathrm{i}\boldsymbol{y}^\top \boldsymbol{x}\right) |\beta|^m \,\mathrm{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^m} \left[\boldsymbol{y}^r |\beta|^m \widehat{f}(\beta\boldsymbol{y})\right] \left[(2\pi\beta\,\mathrm{i})^r \exp\left(2\pi\beta\,\mathrm{i}\boldsymbol{y}^\top \boldsymbol{x}\right)\right] \mathrm{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^m} \frac{\boldsymbol{y}^r |\beta|^m \widehat{f}(\beta\boldsymbol{y})}{\widehat{f}_e(\beta)} \left[\widehat{D^r f_e}(\beta) \exp\left(2\pi\beta\,\mathrm{i}\boldsymbol{y}^\top \boldsymbol{x}\right)\right] \mathrm{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^m} \frac{\boldsymbol{y}^r |\beta|^m \widehat{f}(\beta\boldsymbol{y})}{\widehat{f}_e(\beta)} \left[\int_{\mathbb{R}} D^r f_e(\alpha) \exp\left(-2\pi\,\mathrm{i}\beta\alpha\right) \mathrm{d}\alpha\right] \exp\left(2\pi\beta\,\mathrm{i}\boldsymbol{y}^\top \boldsymbol{x}\right) \mathrm{d}\boldsymbol{y}\ ,
\end{aligned}
\tag{7}
$$

where $\alpha, \beta \in \mathbb{R}$, and the above equations hold from the Fourier transforms and some of their properties. By taking the real part of Eq. (7), we have

$$
D^r f(\boldsymbol{x}) = \int_{\mathbb{R}^m} \int_{\mathbb{R}} \boldsymbol{y}^r D^r f_e(\alpha) \mathcal{K}(\alpha, \beta, \boldsymbol{y}) \,\mathrm{d}\alpha\,\mathrm{d}\boldsymbol{y}\ , \tag{8}
$$

where

$$
\mathcal{K}(\alpha, \beta, \boldsymbol{y}) = \frac{|\beta|^m \widehat{f}(\beta\boldsymbol{y}) \exp\left[2\pi\beta\,\mathrm{i}(\boldsymbol{y}^\top \boldsymbol{x} - \alpha)\right]}{\widehat{f}_e(\beta)}\ .
$$

In this proof, we set

$$
\alpha = \boldsymbol{y}^\top \boldsymbol{x} + \hbar, \quad \boldsymbol{y} = \mathbf{W}_{i,[m]}, \quad \hbar = \sum_{j \in [n]} -\frac{1}{\tau_m} \exp\left(-\frac{s - t'}{\tau_m}\right) \mathbf{V}_{i,j} \boldsymbol{s}_j(t')\ ,
$$

and the k-th element of vector $\boldsymbol{x}$ equals to a temporal-weighted average of $\mathbf{I}_k(t)$ at time interval $[t', t]$

$$
\boldsymbol{x}_k = \int_{t'}^{t} \exp\left(-\frac{s - t'}{\tau_m}\right) \mathbf{I}_k(s) \,\mathrm{d}s.
$$

Thus, we have

$$
\mathcal{K}(\alpha, \beta, \boldsymbol{y}) = \frac{|\beta|^m \widehat{f}(\beta\boldsymbol{y}) \exp\left(2\pi\beta\hbar\,\mathrm{i}\right)}{\widehat{f}_e(\beta)} \triangleq \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \quad \text{and} \quad \sup_{\boldsymbol{x} \in K} |\boldsymbol{x}| \leq C_x\ .
$$

Based on Eq. (8), we can construct a family of approximation functions of the form

$$
f_\kappa(\boldsymbol{x}) = \int_{\mathcal{B}_1} \int_{\mathcal{B}_2} f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \,\mathrm{d}\hbar\,\mathrm{d}\boldsymbol{y}\ , \tag{9}
$$

where $\mathcal{B}_1 = \{\boldsymbol{x} \mid \boldsymbol{x} \preccurlyeq \kappa\}$ and $\mathcal{B}_2 = \{x \mid x \preccurlyeq (C_x m + 1)\kappa\}$. Thus, we have

$$
D^r f_\kappa(\boldsymbol{x}) = \int_{\mathcal{B}_1} \int_{\mathcal{B}_2} \boldsymbol{y}^r D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \,\mathrm{d}\hbar\,\mathrm{d}\boldsymbol{y}\ . \tag{10}
$$

It suffices to prove that $D^r f_\kappa \to D^r f$ uniformly on K, as $\kappa \to \infty$. Now

$$
\begin{aligned}
D^r f_\kappa(\boldsymbol{x}) - D^r f(\boldsymbol{x}) = &\int_{\mathbb{R}^m / \mathcal{B}_1} \int_{\mathcal{R}} \boldsymbol{y}^r D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \,\mathrm{d}\hbar\,\mathrm{d}\boldsymbol{y} \\
&+ \int_{\mathcal{B}_1} \int_{\mathbb{R} / \mathcal{B}_2} \boldsymbol{y}^r D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \,\mathrm{d}\hbar\,\mathrm{d}\boldsymbol{y} \\
\triangleq &\ \mathcal{R}_1 + \mathcal{R}_2\ .
\end{aligned}
$$

For R_1, one has

$$
\begin{aligned}
|\mathcal{R}_1| &= \left| \int_{\mathbb{R}^m/\mathcal{B}_1} \int_{\mathcal{R}} \boldsymbol{y}^r D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \, \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y} \right| \\
&\leq \int_{\mathbb{R}^m/\mathcal{B}_1} |\boldsymbol{y}^r| \left| \int_{\mathcal{R}} D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \, \mathrm{d}\hbar \right| \mathrm{d}\boldsymbol{y} \\
&\leq \int_{\mathbb{R}^m/\mathcal{B}_1} |\boldsymbol{y}^r| \left| \int_{\mathcal{R}} D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \, \mathrm{d}\hbar \right| \left| \frac{|\beta|^m \widehat{f}(\beta \boldsymbol{y})}{\widehat{f}_e(\beta)} \right| \mathrm{d}\boldsymbol{y} \\
&\leq \left\| D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \right\|_{1,\mathbb{R}} \int_{\mathbb{R}^m/\mathcal{B}_1} \left| \frac{|\beta|^m \boldsymbol{y}^r \widehat{f}(\beta \boldsymbol{y})}{\widehat{f}_e(\beta)} \right| \mathrm{d}\boldsymbol{y} \\
&\leq \left\| D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \right\|_{1,\mathbb{R}} \int_{\mathbb{R}/\tilde{\mathcal{B}}_1} \left| \frac{|\beta \boldsymbol{y}|^r \widehat{f}(\beta \boldsymbol{y})}{\widehat{f}_e(\beta) |\beta|^r} \right| \mathrm{d}(\beta \boldsymbol{y}) \\
&\leq \frac{\left\| D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \right\|_{1,\mathbb{R}}}{\left| \widehat{f}_e(\beta) |\beta|^r \right|} \int_{\mathbb{R}/\tilde{\mathcal{B}}_1} \left| \boldsymbol{y}^r \widehat{f}(\boldsymbol{y}) \right| \mathrm{d}\boldsymbol{y} \ ,
\end{aligned}
$$

where $\tilde{\mathcal{B}}_1 = \{\beta \boldsymbol{x} \mid \beta \boldsymbol{x} \preccurlyeq \beta \kappa\}$. For R_2, one has

$$
\begin{aligned}
|\mathcal{R}_2| &= \left| \int_{\mathcal{B}_1} \int_{\mathbb{R}/\mathcal{B}_2} \boldsymbol{y}^r D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \, \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y} \right| \\
&\leq \int_{\mathcal{B}_1} \left| \int_{\mathbb{R}/\mathcal{B}_2} D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \, \mathrm{d}\hbar \right| \left| \frac{|\beta|^m \boldsymbol{y}^r \widehat{f}(\beta \boldsymbol{y})}{\widehat{f}_e(\beta)} \right| \mathrm{d}\boldsymbol{y} \\
&\leq \int_{\mathbb{R}/\tilde{\mathcal{B}}_2} |D^r f_e(\mu)| \, \mathrm{d}\mu \cdot \int_{\tilde{\mathcal{B}}_1} \left| \frac{|\beta \boldsymbol{y}|^r \widehat{f}(\beta \boldsymbol{y})}{\widehat{f}_e(\beta) |\beta|^r} \right| \mathrm{d}(\beta \boldsymbol{y}) \\
&\leq \int_{\mathbb{R}/\tilde{\mathcal{B}}_2} |D^r f_e(\mu)| \, \mathrm{d}\mu \, \frac{\| D^r f_e(\mu) \|_{1,\tilde{\mathcal{B}}_1}}{\left| \widehat{f}_e(\beta) |\beta|^r \right|} \ ,
\end{aligned}
$$

where $\mu = \boldsymbol{y}^\top \boldsymbol{x} + \hbar$ and $\tilde{\mathcal{B}}_2 = \{x \mid x \preccurlyeq \kappa\}$ since $|\mu| \geq \kappa$. Summing up the inequalities above, we have

$$
\sup_{\boldsymbol{x} \in K} |D^r f_\kappa(\boldsymbol{x}) - D^r f(\boldsymbol{x})| \leq \frac{C_\kappa^1 + C_\kappa^2}{\left| \widehat{f}_e(\beta) |\beta|^r \right|}
$$

with

$$
C_\kappa^1 = \left\| D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \right\|_{1,\mathbb{R}} \int_{\mathbb{R}/\tilde{\mathcal{B}}_1} \left| \boldsymbol{y}^r \widehat{f}(\boldsymbol{y}) \right| \mathrm{d}\boldsymbol{y} \quad \text{and} \quad C_\kappa^2 = \| D^r f_e(\mu) \|_{1,\tilde{\mathcal{B}}_1} \int_{\mathbb{R}/\tilde{\mathcal{B}}_2} |D^r f_e(\mu)| \, \mathrm{d}\mu \ ,
$$

which tends to 0 as $\kappa \to \infty$.

Given κ, it suffices to construct a series of approximations to f_κ in Eq. (9). Formally, we define

$$
\tilde{f}_\kappa^n(\boldsymbol{x}) = \sum_{\boldsymbol{\mu} \in \mathcal{U}} \tilde{\beta} f_e(\tilde{\boldsymbol{y}}^\top \boldsymbol{x} + \tilde{\hbar}) \ ,
$$

where

$$
\begin{cases}
\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_m)^\top \quad \text{with} \quad \mu_i \in [-n, n] \cap \mathbb{Z} \quad \text{for} \quad i \in [m], \\
\tilde{\beta} = (C_x m + 1)(\kappa/n)^{m+1} \mathcal{K}_\beta(\tilde{\hbar}, \tilde{\boldsymbol{y}}) \ , \\
\tilde{\boldsymbol{y}} = \boldsymbol{\mu} \kappa / n \ , \\
\tilde{\hbar} = \mu^* (C_x m + 1) \kappa / n \quad \text{with} \quad \mu^* \in [-n, n] \cap \mathbb{Z} \ .
\end{cases}
$$

It is observed that $\tilde{f}_\kappa^n$ belongs to the set of IFR functions, and

$$
D^r \tilde{f}_\kappa^n(\boldsymbol{x}) = \sum_{\boldsymbol{\mu} \in \mathcal{U}} (C_x m + 1)(\kappa/n)^{m+1} \, \tilde{\boldsymbol{y}}^r D^r f_e(\tilde{\boldsymbol{y}}^\top \boldsymbol{x} + \tilde{\hbar}) \, \mathcal{K}_\beta(\tilde{\hbar}, \tilde{\boldsymbol{y}}) \ . \tag{11}
$$

Next, we are going to prove that $D^r \tilde{f}_\kappa^n \to D^r f_\kappa$ uniformly on K, as $n \to \infty$. For simplicity, we define the following function

$$G_\beta(\boldsymbol{x}, \boldsymbol{y}, \hbar) = \boldsymbol{y}^r D^r f_e(\boldsymbol{y}^\top \boldsymbol{x} + \hbar) \mathcal{K}_\beta(\hbar, \boldsymbol{y}) \ .$$

Thus, Eq. (10) and Eq. (11) become

$$D^r f_\kappa(\boldsymbol{x}) = \sum_{\boldsymbol{\mu} \in \mathcal{U}} \int_{\mathcal{B}_3} G_\beta(\boldsymbol{x}, \boldsymbol{y}, \hbar) \, \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y}$$

and

$$D^r \tilde{f}_\kappa^n(\boldsymbol{x}) = \sum_{\boldsymbol{\mu} \in \mathcal{U}} \int_{\mathcal{B}_3} G_\beta(\boldsymbol{x}, \tilde{\boldsymbol{y}}, \tilde{\hbar}) \, \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y}$$

respectively, where $\cup_{\boldsymbol{\mu} \in \mathcal{U}} \mathcal{B}_3 = \{(x_0, x_1, \ldots, x_m) \mid x_0 \in \mathcal{B}_2, \ \ (x_1, \ldots, x_m)^\top \in \mathcal{B}_1\} \subset \mathbb{R}^{m+1}$. Hence, one has

$$\sup_{(\hbar, \boldsymbol{y}), (\tilde{\hbar}, \tilde{\boldsymbol{y}}) \in \mathcal{B}_3} \left| G_\beta(\boldsymbol{x}, \boldsymbol{y}, \hbar) - G_\beta(\boldsymbol{x}, \tilde{\boldsymbol{y}}, \tilde{\hbar}) \right| < \infty \ .$$

Let

$$C_\kappa^n(\delta) \quad \triangleq \sup_{\substack{(\hbar, \boldsymbol{y}), (\tilde{\hbar}, \tilde{\boldsymbol{y}}) \in \mathcal{B}_3 \\ |(\hbar, \boldsymbol{y}) - (\tilde{\hbar}, \tilde{\boldsymbol{y}})| \leq \delta^{m+1}}} \left| G_\beta(\boldsymbol{x}, \boldsymbol{y}, \hbar) - G_\beta(\boldsymbol{x}, \tilde{\boldsymbol{y}}, \tilde{\hbar}) \right| .$$

Thus, we have

$$\begin{aligned}
\left| D^r \tilde{f}_\kappa^n(\boldsymbol{x}) - D^r f_\kappa(\boldsymbol{x}) \right| &\leq \sum_{\boldsymbol{\mu} \in \mathcal{U}} \int_{\mathcal{B}_3} \left| G_\beta(\boldsymbol{x}, \boldsymbol{y}, \hbar) - G_\beta(\boldsymbol{x}, \tilde{\boldsymbol{y}}, \tilde{\hbar}) \right| \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y} \\
&\leq \sum_{\boldsymbol{\mu} \in \mathcal{U}} \int_{\mathcal{B}_3} C_\kappa^n(\kappa/n) \, \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y} \\
&\leq C_\kappa^n(\kappa/n) \sum_{\boldsymbol{\mu} \in \mathcal{U}} \int_{\mathcal{B}_3} \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y} \\
&\leq C_\kappa^n(\kappa/n) \ (2n)^{m+1} \ (C_x m + 1)(\kappa/n)^{m+1},
\end{aligned}$$

where the last inequality holds from

$$\int_{\mathcal{B}_3} \mathrm{d}\hbar \, \mathrm{d}\boldsymbol{y} = (C_x m + 1)(\kappa/n)^{m+1} \quad \text{and} \quad |\mathcal{U}|_\# = (2n)^{m+1}.$$

Further, we can obtain

$$\sup_{\boldsymbol{x} \in K} \left| D^r \tilde{f}_\kappa^n(\boldsymbol{x}) - D^r f_\kappa(\boldsymbol{x}) \right| \leq (C_x m + 1)(2\kappa)^{m+1} \ C_\kappa^n(\kappa/n) \ ,$$

which tends to 0 as $n \to \infty$.

Finally, we finish the proof by taking double limits $n \to \infty$ before $\kappa \to \infty$. □

B Full Proof for Theorem 2

There are two proof methods for Theorem 2, one based on the rotation group action and one based on the Fourier transformation, which are detailed in this and the following sections, respectively.

This proof idea can be summarized as follows. It is observed that the IFR function comprises the radius and phase since $\mathbf{V}$ is a symmetric matrix. Thus, there exist some linear connections (including rotation transformations) such that the combination of these spiking neurons is invariant to rotations. In other words, the IFR function led by scSNNs with one-hidden layer can easily and well approximate some radial functions, see Lemma 1, since the radial function is invariant to rotations, and is dependent on the input norm.

B.1 Proof of Lemma 1

Let $f' : \mathbb{R}^m \to \mathbb{R}$ be a radial function with $f'(\boldsymbol{x}) = \|\boldsymbol{x}\|$. According to Theorem 1, for any $\delta > 0$, $i \in [n]$, and $m \geq 1$, there exists some time t such that

$$\sup_{\boldsymbol{x}\in\mathbb{R}^d} \Big| f'(\boldsymbol{x}) - |f_i(\boldsymbol{x}, t)| \Big| \leq \delta/2 \ , \tag{12}$$

where f_i denotes the IFR function of the i-th hidden spiking neuron, defined by Eq. (4). Further, we define a new function $g' : \mathbb{R} \to \mathbb{R}$ as follows.

$$g'(s) = \sum_{i=1}^{n'} \alpha'_i f_e(s) \ ,$$

where $\alpha'_i, a'_i \in \mathbb{R}$. For Lipschitz continuous function $\sqrt[r]{\cdot}$ and from [42, Lemma 1], we have

$$\sup_{s\in[r^k, R^k]} \big| g(\sqrt[k]{s}) - g'(s) \big| \leq \delta/2 \ , \tag{13}$$

where $n' \leq C' L(R^k - r^k)/(\sqrt[k]{r}\delta)$ for some constant $C' > 0$ and integer $k \geq 2$. Further, we have

$$|g'(s) - f(\boldsymbol{x}, t)| \leq |g'(s) - f'(\boldsymbol{x})| + |f'(\boldsymbol{x}) - f(\boldsymbol{x}, t)| \ , \tag{14}$$

where

$$f'(\boldsymbol{x}) = \sum_{i=1}^{n'} w'_i \ |f_i(\boldsymbol{x}, t))| \ ,$$

in which $\{w'_i\}$ denotes another collection of weighted parameters that corresponds to $f(\cdot, t) = \sum_{i\in[n]} w_i f_i(\cdot, t)$. The first term of Eq. (14) can be bounded by $\delta/4$ from [42, Lemma 1] for any $s \in [r^k, R^k]$. The second term is at most $\delta/4$ when $n \geq n'$ from Eq. (12). This follows that

$$|g'(s) - f(\boldsymbol{x}, t)| \leq \delta/2 \ . \tag{15}$$

Combining with Eqs. (13) and (15), we have

$$|g(\|\boldsymbol{x}\|) - f(\boldsymbol{x}, t)| \leq \big| g(\sqrt[k]{s}) - g'(s) \big| + |g'(s) - f(\boldsymbol{x}, t)| \leq \delta \ ,$$

where $\boldsymbol{x} \in \mathbb{R}^m$ and $s \in [r^k, R^k]$. We finally obtain

$$n \leq C_s(R^k - r^k)Lm/(\sqrt[k]{r}\delta) \ ,$$

provided $n \leq mn'$ and $C' \leq C_s$. Finally, we can complete the proof by setting $k = 2$ in the above upper bound. □

B.2 Proof of Lemma 2

Let $r = C_2\sqrt{m}$, $R = 2C_2\sqrt{m}$, and $m \geq 1$, then we have $r \geq 1$, which satisfies the condition of [42, Lemma 1]. Invoke Lemma 1 to construct the concerned spiking neural networks and define $\delta' \leq \delta/d$. Then for any L-Lipschitz radial function $g : \mathbb{R}^m \to \mathbb{R}$ supported on $\mathcal{S}_\triangle$, we have

$$\sup_{\boldsymbol{x}\in\mathbb{R}^m} |g(\boldsymbol{x}) - f(\boldsymbol{x}, t)| \leq \delta' \ ,$$

where the width of the hidden layer is bounded by

$$n \leq \frac{C_s(C_2)^{3/2} mL}{\delta}(m)^{3/4} \leq \frac{C_s(C_2)^{3/2} L}{\delta}(m)^{7/4} \ .$$

This completes the proof. □

B.3 Proof of Lemma 3

Define a branch function

$$h_i(\boldsymbol{x}) = \begin{cases} \max\{\mathbb{I}\{\|\boldsymbol{x}\| \in \Omega_i\}, ND_i\}, & \text{if} \quad B_i = 1, \\ \mathbf{0} \qquad\qquad\qquad\qquad , & \text{if} \quad B_i = 0, \end{cases}$$

with

$$D_i = \min\left\{\left|\|\boldsymbol{x}\| - \left(1 + \frac{i-1}{N}\right) C_2\sqrt{m}\right|, \left|\|\boldsymbol{x}\| - \left(1 + \frac{i}{N}\right) C_2\sqrt{m}\right|\right\} .$$

Let

$$h(\boldsymbol{x}) = \sum_{i=1}^{N} \epsilon_i h_i(\boldsymbol{x}) ,$$

$B_i = 1$, Ω_i's are disjoint intervals, $h_i(\boldsymbol{x})$ is an N-Lipschitz function. Thus, h is also an N-Lipschitz function. So we have

$$\begin{aligned} \int_{\mathbb{R}^m} \left(h(\boldsymbol{x}) - \sum_{i=1}^{N} \epsilon_i g_i(\boldsymbol{x})\right)^2 \phi^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_{\mathbb{R}^m} \sum_{i=1}^{N} \epsilon_i^2 \left(h_i(\boldsymbol{x}) - g_i(\boldsymbol{x})\right)^2 \phi^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &= \sum_{i=1}^{N} \int_{\mathbb{R}^m} \left(h_i(\boldsymbol{x}) - g_i(\boldsymbol{x})\right)^2 \phi^2(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\ &\leq (3/(C_2)^2\sqrt{m}) , \end{aligned}$$

where the last inequality holds from [10, Lemma 22]. This completes the proof. □

Based on the lemmas above, we can finish the proof of Theorem 2.

Finishing the Proof of Theorem 2. Let $g(\boldsymbol{x}) = \sum_{i=1}^{N} \epsilon_i g_i(\boldsymbol{x})$ be defined by Eq. (5) and $N \geq 4C_2^{5/2}m^2$. According to Lemma 3, there exists a Lipschitz function h with range $[-1, +1]$ such that

$$\|h(\boldsymbol{x}) - g(\boldsymbol{x})\|_{L_2(\mu)} \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} .$$

Based on Lemmas 1 and 2, any Lipschitz radial function supported on $\mathcal{S}_\Delta$ can be approximated by an IFR function $f(\boldsymbol{x}, t)$ led by scSNNs of one-hidden layer with width at most $C_3C_s(m)^{15/4}$, where C_3 is a constant relative to C_2 and δ. This means that there exists some time t such that

$$\sup_{\boldsymbol{x}\in\mathbb{R}^m} |h(\boldsymbol{x}) - f(\boldsymbol{x}, t)| \leq \delta .$$

Thus, we have

$$\|h(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \delta .$$

Hence, the range of $f(\cdot, t)$ is in $[-1-\delta, +1+\delta] \subseteq [-2, +2]$. Provided the radial function, defined by Eq. (5), we have

$$\|g(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \|g(\boldsymbol{x}) - h(\boldsymbol{x})\|_{L_2(\mu)} + \|h(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} + \delta .$$

This implies that given constants $m > C_2 > 0$ and $C_3 > 0$, for any $\delta > 0$ and $\epsilon_i \in \{-1, +1\}$ ($i \in [N]$), there exists some time t, such that the target radial function g can be approximated by an IFR function $f(\boldsymbol{x}, t)$ led by scSNNs of one-hidden layer with range in $[-2, +2]$ and width at most $C_3C_s(m)^{15/4}$, that is,

$$\|g(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} + \delta < \delta_1 .$$

This completes the proof. □

C Another Proof for Theorem 2

The another proof idea can be summarized as follows. Given any permutation operation on the input channels, there are some permutation operations on the self-connection and final connection weights such that the output of SNNs are invariant. In other words, the IFR function led by scSNNs with one-hidden layer admits the rotation transformations. Therefore, it suffices to show that before final weighted aggregation, the component IFR functions of hidden spiking neurons can approximate any L-Lipschitz radial function, similar to the proof line of Lemma 1, and then find a special radial function that can be well approximated by these IFR functions within the polynomial parameter complexity, similar to the proof line of Lemma 3.

Lemma 5 *Let f_e and $f(\boldsymbol{x}, t)$ be an l-finite function and the IFR function led by scSNNs with one-hidden layer, respectively. For any $\delta > 0$ and $\mathbf{A} \in \mathbf{SO}(m, \mathbb{R})$, there exists some time t such that*

$$|f(\mathbf{A}\boldsymbol{x}, t) - \|\boldsymbol{x}\|| < \delta \ .$$

Lemma 6 *Let f_e be an l-finite function, $g : \mathbb{R} \to \mathbb{R}$ is a scalar function, and $f(\boldsymbol{x}, t)$ is the IFR function led by scSNNs with one-hidden layer, where $\boldsymbol{x} \in \mathcal{S}(r)$ and $\mathcal{S}(r)$ is a sphere supported with density ϕ^2 for $0 < r < \infty$. For any $\delta > 0$, if there exists some time $t \in \mathbb{R}$ such that it holds*

$$|f(\boldsymbol{x}, t) - g(\|\boldsymbol{x}\|)| < \delta \quad \textit{if and only if} \quad |f(\mathbf{A}\boldsymbol{x}, t) - g(\|\boldsymbol{x}\|)| < \delta \ ,$$

for any $\mathbf{A} \in \mathbf{SO}(m, \mathbb{R})$.

Lemma 7 *Define a radial function $g' : \mathbb{R} \to \mathbb{R}$ with the form of*

$$g'(\|\boldsymbol{x}\|) \stackrel{\text{def}}{=} \sum_{i=1}^{N} \epsilon_i \, \mathbb{I}\{\|\boldsymbol{x}\| \in \Omega_i\} \ ,$$

where N is a polynomial function of m, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N) \in \{-1, +1\}^N$, and Ω_i's are disjoint intervals of width $\mathcal{O}(1/N)$ on values in the range $\Theta(\sqrt{m})$. For $C_2, C_3 > 0$ with $d > C_2$, any $\delta > 0$, and any choice of $\epsilon_i \in \{-1, +1\}$ $(i \in [N])$, there exist some time t and $\mathbf{A} \in \mathbf{SO}(m, \mathbb{R})$ such that

$$|g'(\|\boldsymbol{x}\|) - f(\mathbf{A}\boldsymbol{x}, t)| \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} + \delta \ ,$$

where $f(\boldsymbol{x}, t)$ indicates the IFR function led by scSNNs of one-hidden layer with range in $[-2, +2]$ and at most $C_3 C_s(m)^{15/4}$ hidden spiking neurons.

Finishing the Proof of Theorem 2. The rest proof is a straightforward combination of Lemmas 5, 6 and 7. Define a radial function $g' : \mathbb{R} \to \mathbb{R}$ with the form of

$$g'(\|\boldsymbol{x}\|) \stackrel{\text{def}}{=} \sum_{i=1}^{N} \epsilon_i \, \mathbb{I}\{\|\boldsymbol{x}\| \in \Omega_i\} \ ,$$

where N is a polynomial function of m, $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_N) \in \{-1, +1\}^N$, and Ω_i's are disjoint intervals of width $\mathcal{O}(1/N)$ on values in the range $\Theta(\sqrt{m})$. Obviously, g' is equivalent to another radial function defined in Eq. (5) as follows

$$g(\boldsymbol{x}) = \sum_{i=1}^{N} \epsilon_i g_i(\boldsymbol{x}) \quad \text{with} \quad g_i(\boldsymbol{x}) = \mathbb{I}\{\|\boldsymbol{x}\| \in \Omega_i\} \ ,$$

provided $\boldsymbol{x} \in \mathcal{S}(r)$. From Lemma 7, it is observed that provided the concerned radial function $g(\boldsymbol{x})$ with a collection of choices $\boldsymbol{\epsilon}$, there exist some time t and $n \in \mathcal{O}(C_3 C_s(m)^{15/4})$ such that $g(\boldsymbol{x})$ can be approximated a IFR function $f(\boldsymbol{x}, t)$ led by one-hidden-layer scSNNs of with range in $[-2, +2]$ and at most n hidden spiking neurons, such that

$$|g'(\|\boldsymbol{x}\|) - f(\mathbf{A}\boldsymbol{x}, t)| \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} + \delta \ .$$

Let $\mathbf{A} \in \mathbf{SO}(m, \mathbb{R})$. If $\|\boldsymbol{x}\| \in \Omega_i \subseteq \mathcal{S}(r)$ for $0 < r < \infty$, then it holds $\|\mathbf{A}\boldsymbol{x}\| \in \Omega_i \subseteq \mathcal{S}(r)$ and $g'(\|\boldsymbol{x}\|) = g'(\|\mathbf{A}\boldsymbol{x}\|)$. According to Lemma 6, we have

$$\|g(\boldsymbol{x}) - f(\boldsymbol{x}, t)\|_{L_2(\mu)} = |g'(\|\boldsymbol{x}\|) - f(\boldsymbol{x}, t)| = |g'(\|\boldsymbol{x}\|) - f(\mathbf{A}\boldsymbol{x}, t)| \leq \frac{\sqrt{3}}{C_2(m)^{1/4}} + \delta \ ,$$

without any incremental change in the parameter complexity. This completes the proof. □

D Full Proof for Theorem 3

We begin this proof with an investigation that acting SVD on the matrix $\mathbf{G}$. Let $\mathbf{G} = \mathbf{P}\ \mathbf{\Lambda}_G\ \mathbf{Q}^\top$, where $\mathbf{P} \in \mathbb{R}^{m\times m}$ and $\mathbf{Q} \in \mathbb{R}^{n\times n}$ are two unitary matrices, $\mathbf{\Lambda}_G \in \mathbb{R}^{m\times n}$ is a diagonal matrix. Thus, we have

$$\left\|\mathbb{E}_{\boldsymbol{x}}\left[\mathbf{V}\boldsymbol{f}(\boldsymbol{x},t)\right] - \mathbf{G}\boldsymbol{\lambda}\right\|_2 = \left\|\mathbf{Q}\widetilde{\mathbf{V}}\ \mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{f}(\boldsymbol{x},t)\right] - \mathbf{P}\mathbf{\Lambda}_G\boldsymbol{\lambda}\right\|_2 \left\|\mathbf{Q}^\top\right\|_2 \leq 2\left\|\widetilde{\mathbf{V}}\ \mathbb{E}_{\boldsymbol{x}}\left[\boldsymbol{f}(\boldsymbol{x},t)\right] - \mathbf{\Lambda}_G\boldsymbol{\lambda}\right\|_2 . \tag{16}$$

If the matrix $\mathbf{G}$ is degenerate, then $\mathbf{\Lambda}_G$ contains some null diagonal elements. Then there must exists a non-degenerate sub-matrix $\mathbf{G}_{\text{sub}}$ of $\mathbf{G}$ such that Inequality (6) degenerates to one about $\mathbf{G}_{\text{sub}}$ in a lower dimensional space within at most constant time cost. Thus, we only care about the case that $\mathbf{G}$ is a non-degenerate matrix. When $m \leq n$, the vector $\mathbf{\Lambda}_G\boldsymbol{\lambda}$ becomes a constant vector, denoted as $\tilde{\boldsymbol{\lambda}}$, whose elements consist of $\lambda_1, \ldots, \lambda_m$ and $n-m$ zeros.

Notice that the $n\times n$-dimensional matrix $\widetilde{\mathbf{V}}$ is symmetric and non-degenerate, since $\mathbf{V}$ is a symmetric and non-degenerate matrix and $\mathbf{Q} \in \mathbf{O}(n)$. Thus, we have $\widetilde{\mathbf{V}} = \widetilde{\mathbf{U}}\ \mathbf{\Lambda}_U\ \widetilde{\mathbf{U}}^\top$, where $\widetilde{\mathbf{U}} \in \mathbf{O}(n)$ and $\mathbf{\Lambda}_U = \operatorname{diag}\{\rho_1, \ldots, \rho_n\}$ in which $\rho_k \neq 0$ for any $k \in [n]$. Then we have the following lemma.

Lemma 8 (Lemma 4 as aforementioned) *Given* $\mathbf{U} = \sqrt{\mathbf{\Lambda}_U}\ \widetilde{\mathbf{U}}^\top, \tilde{\boldsymbol{\lambda}} \in \mathbb{R}^n$*, and* $\epsilon > 0$*, when* $t \geq \Omega\left(\frac{\sqrt{n}\ \|\Lambda_G\|_2}{\epsilon\ \|\mathbf{U}\|_2}\right)$*, it holds* $\|\mathbf{U}\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\lambda}},t) - \tilde{\boldsymbol{\lambda}}\| \leq \epsilon$*, where the modified IFR function* $\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\lambda}},t)$ *is led by a scSNN without connection matrix* $\mathbf{W}$ *and fed up to the constant spike sequence* $\tilde{\boldsymbol{\lambda}}$ *at every timestamp.*

This lemma is a straightforward derivation of [8, Theorem 1] due to the following two facts. (1) For any $0 < \gamma \leq \min\{\rho_k, k \in [n]\}$, the matrix $\mathbf{U}$ meets the regular conditions of [8, Definition 1]. (2) The optimal solution of $\mathbf{U}\boldsymbol{z} = \tilde{\boldsymbol{\lambda}}$ becomes $\boldsymbol{z}^* = \mathbf{U}^{-1}\tilde{\boldsymbol{\lambda}}$. Hence, we omit the detailed proof of this lemma.

Finally, it suffices to prove that the concerned IFR function $\boldsymbol{f}(\boldsymbol{x},t)$ can approximate $\tilde{\boldsymbol{f}}(\tilde{\boldsymbol{\lambda}},t^*)$ within time interval $[0,T]$ where $t^* \leq T$ and $T \geq \Omega(\frac{\sqrt{n}\ \|\mathbf{G}\|_2}{\epsilon\sqrt{\|\mathbf{V}\|_2}})$. Since f_e is the linear function defined by Eq. (3). The approximation above can be converted into another one between

$$\boldsymbol{u}_i(t) = \int_{t'}^{t} \exp\left(\frac{s-t'}{\tau_m}\right)\left(\sum_{j\in[n]} \mathbf{V}_{ij}\boldsymbol{s}_j(t) + \sum_{k\in[m]} \mathbf{W}_{ik}\boldsymbol{x}_k(t)\right)\mathrm{d}s$$

and

$$\tilde{\boldsymbol{u}}_i(t^*) = \int_{t''}^{t^*} \exp\left(\frac{s-t'}{\tau_m}\right)\tilde{\boldsymbol{\lambda}}_i(s)\,\mathrm{d}s\ .$$

Since

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{u}_i(t)\right] &= \int_{t'}^{t} \exp\left(\frac{s-t'}{\tau_m}\right)\left(\sum_{j\in[n]} \mathbf{V}_{ij}\mathbb{E}\left[\boldsymbol{s}_j(t)\right] + \sum_{k\in[m]} \mathbf{W}_{ik}\mathbb{E}\left[\boldsymbol{x}_k(t)\right]\right)\mathrm{d}s \\ &= \int_{t'}^{t} \exp\left(\frac{s-t'}{\tau_m}\right)\left(\sum_{j\in[n]} \mathbf{V}_{ij}\mathbb{E}\left[\boldsymbol{s}_j(t)\right] + \sum_{k\in[m]} \mathbf{W}_{ik}\lambda_k(t)\right)\mathrm{d}s\ ,\end{aligned}$$

we have

$$\|\mathbb{E}\left[\boldsymbol{u}(T)\right] - \tilde{\boldsymbol{u}}(t^*)\| \leq \epsilon, \quad \left\|\widetilde{\mathbf{U}}\ \mathbb{E}\left[\boldsymbol{u}(T)\right] - \tilde{\boldsymbol{u}}(t^*)\right\| \leq \epsilon, \quad \text{and} \quad \|\mathbf{U}\ \mathbb{E}\left[\boldsymbol{u}(T)\right] - \tilde{\boldsymbol{u}}(t^*)\| \leq \epsilon\ .$$

Thus, we have

$$\left\|\mathbf{U}\ \mathbb{E}\left[\boldsymbol{f}(\boldsymbol{x},T)\right] - \tilde{\boldsymbol{f}}(\boldsymbol{x},t^*)\right\| \leq \epsilon\ .$$

We finally complete the proof according to $\|\mathbf{G}\|_2 = \|\Lambda_G\|_2$ and $\sqrt{\|\mathbf{V}\|_2} = \|\mathbf{U}\|_2$. □