

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 $[N] = \{1, 2, \dots, N\}$ be an integer set for $N \in \mathbb{N}^+$, and $|\cdot|_{\#}$ denotes the number of elements in a collection, e.g., $|[N]|_{\#} = N$. Let $i = \sqrt{-1}$ be the imaginary unit, and $\mathbf{x} \preceq 0$ means that every element $x_i \leq 0$ for any i . Let the sphere $\mathcal{S}(r)$ and globe $\mathcal{B}(r)$ be $\mathcal{S}(r) = \{\mathbf{x} \mid \|\mathbf{x}\| = r\}$ and $\mathcal{B}(r) = \{\mathbf{x} \mid \|\mathbf{x}\| \leq r\}$ 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 $h_2(n) \leq c g(n)$ for every $n \geq n_0$; $h_3(n) = \Omega(g(n))$ if there exist positive constants c and n_0 such that $h_3(n) \geq c g(n)$ for every $n \geq n_0$; $h_4(n) = o(g(n))$ if there exists positive constant n_0 such that $h_4(n) < c g(n)$ for every $c > 0$ and $n \geq n_0$.

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

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

where $\mathbf{x} = (x_1, x_2, \dots, x_n) \in K$. 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

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

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,

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

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

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

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

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

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

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

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

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

where

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

In this proof, we set

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

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

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

Thus, we have

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

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

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

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

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

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

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

For R_1 , one has

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

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

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

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

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

with

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

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

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

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

where

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

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

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

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

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

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

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

and

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

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

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

Let

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

Thus, we have

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

where the last inequality holds from

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

Further, we can obtain

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

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

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

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

$$\sup_{\mathbf{x} \in \mathbb{R}^d} \left| f'(\mathbf{x}) - |f_i(\mathbf{x}, t)| \right| \leq \delta/2, \quad (12)$$

where f_i denotes the IFR function of the i -th hidden spiking neuron, defined by Eq. (4). Further, we define a new function $g' : \mathbb{R} \rightarrow \mathbb{R}$ 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 $\sqrt[k]{\cdot}$ and from [42, Lemma 1], we have

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

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

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

where

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

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

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

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

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

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

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

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

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 $g : \mathbb{R}^m \rightarrow \mathbb{R}$ supported on \mathcal{S}_Δ , we have

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

where the width of the hidden layer is bounded by

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

This completes the proof. \square

B.3 Proof of Lemma 3

Define a branch function

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

with

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

Let

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

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

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

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

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

Finishing the Proof of Theorem 2. Let $g(\mathbf{x}) = \sum_{i=1}^N \epsilon_i g_i(\mathbf{x})$ 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

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

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

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

Thus, we have

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

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

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

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 $f(\mathbf{x}, t)$ led by scSNNs of one-hidden layer with range in $[-2, +2]$ and width at most $C_3 C_s(m)^{15/4}$, that is,

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

This completes the proof. \square

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 $f(\mathbf{x}, t)$ 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*

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

Lemma 6 *Let f_e be an l -finite function, $g : \mathbb{R} \rightarrow \mathbb{R}$ is a scalar function, and $f(\mathbf{x}, t)$ is the IFR function led by scSNNs with one-hidden layer, where $\mathbf{x} \in \mathcal{S}(r)$ 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*

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

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

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

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

where N is a polynomial function of m , $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_N) \in \{-1, +1\}^N$, 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

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

where $f(\mathbf{x}, t)$ 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 $g' : \mathbb{R} \rightarrow \mathbb{R}$ with the form of

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

where N is a polynomial function of m , $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_N) \in \{-1, +1\}^N$, 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

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

provided $\mathbf{x} \in \mathcal{S}(r)$. From Lemma 7, it is observed that provided the concerned radial function $g(\mathbf{x})$ 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 $g(\mathbf{x})$ can be approximated a IFR function $f(\mathbf{x}, t)$ led by one-hidden-layer scSNNs of with range in $[-2, +2]$ and at most n hidden spiking neurons, such that

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

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

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

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

D Full Proof for Theorem 3

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

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

If the matrix \mathbf{G} is degenerate, then Λ_G contains some null diagonal elements. Then there must exists a non-degenerate sub-matrix \mathbf{G}_{sub} of \mathbf{G} such that Inequality (6) degenerates to one about \mathbf{G}_{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 $\Lambda_G \boldsymbol{\lambda}$ becomes a constant vector, denoted as $\tilde{\boldsymbol{\lambda}}$, whose elements consist of $\lambda_1, \dots, \lambda_m$ and $n - m$ zeros.

Notice that the $n \times n$ -dimensional matrix $\tilde{\mathbf{V}}$ 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 $\tilde{\mathbf{V}} = \tilde{\mathbf{U}} \Lambda_U \tilde{\mathbf{U}}^\top$, where $\tilde{\mathbf{U}} \in \mathbf{O}(n)$ and $\Lambda_U = \text{diag}\{\rho_1, \dots, \rho_n\}$ in which $\rho_k \neq 0$ for any $k \in [n]$. Then we have the following lemma.

Lemma 8 (Lemma 4 as aforementioned) *Given $\mathbf{U} = \sqrt{\Lambda_U} \tilde{\mathbf{U}}^\top$, $\tilde{\boldsymbol{\lambda}} \in \mathbb{R}^n$, and $\epsilon > 0$, when $t \geq \Omega\left(\frac{\sqrt{n} \|\Lambda_G\|_2}{\epsilon \|\mathbf{U}\|_2}\right)$, it holds $\|\mathbf{U} \tilde{\mathbf{f}}(\tilde{\boldsymbol{\lambda}}, t) - \tilde{\boldsymbol{\lambda}}\| \leq \epsilon$, where the modified IFR function $\tilde{\mathbf{f}}(\tilde{\boldsymbol{\lambda}}, t)$ 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 $\mathbf{U} \mathbf{z} = \tilde{\boldsymbol{\lambda}}$ becomes $\mathbf{z}^* = \mathbf{U}^{-1} \tilde{\boldsymbol{\lambda}}$. Hence, we omit the detailed proof of this lemma.

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

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

and

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

Since

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

we have

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

Thus, we have

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

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