Universal Approximation Property of Quantum Feature Map

Takahiro Goto*
Nagase Hongo Building F5, 5-24-5, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Quoc Hoan Tran† and Kohei Nakajima‡
Graduate School of Information Science and Technology, The University of Tokyo, Tokyo 113-8656, Japan

(Dated: September 2, 2020)

Encoding classical inputs into quantum states is considered as a quantum feature map to map the classical data into the quantum Hilbert space. This feature map paves opportunities to merge the advantages of quantum mechanics into machine learning algorithms to perform on the near-term intermediate-scale quantum computers. While the quantum feature map has demonstrated its capability when combining with linear classification models in some specific applications, its expressive power from the theoretical perspective remains unknown. We prove that the quantum feature map is a universal approximator of continuous functions under its typical settings in many practical applications. We further study the capability of the quantum feature map in the classification of disjoint regions. Our work enables a theoretical analysis of the feasibility of quantum-enhanced machine learning algorithms. In light of this, one can utilize knowledge to design a quantum machine learning model with more powerful expressivity.

Introduction.— The rapidly increasing volume and complexity of data have led to the notable progress of machine learning (ML) techniques that employ von Neumann architectures to build sophisticated models for finding patterns in data. The core of its interest lies in the ability to recognize the patterns that it can produce. If a physical computation model can produce atypical patterns that cannot be generated by a classical computer, it may reveal patterns that are difficult to recognize in the classical regime [1]. This expectation has led to the advent of quantum machine learning (QML), a field that takes advantages of quantum effects to surpass the classical ML techniques in terms of computational complexity or pattern recognition. QML is currently benefiting from the arrival of noisy intermediate-scale quantum (NISQ) devices which may include a few tens to hundreds of qubits with no error correction capability [2, 3]. Such devices have ushered in the era of hybrid quantum-classical algorithms such as variational quantum eigensolver [4], quantum neural networks [5, 6], quantum classifiers [7–9], and quantum reservoir computing [10–13].

Since a quantum computer can efficiently access and manipulate quantum states, the quantum Hilbert space can be used as a feature space for classical data. The input data is encoded in a quantum state via a quantum feature map, a nonlinear feature map that maps data to the quantum Hilbert space (Fig. 1). A quantum computer can analyze the input data in this feature space, where a classifier, such as a linear support vector machine, can gain its power in finding a hyperplane to separate the data. In Refs. [7, 8], the quantum feature map is set as a fixed quantum circuit, and the adaptive training for classification tasks in ML is performed on a variational circuit that adapts the measurement basis. Here, the variational circuit is a set of quantum gates consisting of parameters which can be optimized via a hybrid quantum-classical optimization procedure [14]. The quantum feature map can also be trained to maximize the separation of mapped quantum states that originated from different data classes [15].

It is challenging to find a quantum feature map that is classically intractable but can substantially improve real-world ML tasks. Another interesting question is whether a quantum feature map can obtain the same or more powerful expressivity as classical ML schemes from the aspect of the universal approximation property (UAP) and the classification capability. Here, UAP refers to the ability to approximate any continuous function [16, 17]. The classification capability implies that the function constructed from quantum feature maps can form disjoint decision regions [18]. UAP and the classification capability have been extensively explored in feedforward classical neural networks [19–21]. It is then natural to derive the UAP of quantum neural networks employing qubits as quantum perceptrons with nonlinear excitation responses [22]. Such neural networks can be also emulated on a photonic quantum computer, which is expected to obtain the UAP [23]. Furthermore, it is conjectured that under a special kind of classical data preprocessing, the sequentially repeated quantum feature maps can become universal function approximators [24]. However, the study of UAP and classification capability in typical settings of the quantum feature map [7, 8] remains a challenge [25].

In this Letter, we formulate the universal approximation problem in terms of quantum feature maps. We present a comprehensive provable UAP and classification capability in two typical scenarios when setting the quantum feature map. Both of these scenarios have been utilized in prior QML applications [14, 24, 26]. In the

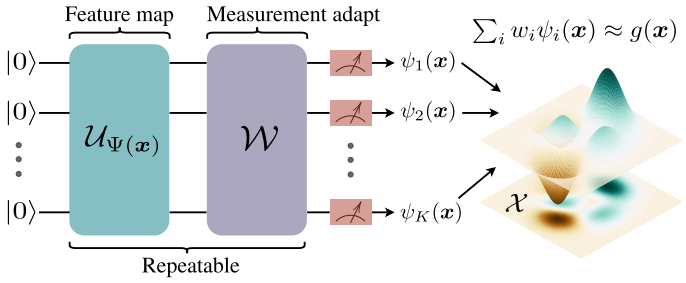

FIG. 1. A quantum feature framework consists of a feature map circuit $\mathcal{U}_{\Psi(\boldsymbol{x})}$ that realizes $\Psi(\boldsymbol{x})$ to map the classical data $\boldsymbol{x} \in \mathcal{X}$ to a quantum state in the Hilbert space, and a quantum circuit $\mathcal{W}$ to adapt the measurement basis. The combination of $\mathcal{U}_{\Psi(\boldsymbol{x})}$ and Ψ can be repeated as a sequence with different parameters. This framework has the universal approximation property if the linear combining of measurement results can approximate any continuous function $g : \mathcal{X} \to \mathbb{R}$.

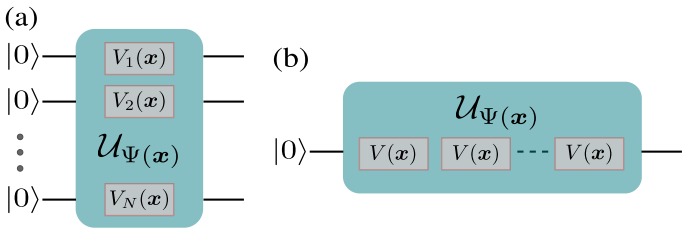

FIG. 2. The quantum circuit for a quantum feature map $U_{\Psi(\boldsymbol{x})}$. (a) The circuit is the tensor product of multiple circuits, where each circuit $V_i(\boldsymbol{x})$ acts on a subsystem. (b) The circuit is the repeated of a simple circuit $V(\boldsymbol{x})$ (for example, a single Pauli-Y rotation) acting on the same qubits.

first scenario, which is defined as the *parallel scenario*, the quantum feature map is a tensor product of multiple quantum circuits where each circuit acts on a subsystem and the number of qubits can be set freely [Fig. 2(a)]. In the second scenario, which is defined as the *sequential scenario*, the quantum feature map is the repetition of a simple fixed quantum circuit with the number of qubits being fixed [Fig. 2(b)]. We obtain the UAP in the first scenario and prove the UAP for the second in single-qubit circuits of the finite input space. Our work paves an intriguing theoretical layout to provide insights into the expressive power of QML techniques.

Quantum feature map.— We will now define the quantum feature map which has been mentioned in Refs. [7, 8]. Let $\mathcal{H}$ be a Hilbert space, $\mathcal{X} \subset \mathbb{R}^d$ be an input set. The quantum feature map $\Psi : \mathcal{X} \to \mathcal{H}$ is a procedure of input encoding that encodes some input $\boldsymbol{x} \in \mathcal{X}$ into a quantum feature state $|\Psi(\boldsymbol{x})\rangle \in \mathcal{H}$. This mapping action is equivalent to applying the quantum circuit $\mathcal{V}(\boldsymbol{x}) = \mathcal{U}_{\Psi(\boldsymbol{x})}$ to the initial state $|0\rangle^{\otimes N}$, where N is the number of qubits. A quantum classifier can be constructed from the quantum feature map by two approaches: the variational circuit approach and the kernel induced approach. In the former approach, a short-depth quantum circuit $\mathcal{W}$ is applied to the quantum feature state to adapt the measure basic [7, 8] (Fig. 1). The circuit $\mathcal{W}$ is often parameterized with the parameters are optimized during the training. The quantum circuit's output is measured to obtain a complex nonlinear output. This output can be represented as a linear combination of exponentially many nonlinear functions. In the kernel induced approach, the quantum computer estimates the inner product between quantum feature states giving rise to a kernel $\kappa(\boldsymbol{x}, \boldsymbol{x}') = \langle\Psi(\boldsymbol{x})|\Psi(\boldsymbol{x}')\rangle = \langle 0\ldots 0|\,\mathcal{V}^\dagger(\boldsymbol{x})\mathcal{V}(\boldsymbol{x}')\,|0\ldots 0\rangle$ to feed into classical kernel methods [8].

We unify two above approaches into a quantum framework combining quantum feature maps with the appropriate possible set of observables. We introduce a set of observables which are Hermitian operators $\boldsymbol{O} = \{O_1, O_2, \ldots, O_K\}$ applied to the quantum feature state $|\Psi(\boldsymbol{x})\rangle$. If we measure the operator O_i, we can obtain the expectation value of this operator and consider it as the ith basic function $\psi_i(\boldsymbol{x}) : \mathcal{X} \to \mathbb{R}$ which is defined as

$$\psi_i(\boldsymbol{x}) = \langle\Psi(\boldsymbol{x})|O_i|\Psi(\boldsymbol{x})\rangle = \mathrm{Tr}[O_i\,|\Psi(\boldsymbol{x})\rangle\,\langle\Psi(\boldsymbol{x})|]. \quad (1)$$

The key idea is that if these basic functions have nonlinearity properties with sufficient high dimension, we can solve a complex task by the linear regression on the output function $f : \mathcal{X} \to \mathbb{R}$, which is the linear combination of the basic functions

$$f(\boldsymbol{x}; \Psi, \boldsymbol{O}, \boldsymbol{w}) = \sum_{i=1}^{K} w_i\psi_i(\boldsymbol{x}). \quad (2)$$

Here, the weights $\boldsymbol{w} = \{w_i\}_{i=1}^K (w_i \in \mathbb{R})$ can be optimized via the training. This scheme is analogous with the classical extreme learning machine (ELM) framework. ELM was originally developed for the single-hidden-layer feedforward neural networks and then extended to multi-layer feedforward neural networks [19, 27]. In the ELM, the input data $\boldsymbol{x}$ is fed into a multi-layer perceptron where all weights between layers are fixed and selected randomly. The states of hidden nodes at some layer are regarded as basic functions that plays similar roles with $\psi_i(\boldsymbol{x})$.

The observables $\{O_i\}$ should be chosen for easy implementation in a quantum computer but can produce the nonlinearity with sufficient high dimensional basic functions. We note that the adaptive training on the variational circuits in Ref. [7, 8] plays a role in adapting the measurement based on the quantum feature space. Furthermore, the kernel induced by the inner product between quantum feature states in Ref. [8] can be regarded as a situation with the number of observables goes to infinity, where the optimization problems are performed over functions in a reproducing kernel Hilbert space. Therefore the expressive power of these well-known approaches depends on the expressive power of the linear combination of basic functions.

Universal approximation property and classification capability.— From the set of feature maps and the collection of observables, we define the quantum framework used in this study.

Definition 1. *Let $\mathcal{P}$ be a set of quantum feature maps from the input set $\mathcal{X}$ to the Hilbert space $\mathcal{H}$, $\mathcal{O}$ be the collection of possible observables on $\mathcal{H}$. A quantum feature framework $\mathcal{F}$ based on $(\mathcal{P}, \mathcal{O})$ is defined as the collection of function $f : \mathcal{X} \to \mathbb{R}$, where each f has the form $f(\boldsymbol{x}; \Psi, \boldsymbol{O}, \boldsymbol{w})$ in Eq. (2). Here, $\Psi \in \mathcal{P}, \boldsymbol{O} = \{O_1, \ldots, O_K\} \in \mathcal{O}$, and $\boldsymbol{w} = \{w_i\}_{i=0}^{K} (w_i \in \mathbb{R})$.*

Next, we define the UAP and classification capability of a given quantum feature framework.

Definition 2. *Let $\mathcal{G}$ be a space of continuous functions $g : \mathcal{X} \to \mathbb{R}$. A quantum feature framework $\mathcal{F}$ based on $(\mathcal{P}, \mathcal{O})$ has the universal approximation property (UAP) w.r.t. $\mathcal{G}$ and a norm $\|\cdot\|$ on $\mathcal{G}$ if given any function $g \in \mathcal{G}$, there exists a series of $f_n(\boldsymbol{x}) \in \mathcal{F}$ such that*

$$\lim_{n\to\infty} \|f_n(\cdot) - g(\cdot)\| = 0. \tag{3}$$

Definition 3. *Given a compact set $\mathcal{M}$ of $\mathbb{R}^d$. A quantum feature framework $\mathcal{F}$ based on $(\mathcal{P}, \mathcal{O})$ has the classification capability if for arbitrary disjoint regions, i.e., closed sets $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ in $\mathcal{M}$, there exist $f \in \mathcal{F}$ such that f can separate these regions [19].*

The classification capability is satisfied if we can approximate a continuous function in terms of the supremum norm [28], where the existence of this function is guaranteed by the following lemma ([19], Theorem 2.1).

Lemma 1. *([19], Lemma 2.1) Given disjoint regions $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m \subset \mathbb{R}^d$ with the corresponding m arbitrary real values $c_1, c_2, \ldots, c_m$, and an arbitrary region $\mathcal{E}$ disjointed from any K_i, there exists a continuous function $f_c(\boldsymbol{x})$ such that $f_c(\boldsymbol{x}) = c_i$ if $\boldsymbol{x} \in K_i$ and $f_c(\boldsymbol{x}) = c_0$ if $\boldsymbol{x} \in \mathcal{E}$, where c_0 is an arbitrary real value different from $c_1, c_2, \ldots, c_m$.*

We investigate the UAP and the classification capability in two typical scenarios in setting the quantum feature map. We assume that $\mathcal{X}$ is a compact set. For the sake of readability, we present some definitions for notations used in this study. Let $L^2(\mathcal{X})$ be a space of functions $g : \mathcal{X} \to \mathbb{R}$ that is square integrable, i.e., $\int_{\mathcal{X}} |g(\boldsymbol{x})|^2 d\boldsymbol{x} < \infty$. The norm of function g in $L^2(\mathcal{X})$ space is defined as $\|g\|_{L^2(\mathcal{X})} = \left[\int_{\mathcal{X}} |g(\boldsymbol{x})|^2 d\boldsymbol{x}\right]^{1/2}$.

Parallel scenario.— We examine the first scenario where the quantum feature map is a tensor product of multiple quantum circuits acting on subsystems with the number of qubits can be set freely [Fig. 2(a)]. We consider the class $\mathcal{P}_N$ of the quantum feature map $\Psi_N^{\mathcal{V}}$ via the following circuit applied to $|0\rangle^{\otimes N}$

$$\mathcal{V}_N(\boldsymbol{x}) = V_1(\boldsymbol{x}) \otimes V_2(\boldsymbol{x}) \otimes \ldots \otimes V_N(\boldsymbol{x}), \tag{4}$$

where $V_j(\boldsymbol{x})$ is the Y-basis rotation $e^{-i\theta_j(\boldsymbol{x})Y}$ applied to the jth qubit. Here, $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $Y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}$, and $Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ are the Pauli matrices give rise to the rotation operators on the Bloch sphere.

First, we investigate a special kind of data preprocessing, where the nonlinearity can be introduced by implementing an activation function into $\theta_j(\boldsymbol{x})$. Given an activation function $\sigma : \mathbb{R} \to [-1, 1]$, we consider $\theta_j(\boldsymbol{x}) = \arccos\left(\frac{1 + \sigma(\boldsymbol{a}_j \cdot \boldsymbol{x} + b_j)}{2}\right)$, where $\boldsymbol{a}_j \in \mathbb{R}^d, b_j \in \mathbb{R}$, and $\boldsymbol{a}_j \cdot \boldsymbol{x}$ denotes the inner product of vectors $\boldsymbol{a}_j$ and $\boldsymbol{x}$ in $\mathbb{R}^d$. We consider the observables on the following set of Hermitian operators $\boldsymbol{Z}_N = \{\boldsymbol{I} \otimes \ldots \otimes \underbrace{Z}_{j\text{-index}} \otimes \ldots \otimes \boldsymbol{I} \mid 1 \le j \le N\}$. The following theorem provides the sufficient condition for the activation function σ to obtain the UAP for our quantum feature map based on $(\mathcal{P}, \boldsymbol{O}_{\text{Pauli}})$, where $\mathcal{P} = \cup_{N=1}^{\infty} \mathcal{P}_N$ and $\boldsymbol{Z}_{\text{Pauli}} = \cup_{N=1}^{\infty} \boldsymbol{Z}_N$.

Theorem 1 (UAP when implementing activation functions in pre-processing)**.** *Given any nonconstant piecewise continuous function [29] $\sigma : \mathbb{R} \to [-1, 1]$ such that $\sigma(\boldsymbol{a} \cdot \boldsymbol{x} + b)$ $(\boldsymbol{a} \in \mathbb{R}^d, b \in \mathbb{R})$ is dense in $L^2(\mathcal{X})$, i.e., for any $\epsilon > 0$ and any $g \in L^2(\mathcal{X})$, there exist $\boldsymbol{a} \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $\|g - \sigma(\boldsymbol{a} \cdot \cdot + b)\|_{L^2(\mathcal{X})} < \epsilon$. For any continuous function $g : \mathcal{X} \to \mathbb{R}$ and any function sequence $\{\sigma_j(\boldsymbol{x}) = \sigma(\boldsymbol{a}_j \cdot \boldsymbol{x} + b_j)\}$ where $\boldsymbol{a}_j$ and b_j are randomly generated based on any continuous sampling distribution, the following statement holds with probability 1*

$$\lim_{N\to\infty} \min_{\boldsymbol{w}\in\mathbb{R}^N} \left\| f(\cdot; \Psi_N^{\mathcal{V}}, \boldsymbol{Z}_N, \boldsymbol{w}) - g(\cdot) \right\|_{L^2(\mathcal{X})} = 0. \tag{5}$$

Proof. From Eq. (1), we compute the jth basic function $\psi_j(\boldsymbol{x})$ as $\psi_j(\boldsymbol{x}) = \langle \Psi_N^{\mathcal{V}}(\boldsymbol{x}) | Z_j | \Psi_N^{\mathcal{V}}(\boldsymbol{x}) \rangle = \langle 0|^{\otimes N} \mathcal{V}_N^{\dagger}(\boldsymbol{x}) Z_j \mathcal{V}_N(\boldsymbol{x}) |0\rangle^{\otimes N} = \langle 0| e^{i\theta_j Y} Z e^{-i\theta_j Y} |0\rangle = 2\cos^2(\theta_j) - 1 = \sigma(\boldsymbol{a}_j \cdot \boldsymbol{x} + b_j) = \sigma_j(\boldsymbol{x})$. Then, $\left\| f(\boldsymbol{x}; \Psi_N^{\mathcal{V}}, \boldsymbol{Z}_N, \boldsymbol{w}) - g(\boldsymbol{x}) \right\|_{L^2(\mathcal{X})} = \left\| \sum_{j=1}^{N} w_j \sigma_j(\boldsymbol{x}) - g(\boldsymbol{x}) \right\|_{L^2(\mathcal{X})}$. Equation (5) is obtained from the main result in the UAP of the classical ELM framework in Ref. [21] (Theorem 2.3), which states that $\lim_{N\to\infty} \min_{\boldsymbol{w}\in\mathbb{R}^N} \left\| \sum_{j=1}^{N} w_j \sigma_j(\boldsymbol{x}) - g(\boldsymbol{x}) \right\|_{L^2(\mathcal{X})} = 0$.

Theorem 1 implies that with a sufficient number of qubits, the quantum feature map induced from nonlinear activation function with the selected observables can work as a universal approximator to approximate any continuous target function $g : \mathcal{X} \to \mathbb{R}$ in $L^2(\mathcal{X})$ with any arbitrary precision. We consider m disjoint regions $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ in $\mathcal{X}$ and their corresponding m distinct real values as labels $c_1, c_2, \ldots, c_m$ with the continuous function f_c according to lemma 1. We say that a function $h : \mathcal{X} \to \mathbb{R}$ can separate m disjoint regions $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ at $\boldsymbol{x}_0 \in \mathcal{X}$ if $|f_c(\boldsymbol{x}_0) - h(\boldsymbol{x}_0)| < \delta =$

$\frac{1}{2}\min\{|c_i - c_j| \mid \forall i \neq j\}$. From theorem 1, for arbitrary $\varepsilon > 0$, there exists $f : \mathcal{X} \to \mathbb{R}$ in the form of Eq. (2) such that $\left\|f_c(\boldsymbol{x}) - f(\boldsymbol{x})\right\|_{L^2(\mathcal{X})} < \varepsilon$. Let $\mathcal{Y} = \{\boldsymbol{y} \in \mathcal{X} \mid |f_c(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)| \geq \delta\}$ and $V_{\mathcal{Y}}$ be the volume of $\mathcal{Y}$, we have $V_{\mathcal{Y}}^{1/2}\delta < \varepsilon$ or $V_{\mathcal{Y}} < (\varepsilon/\delta)^2$. Therefore, by selecting the sufficiently small ε, we can reduce $V_{\mathcal{Y}}$ as small as possible to increase the classification capability of function f.

We note that the quantum feature map described above depends on implementing the activation function σ and its randomly generated function sequence $\{\sigma_j(\boldsymbol{x}) = \sigma(\boldsymbol{a}_j \cdot \boldsymbol{x} + b_j)\}$. Next, we examine the UAP of the quantum feature map in more simple form of the circuit $V_j(\boldsymbol{x})$ without relying on the pre-processing with activation function. Since $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$, without the loss of the generality, we assume that $\mathcal{X} = [0,1]^d$. Here, we consider the quantum feature map $\Psi_N^{\mathcal{V}}$ via the circuits in Eq. (4) with $V_j(\boldsymbol{x}) = e^{-i\arccos(\sqrt{x_k})Y}, 1 \leq k \leq d, j \equiv k (\mathrm{mod}\ d), (1 \leq j \leq N)$. We then prove the UAP for the quantum feature framework based on $\mathcal{P} = \cup_{N=1}^{\infty}\{\Psi_N^{\mathcal{V}}\}$ and the class of observables $\boldsymbol{Z}_{\boldsymbol{\alpha}} = Z^{\alpha_1} \otimes Z^{\alpha_2} \otimes \ldots \otimes Z^{\alpha_N}$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N) \in \{0,1\}^N$. The basic functions corresponding with the feature map $\Psi_N^{\mathcal{V}}$ and the observables $\boldsymbol{Z}_{\boldsymbol{\alpha}}$ are calculated as $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \langle\Psi_N^{\mathcal{V}}(\boldsymbol{x})|\boldsymbol{Z}_{\boldsymbol{\alpha}}|\Psi_N^{\mathcal{V}}(\boldsymbol{x})\rangle = \langle 0|^{\otimes N}\mathcal{V}_N^{\dagger}(\boldsymbol{x})\boldsymbol{Z}_{\boldsymbol{\alpha}}\mathcal{V}_N(\boldsymbol{x})|0\rangle^{\otimes N}$. From $\{\psi_{\boldsymbol{\alpha}}\}$, we can construct any polynomial function on $\mathcal{X}$ by using the following lemma (see proof in [30]).

Lemma 2. *Consider the polynomial $P(\boldsymbol{x})$ of the input $\boldsymbol{x} = (x_1, x_2, \ldots, x_d) \in [0,1]^d$ where the degree of x_j in $P(\boldsymbol{x})$ is less than or equal to $\dfrac{N + (d-j)}{d}$ for $j = 1, \ldots, d$, then there exists a collection of output weights $\{w_{\boldsymbol{\alpha}} \in \mathbb{R}^{2^N} \mid \boldsymbol{\alpha} \in \{0,1\}^N\}$ such that*

$$\sum_{\boldsymbol{\alpha}\in\{0,1\}^N} w_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = P(\boldsymbol{x}). \tag{6}$$

Lemma 2 implies that we can represent any polynomial function on $\mathcal{X}$ by using the defined quantum feature map and observables. Since any continuous function on $\mathcal{X}$ can be approximated by polynomial functions with arbitrary precision in terms of the supremum norm [31], we obtain the following UAP (see proof in [30]).

Theorem 2 (UAP in the parallel scenario)**.** *For any continuous function $g : \mathcal{X} \to \mathbb{R}$,*

$$\lim_{N\to\infty} \min_{\boldsymbol{w}\in\mathbb{R}^{2^n}} \left\| \sum_{\boldsymbol{\alpha}\in\{0,1\}^N} w_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) - g(\boldsymbol{x})\right\|_{\infty} = 0. \tag{7}$$

Theorem 2 implies that the induced quantum feature framework has the UAP w.r.t to the supremum norm. Furthermore, we prove the classification capability of this quantum feature framework. We consider the function f_c in lemma 1 with m disjoint regions $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$ in $\mathcal{X}$ and their corresponding m distinct real values as labels $c_1, c_2, \ldots, c_m$. From theorem 2, we can obtain the function $f : \mathcal{X} \to \mathbb{R}$ in the form $\sum_{\boldsymbol{\alpha}} w_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ such that $\|f(\boldsymbol{x}) - f_c(\boldsymbol{x})\|_{\infty} < \delta = \frac{1}{2}\min\{|c_i - c_j| \mid \forall i \neq j\}$. Therefore, $|f_c(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)| < \delta$ for all $\boldsymbol{x}_0 \in \mathcal{X}$, i.e., f can separate all disjoint regions $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_m$.

Sequential scenario.— In the parallel scenario, it is assumed that we can increase the number N of qubits to approximate the output function to a target continuous function with arbitrary precision. However, there is a limitation in implementing a quantum system with a large number of qubits in the current realistic model. Here, we investigate whether the UAP can be obtained by constructing the quantum feature map with only single qubit by repeating the same randomness-free and simple quantum circuit $V(\boldsymbol{x})$. Unlike the previous scenario, the quantum feature map described in the following paragraphs is not capable of approximating a function whose domain is an infinite set [30]. We restrict the input set to a finite set $\mathcal{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_M\}$, for example, in a real-world application, $\mathcal{X}$ can be the set of RGB fixed-size images.

To obtain the UAP, it is important to set the appropriate form of $V(\boldsymbol{x})$ [32]. We consider the unitary operator $V(\boldsymbol{x}) = e^{-\pi i\theta(\boldsymbol{x})Y}$ applied to the single qubit and establish the condition of $\theta(\boldsymbol{x}_1), \theta(\boldsymbol{x}_2), \ldots, \theta(\boldsymbol{x}_M)$ to obtain the UAP. The quantum feature map is constructed by repeating $V(\boldsymbol{x})$–that is, applying $V^n(\boldsymbol{x}) = e^{-n\pi i\theta(\boldsymbol{x})Y}$ ($n \in \mathbb{N}$) to $|0\rangle$, where $\theta : \mathcal{X} \to \mathbb{R}$. The corresponding basic function with the observable Z (Pauli-Z) becomes

$$\begin{aligned}\psi_n(\boldsymbol{x}) &= \langle 0|\,(V^n)^{\dagger}(\boldsymbol{x})ZV^n(\boldsymbol{x})\,|0\rangle = 2\cos^2(\pi n\theta(\boldsymbol{x})) - 1 \\ &= \cos(2\pi n\theta(\boldsymbol{x})) = \cos(2\pi\{n\theta(\boldsymbol{x})\}),\end{aligned} \tag{8}$$

where $\{n\theta(\boldsymbol{x})\} = n\theta(\boldsymbol{x}) - \lfloor n\theta(\boldsymbol{x})\rfloor$ is the fractional part of $n\theta(\boldsymbol{x})$. The output function is modeled by the output weight $w \in \mathbb{R}$ as $f_n(\boldsymbol{x}; w) = w\psi_n(\boldsymbol{x}) = w\cos(2\pi\{n\theta(\boldsymbol{x})\})$. The UAP is studied via the density of the fractional parts $(\{n\theta(\boldsymbol{x}_1)\}, \ldots, \{n\theta(\boldsymbol{x}_M)\})_{n\in\mathcal{N}}$. We prove in [30] the following lemma which can be directly derived from Kronecker–Weyl theorem [33, 34].

Lemma 3. *If real numbers $1, a_1, a_2, \ldots, a_M$ are linearly independent over the field $\mathcal{Q}$ of rational numbers [35], then $(\{na_1\}, \ldots, \{na_M\})_{n\in\mathcal{N}}$ is dense in $[0,1)^M$, i.e., given $(t_1, t_2, \ldots, t_M) \in [0,1)^M$, for every $\varepsilon > 0$, there exist $n \in \mathcal{N}$ such that $|\{na_i\} - t_i| < \varepsilon$ for $i = 1, 2, \ldots, M$.*

Now, we consider the case when $1, \theta(\boldsymbol{x}_1), \ldots, \theta(\boldsymbol{x}_M)$ are linearly independent over $\mathcal{Q}$. Given a continuous real-valued function g on $\mathcal{X}$, we define real numbers $\beta_1, \ldots, \beta_M$, such that $\beta_i = g(\boldsymbol{x}_i)/\beta$, where $\beta = 1 + \max_{i=1,\ldots,M}|g(\boldsymbol{x}_i)|$. Since $-1 < \beta_i < 1$, there exists $\gamma_i \in (0,1)$ such that $\beta_i = \cos(2\pi\gamma_i)$ for $i = 1, \ldots, M$. From lemma 3, for every $\varepsilon > 0$, there exists $n \in \mathcal{N}$ such that $|\{n\theta(\boldsymbol{x}_i)\} - \gamma_i| < \varepsilon/(2\pi\beta)$ for $i = 1, \ldots, M$. If we choose $w = \beta$, then $|f_n(\boldsymbol{x}_i; w) - g(\boldsymbol{x}_i)| = \beta|\psi_n(\boldsymbol{x}_i) - \beta_i| =$

$\beta|\cos(2\pi\{n\theta(\boldsymbol{x}_i)\})-\cos(2\pi\gamma_i)| \leq 2\pi\beta|\{n\theta(\boldsymbol{x}_i)\}-\gamma_i| < \varepsilon$ for $i = 1, \ldots, M$. We obtain the following UAP theorem which states that with the condition of the linearly independence for $1, \theta(\boldsymbol{x}_2), \ldots, \theta(\boldsymbol{x}_M)$, any function in $\mathcal{X}$ can be approximated by repeatedly applying the unitary operator $V(\boldsymbol{x})$ with an appropriate iteration number n.

Theorem 3 (UAP in the sequential scenario)**.** *If $1, \theta(\boldsymbol{x}_1), \ldots, \theta(\boldsymbol{x}_M)$ are linearly independent over $\mathbb{Q}$, then for any function $g : \mathcal{X} \to \mathbb{R}$ and for any $\epsilon > 0$, there exists $n \in \mathbb{N}$ such that $|f_n(\boldsymbol{x}) - g(\boldsymbol{x})| < \epsilon$ for $\forall \boldsymbol{x} \in \mathcal{X}$.*

As similar to the analysis in the parallel scenario, we can also obtain the classification capability in the sequential scenario via theorem 3.

Conclusion.— We present a comprehensive understanding of the UAP of the quantum feature framework induced from quantum feature maps. This study lays a foundation for further theoretical analysis on the expressivity of these quantum frameworks and can provide insights into the design of a good expressive model in the QML applications. The suggestions in practical applications are left for future works, such as finding an efficient scheme with the lowest implementation cost to obtain the necessary approximation accuracy.

* goto.takahiro.2020@gmail.com
† tran_qh@ai.u-tokyo.ac.jp
‡ k_nakajima@mech.t.u-tokyo.ac.jp

[1] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature **549**, 195 (2017).
[2] J. Preskill, Quantum **2**, 79 (2018).
[3] G. Torlai and R. G. Melko, Annu. Rev. Condens. Matter Phys. **11**, 325 (2020).
[4] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, Nat. Commun. **5** (2014), 10.1038/ncomms5213.
[5] E. Farhi and H. Neven, Preprint at arXiv:1802.06002 (2018).
[6] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Nat. Commun. **11** (2020), 10.1038/s41467-020-14454-2.
[7] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019).
[8] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019).
[9] C. Blank, D. K. Park, J.-K. K. Rhee, and F. Petruccione, Npj Quantum Inf. **6** (2020), 10.1038/s41534-020-0272-6.
[10] K. Fujii and K. Nakajima, Phys. Rev. Applied **8**, 024030 (2017).
[11] K. Nakajima, K. Fujii, M. Negoro, K. Mitarai, and M. Kitagawa, Phys. Rev. Applied **11**, 034021 (2019).
[12] J. Chen and H. I. Nurdin, Quantum Inf. Process. **18** (2019), 10.1007/s11128-019-2311-9.
[13] Q. H. Tran and K. Nakajima, Preprint at arXiv:2006.08999 (2020).
[14] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Phys. Rev. A **98**, 032309 (2018).
[15] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, Preprint at arXiv:2001.03622 (2020).
[16] G. Cybenko, Math. Control. Signal **2**, 303 (1989).
[17] K. Hornik, Neural Networks **4**, 251 (1991).
[18] W. Y. Huang and R. P. Lippmann, in *Neural Information Processing Systems*, edited by D. Z. Anderson (American Institute of Physics, 1988) pp. 387–396.
[19] G.-B. Huang, Y.-Q. Chen, and H. Babri, IEEE Trans. Neural Netw. **11**, 799 (2000).
[20] G.-B. Huang, L. Chen, C. K. Siew, *et al.*, IEEE Trans. Neural Netw. **17**, 879 (2006).
[21] G.-B. Huang and L. Chen, Neurocomputing **70**, 3056 (2007).
[22] E. Torrontegui and J. J. García-Ripoll, EPL (Europhysics Letters) **125**, 30004 (2019).
[23] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Phys. Rev. Research **1**, 033063 (2019).
[24] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, Quantum **4**, 226 (2020).
[25] In preparing this manuscript, we notice a very recent and independent work with our study in characterizing the variational quantum model in terms of a partial Fourier series in the data [36]. The asymptotical universality then is investigated via the realization of arbitrary frequency spectra under the repetitions of simple data-encoding gates or large enough dimension of the feature map.
[26] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Phys. Rev. A **101**, 032308 (2020).
[27] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, Neurocomputing **70**, 489 (2006).
[28] The supremum norm of a function h from $\mathcal{X}$ to $\mathbb{R}$ is defined as $\|h\|_\infty = \sup_{\boldsymbol{x}\in\mathcal{X}} |h(\boldsymbol{x})|$.
[29] A function $h : \mathbb{R} \to \mathbb{R}$ is said to be piecewise continuous if it has only a finite number of discontinuities in any interval, and its left and right limits are defined (not necessarily equal) at each discontinuity.
[30] See Supplemental Materials.
[31] This is "Weierstrass's polynomial approximation theorem", which is a special case of Stone-Weierstrass's theorem [37].
[32] It is easy to show a counter-example of $V(\boldsymbol{x})$ that we cannot obtain the UAP. For example, if $V(\boldsymbol{x})$ is a unitary operator such that $|0\rangle$ is a common eigenvector of $V(\boldsymbol{x}_1)$ and $V(\boldsymbol{x}_2)$, we denote λ_1 and λ_2 as the corresponding eigenvalues of $V(\boldsymbol{x}_1)$ and $V(\boldsymbol{x}_2)$, respectively. Then the quantum feature maps at $\boldsymbol{x}_1, \boldsymbol{x}_2$ after repeating $V(\boldsymbol{x})$ for n times are $\Psi(\boldsymbol{x}_1) = V(\boldsymbol{x}_1)^n |0\rangle = \lambda_1^n |0\rangle, \Psi(\boldsymbol{x}_2) = V(\boldsymbol{x}_2)^n |0\rangle = \lambda_2^n |0\rangle$. For an arbitrary set of observable operator $\boldsymbol{O} = \{O_1, O_2, \ldots, O_K\}$, the value of the basic functions at $\boldsymbol{x}_1, \boldsymbol{x}_2$ are $\psi_i(\boldsymbol{x}_1) = \langle\Psi(\boldsymbol{x}_1)| O_i |\Psi(\boldsymbol{x}_1)\rangle = |\lambda_1|^{2n} \langle 0| O_i |0\rangle$, $\psi_i(\boldsymbol{x}_2) = \langle\Psi(\boldsymbol{x}_1)| O_i |\Psi(\boldsymbol{x}_2)\rangle = |\lambda_2|^{2n} \langle 0| O_i |0\rangle$. Since $V(\boldsymbol{x}_1)$ and $V(\boldsymbol{x}_2)$ are unitary, then $|\lambda_1| = |\lambda_2| = 1$ and $\psi_i(\boldsymbol{x}_1) = \psi_i(\boldsymbol{x}_2)$ for all i. Therefore, any linear combination of $\{\psi_i(\boldsymbol{x})\}$ cannot approximate the function which takes different values at $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$.
[33] J. L. King, Am. Math. Mon. **101**, 609 (1994).
[34] E. M. Stein and R. Shakarchi, *Fourier analysis: an introduction*, Vol. 1 (Princeton University Press, 2011) pp. 105–108.
[35] Real numbers $b_1, b_2, \ldots, b_L$ are linearly independent over

$\mathcal{Q}$ if the only integral solution to $\alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_L b_M = 0$ is the all zero $\alpha_1 = \alpha_2 = \ldots = \alpha_L = 0$.

[36] M. Schuld, R. Sweke, and J. J. Meyer, Preprint at arXiv:2008.08605 (2020).

[37] K. Yoshida, *Functional Analysis* (Springer-Verlag Berlin Heidelberg, 1980) pp. 9–10.

Supplementary Material for "Universal Approximation Property of Quantum Feature Map"

Takahiro Goto*
Nagase Hongo Building F5, 5-24-5, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Quoc Hoan Tran† and Kohei Nakajima‡
Graduate School of Information Science and Technology,
The University of Tokyo, Tokyo 113-8656, Japan
(Dated: September 2, 2020)

This supplementary material describes in detail the calculations introduced in the main text. The equation and figure numbers in this section are prefixed with S (e.g., Eq. (S1), Fig. (S1)), while numbers without the prefix (e.g., Eq. (1), Fig. (1)) refer to items in the main text.

PROOF OF LEMMA 2

LEMMA 2

Consider the polynomial $P(\boldsymbol{x})$ of the input $\boldsymbol{x} = (x_1, x_2, \ldots, x_d) \in [0,1]^d$ where the degree of x_j in $P(\boldsymbol{x})$ is less than or equal to $\dfrac{N+(d-j)}{d}$ for $j = 1, \ldots, d$ $(N \geq d)$, then there exists a collection of output weights $\{w_{\boldsymbol{\alpha}} \in \mathbb{R}^{2^N} \mid \boldsymbol{\alpha} \in \{0,1\}^N\}$ such that

$$\sum_{\boldsymbol{\alpha} \in \{0,1\}^N} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = P(\boldsymbol{x}).$$

Proof.— Consider a real number r, we denote $\lfloor r \rfloor$ as the greatest integer less than or equal to r, and $\lceil r \rceil$ as the least integer greater than or equal to r. For integers i, d $(d > 0)$, let $[i]$ denote the integer number k such that $1 \leq k \leq d$, and $i \equiv k (\mathrm{mod}\ d)$. For a nonzero monomial $m(\boldsymbol{x}) = l x_1^{a_1} x_2^{a_2} \cdots x_d^{a_d}$, let $\deg(m(\boldsymbol{x})) = a_1 + a_2 + \cdots a_d$ be the degree of $m(\boldsymbol{x})$. We define $\deg(0) = -1$ for our convenience. Furthermore, for a nonzero polynomonial $P(\boldsymbol{x})$, the degree $\deg(P(\boldsymbol{x}))$ of $P(\boldsymbol{x})$ is defined as the largest degree of monomial terms in $P(\boldsymbol{x})$.

The basic functions induced from $\mathcal{V}_N(\boldsymbol{x}) = V_1(\boldsymbol{x}) \otimes V_2(\boldsymbol{x}) \otimes \ldots \otimes V_N(\boldsymbol{x})$ and $\boldsymbol{Z}_{\boldsymbol{\alpha}} = Z^{\alpha_1} \otimes Z^{\alpha_2} \otimes \ldots \otimes Z^{\alpha_N}$, where $V_j(\boldsymbol{x}) = e^{-i \arccos(\sqrt{x_{[j]}}) Y}, (1 \leq j \leq N)$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N) \in \{0,1\}^N$, are represented as follow:

$$\begin{aligned}
\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) &= \langle 0|^{\otimes N} \mathcal{V}_N^\dagger(\boldsymbol{x}) \boldsymbol{Z}_{\boldsymbol{\alpha}} \mathcal{V}_N(\boldsymbol{x}) |0\rangle^{\otimes N} \\
&= \langle 0|^{\otimes N} \left(V_1(\boldsymbol{x}) \otimes V_2(\boldsymbol{x}) \otimes \ldots \otimes V_N(\boldsymbol{x})\right)^\dagger \left(Z^{\alpha_1} \otimes Z^{\alpha_2} \otimes \cdots \otimes Z^{\alpha_N}\right) \left(V_1(\boldsymbol{x}) \otimes V_2(\boldsymbol{x}) \otimes \ldots \otimes V_N(\boldsymbol{x})\right) |0\rangle^{\otimes N} \\
&= \prod_{i=1}^{N} \langle 0| V_i^\dagger(\boldsymbol{x}) Z_i^{\alpha_i} V_i(\boldsymbol{x}) |0\rangle \\
&= \prod_{i=1}^{N} \left(2 \left(\arccos\left(\cos \sqrt{x_{[i]}}\right)\right)^2 - 1\right)^{\alpha_i} \\
&= \prod_{i=1}^{N} (2x_{[i]} - 1)^{\alpha_i} = 2^{|\boldsymbol{\alpha}|} x_{[1]}^{\alpha_1} x_{[2]}^{\alpha_2} \cdots x_{[N]}^{\alpha_N} + Q_{\boldsymbol{\alpha}}(\boldsymbol{x}).
\end{aligned} \tag{S1}$$

Here, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N) \in \{0,1\}^N$, $|\boldsymbol{\alpha}| = \alpha_1 + \alpha_2 + \ldots + \alpha_N$, and $Q_{\boldsymbol{\alpha}}(\boldsymbol{x})$ is the total of other monomial terms in $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ with the degree less than $|\boldsymbol{\alpha}|$.

For each j $(1 \leq j \leq d)$, we find the number of integer i such that $1 \leq i \leq N$ and $[i] \equiv j (\mathrm{mod}\ d)$. It is equivalent to finding the number of integer p such that $1 - j \leq dp \leq N - j$ (then $i = j + dp$). Since $1 - j \geq 1 - d > -d$, then $0 \leq dp < N - j + 1$, or $0 \leq p < \dfrac{N-j+1}{d}$. Therefore the number of such p is $\lceil \dfrac{N-j+1}{d} \rceil$. From this result and Eq. (S1), the degree of x_j $(1 \leq j \leq d)$ in $\psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$ is not larger than $\lceil \dfrac{N-j+1}{d} \rceil = \lfloor \dfrac{N+(d-j)}{d} \rfloor$.

Next, to prove lemma 2, we prove that if $P(\boldsymbol{x})$ is the polymonial where the degree of x_j in $P(\boldsymbol{x})$ is less than or equal to $\lfloor \dfrac{N+(d-j)}{d} \rfloor$, then $P(\boldsymbol{x})$ can be represented as the form $\sum_{\boldsymbol{\alpha}} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$. We present the proof by induction technique on the degree of $P(\boldsymbol{x})$. The statement is trivial for the case $\deg(P(\boldsymbol{x})) = -1, 0$. We assume that the statement is established if $\deg(P(\boldsymbol{x})) = -1, 0, \ldots, D-1$ (induction assumption), then we need to establish the truth of the statement for the case $\deg(P(\boldsymbol{x})) = D > 0$ (induction hypothesis). Let $m_1(\boldsymbol{x}), m_2(\boldsymbol{x}), \ldots, m_K(\boldsymbol{x})$ denote all monomial terms in $P(\boldsymbol{x})$ such that $\deg(m_k)(\boldsymbol{x}) = \deg(P(\boldsymbol{x}))$ for all k $(1 \leq k \leq K)$. Then, $P(\boldsymbol{x})$ can be represented as

$$P(\boldsymbol{x}) = m_1(\boldsymbol{x}) + m_2(\boldsymbol{x}) + \ldots + m_K(\boldsymbol{x}) + P'(\boldsymbol{x}), \tag{S2}$$

where $\deg(P'(\boldsymbol{x})) < \deg(P(\boldsymbol{x}))$.

For each k $(1 \leq k \leq K)$, let $\boldsymbol{\alpha}^{(k)}$ denote the value of $\boldsymbol{\alpha}$ such that $m_k(\boldsymbol{x}) = w_{\boldsymbol{\alpha}} 2^{|\boldsymbol{\alpha}|} x_{[1]}^{\alpha_1} x_{[2]}^{\alpha_2} \cdots x_{[N]}^{\alpha_N}$. We consider the difference polynomial $R(\boldsymbol{x})$ between $P(\boldsymbol{x})$ and $\sum_{k=1}^{K} w_{\boldsymbol{\alpha}^{(k)}} \psi_{\boldsymbol{\alpha}^{(k)}}(\boldsymbol{x})$ as

$$R(\boldsymbol{x}) = P(\boldsymbol{x}) - \sum_{k=1}^{K} w_{\boldsymbol{\alpha}^{(k)}} \psi_{\boldsymbol{\alpha}^{(k)}}(\boldsymbol{x}) = P'(\boldsymbol{x}) - \sum_{k=1}^{K} w_{\boldsymbol{\alpha}^{(k)}} Q_{\boldsymbol{\alpha}^k}(\boldsymbol{x}). \tag{S3}$$

Since $\deg(R(\boldsymbol{x})) < \deg(P(\boldsymbol{x})) = D$ and the degree of x_j in $R(\boldsymbol{x})$ is not larger than $\lfloor \frac{N+(d-j)}{d} \rfloor$, following the induction assumption, $R(\boldsymbol{x})$ can be written in the form $\sum_{\boldsymbol{\alpha}} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$. From this result and Eq. (S3), $P(\boldsymbol{x})$ is also represented by the form $\sum_{\boldsymbol{\alpha}} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x})$. Therefore, the induction hypothesis is proved for $\deg(P(\boldsymbol{x})) = D$. The statement in lemma 2 is established for all values of $\deg(P(\boldsymbol{x}))$.

PROOF OF THEOREM 2

THEOREM 2

For any continuous function $g : \mathcal{X} \to \mathbb{R}$,

$$\lim_{N\to\infty} \min_{\boldsymbol{w}\in\mathbb{R}^{2^n}} \Big\| \sum_{\boldsymbol{\alpha}\in\{0,1\}^N} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) - g(\boldsymbol{x}) \Big\|_\infty = 0.$$

Theorem 2 is obtained from lemma 2 and the Weierstrass's polynomial approximation theorem, which states that any continuous function on $\mathcal{X}$ can be approximated by polynomial functions with arbitrary precision in terms of the supremum norm. Let us mention here the general form of the Weierstrass's polynomial approximation theorem, which is known as the Stone–Weierstrass theorem [1].

Theorem S1 (Stone–Weierstrass [1]) *Let $\mathcal{X}$ be a compact space and $C(\mathcal{X})$ be the set of real-valued continuous functions defined on $\mathcal{X}$. Let a subset B of $C(\mathcal{X})$ satisfy the three conditions:*

(i) If $f, g \in B$, then the product $f \cdot g$ and linear combination $\alpha f + \beta g$, with real coefficients α, β, belong to B.

(ii) The constant function 1 belongs to B.

(iii) The uniform limit f_∞ of any sequence f_n of functions in B also belongs to B.

Then $B = C(\mathcal{X})$ iff B separates the points of $\mathcal{X}$, i.e. iff, for every distinct points $\boldsymbol{x}, \boldsymbol{y}$ in $\mathcal{X}$, there exists a function h in B such that $h(\boldsymbol{x}) \neq h(\boldsymbol{y})$.

Let $\mathcal{P}$ denote the set of all polynomials on the compact set $\mathcal{X} \subset \mathbb{R}^d$. We apply theorem S1 for the uniform closure $\bar{\mathcal{P}}$, which is the space of all functions that can be approximated by a sequence of uniformly-converging polynomials in $\mathcal{P}$. We check three conditions given in theorem S1. From the definition of $\bar{\mathcal{P}}$, the conditions $(ii), (iii)$ are satisfied. Let us check the condition (i). For arbitrary functions $f, g \in \bar{\mathcal{P}}$, there exist sequences of polynomials $\{f_n\}, \{g_n\} \in \mathcal{P}$, such that $\lim_{n\to\infty} \|f_n - f\|_\infty = 0$ and $\lim_{n\to\infty} \|g_n - g\|_\infty = 0$. We define two functions $p = f \cdot g$, and $q = \alpha f + \beta g$, and two sequences $\{p_n\}$ and $\{q_n\}$, where $p_n = f_n \cdot g_n$ and $q_n = \alpha f_n + \beta g_n$. Because f and g are defined on the compact set $\mathcal{X}$, there are the maximum values and minimum values for f and g. Let M be a real number such that $\|f\|_\infty \leq M$ and $\|g\|_\infty \leq M$, then

$$\begin{aligned} \|p_n - p\|_\infty &= \|f_n \cdot g_n - f \cdot g\|_\infty \\ &= \|(f_n - f) \cdot g + (g_n - g) \cdot f_n\|_\infty \\ &\leq \|f_n - f\|_\infty \|g\|_\infty + \|g_n - g\|_\infty \|f_n\|_\infty \\ &\leq M\|f_n - f\|_\infty + M\|g_n - g\|_\infty. \end{aligned} \tag{S4}$$

Similarly, we have the following inequality

$$\begin{aligned} \|q_n - q\|_\infty &= \|(\alpha f_n + \beta g_n) - (\alpha f + \beta g)\|_\infty \\ &= \|\alpha(f_n - f) + \beta(g_n - g)\|_\infty \\ &\leq |\alpha| \|f_n - f\|_\infty + |\beta| \|g_n - g\|_\infty. \end{aligned} \tag{S5}$$

Equations (S4)(S5) imply that $\lim_{n\to\infty} \|p_n - p\|_\infty = 0$ and $\lim_{n\to\infty} \|q_n - q\|_\infty = 0$, therefore $p \in \mathcal{P}$ and $q \in \mathcal{P}$; the condition (i) is satisfied.

Furthermore, we consider arbitrary distinct points $\boldsymbol{x} = (x_1, \ldots, x_d), \boldsymbol{y} = (y_1, \ldots, y_d) \in \mathcal{X}$. Without the loss of generality, we assume that $x_1 \neq y_1$. We consider a polynomial h in $\mathcal{P}$ ($h \in \bar{\mathcal{P}}$) such that $h(\boldsymbol{z}) = z_1 - x_1$ for all

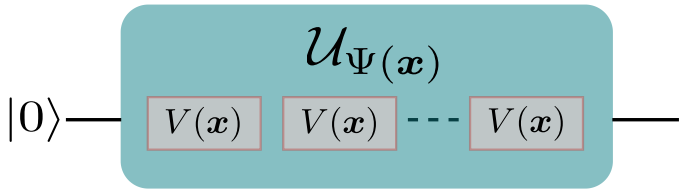

FIG. S1. The sequential scenario in setting the quantum feature map. The circuit is the repeated of a simple circuit $V(\boldsymbol{x})$ (for example, a single Pauli-Y rotation) acting on the same qubits.

$\boldsymbol{z} = (z_1, \ldots, z_d) \in \mathcal{X}$. Then, $h(\boldsymbol{x}) = 0$ and $h(\boldsymbol{y}) = y_1 - x_1 \neq 0$. Therefore, $\bar{\mathcal{P}}$ can separate the points of $\mathcal{X}$. From the Stone–Weierstrass theorem, we obtain $\bar{\mathcal{P}} = C(\mathcal{X})$, or $\mathcal{P}$ is dense in $C(\mathcal{X})$.

Now, let g be a continuous real-valued function defined on $\mathcal{X}$. Since $\mathcal{P}$ is dense in $C(\mathcal{X})$, there exists a sequence of polynomials $P_k(\boldsymbol{x}) \in \mathcal{P}$ such that $\lim_{k\to\infty} \|g(\boldsymbol{x}) - P_k(\boldsymbol{x})\|_\infty = 0$. Then, for any $\varepsilon > 0$, there exists k_0 such that $\|P_k(\boldsymbol{x}) - g(\boldsymbol{x})\|_\infty < \varepsilon$ for all $k \geq k_0$. We choose the number N_{k_0} of qubits such that the degree of x_j in $P_{k_0}(\boldsymbol{x})$ is less than or equal to $\dfrac{N + (d-j)}{d}$ for all $N \geq N_{k_0}$. From lemma 2, for $N \geq N_{k_0}$, there exists a collection of output weights $\{w_{\boldsymbol{\alpha}} \in \mathbb{R}^{2^N} \mid \boldsymbol{\alpha} \in \{0,1\}^N\}$ such that $\sum_{\boldsymbol{\alpha}\in\{0,1\}^N} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) = P_k(\boldsymbol{x})$. Then,

$$\| \sum_{\boldsymbol{\alpha}\in\{0,1\}^N} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) - g(\boldsymbol{x})\|_\infty < \varepsilon, \tag{S6}$$

for $N \geq N_{k_0}$. Therefore, we have the following result of theorem 2

$$\lim_{N\to\infty} \min_{\boldsymbol{w}\in\mathbb{R}^{2^n}} \Big\| \sum_{\boldsymbol{\alpha}\in\{0,1\}^N} w_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) - g(\boldsymbol{x}) \Big\|_\infty = 0.$$

UAP OF THE SEQUENTIAL SCENARIO ON AN INFINITE SET

In the sequential scenario (Fig. S1), we investigate whether the UAP can be obtained by constructing the quantum feature map with only single qubit by repeating the same randomness-free and simple quantum circuit. In this scenario, we consider the unitary operator $V(\boldsymbol{x}) = e^{-\pi i \theta(\boldsymbol{x}) Y}$ applied to the single qubit. The quantum feature map is constructed by repeating $V(\boldsymbol{x})$–that is, applying $V^n(\boldsymbol{x}) = e^{-n\pi i \theta(\boldsymbol{x}) Y}$ $(n \in \mathbb{N})$ to $|0\rangle$, where $\theta : \mathcal{X} \to \mathbb{R}$. We consider the observable Z (Pauli-Z) on the state after applied the quantum feature map. Therefore, the corresponding basic function is

$$\begin{aligned} \psi_n(\boldsymbol{x}) &= \langle 0| \left(V^n\right)^\dagger(\boldsymbol{x}) Z V^n(\boldsymbol{x}) |0\rangle \\ &= 2\cos^2(\pi n\theta(\boldsymbol{x})) - 1 = \cos(2\pi\{n\theta(\boldsymbol{x})\}), \end{aligned} \tag{S7}$$

where $\{n\theta(\boldsymbol{x})\} = n\theta(\boldsymbol{x}) - \lfloor n\theta(\boldsymbol{x})\rfloor$ is the fractional part of $n\theta(\boldsymbol{x})$. The output function is modeled by the output weight $w \in \mathbb{R}$ as $f_n(\boldsymbol{x}; w) = w\psi_n(\boldsymbol{x}) = w\left(\cos(2\pi\{n\theta(\boldsymbol{x})\}) - 1\right)$.

We prove that the quantum feature map described above is not capable of approximating a function whose domain is an infinite set. Here, for simplicity we consider $\mathcal{X}$ as a closed interval $\mathcal{X} = [a, b] \subset \mathbb{R}$, where $b > a$, and g is an arbitrary continuous real-valued function defined on $[a, b]$. We assume that θ is a continuous real-valued function defined on $[a, b]$. If $\theta(a) = \theta(b)$, any basic function ψ constructed from any observable takes the same value on a and b, i.e., $\psi(a) = \psi(b)$, therefore the linear combination of these basic functions cannot approximate g if $g(a) \neq g(b)$. Therefore, we only need to consider the case $\theta(a) \neq \theta(b)$ to investigate the UAP. Without the loss of generality, we assume that $\theta(a) < \theta(b)$. Because the set of rational numbers are a dense subset of the real numbers, there exist ration numbers $q_1, q_2 \in \mathbb{Q}$ such that $\theta(a) < q_1 < q_2 < \theta(b)$. From the intermediate value theorem, there exist $c_1, c_2 \in [a, b]$ such that $\theta(c_1) = q_1, \theta(c_2) = q_2$. If we write $q_1 = n_1/s, q_2 = n_2/s$ for $n_1, n_2, s \in \mathbb{Z}$, then the fractional sequence $(\{n\theta(c_1)\}, \{n\theta(c_2)\})_{n\in\mathbb{N}}$ has the period s and takes the finite number of values. Therefore, $(\psi_n(c_1), \psi_n(c_2))_{n\in\mathbb{N}}$ also takes the finite number of values.

We assume that the quantum feature framework defined by this setting has the UAP. Let choose a continuous function $g : [a, b] \to \mathbb{R}$ such that $g(c_1) = kg(c_2) > 0$ for $k > 1$. For $\varepsilon_k = g(c_2)/k$, there exist $w_k \in \mathbb{R}$ and $n_k \in \mathcal{N}$ such

that $|f_{n_k}(c_1; w_k) - g(c_1)| < \varepsilon_k$ and $|f_{n_k}(c_2; w_k) - g(c_2)| < \varepsilon_k$. Therefore, we have the following inequalities

$$g(c_1) - \varepsilon_k < f_{n_k}(c_1; w_k) = w_k \psi_{n_k}(c_1) < g(c_1) + \varepsilon_k, \tag{S8}$$
$$g(c_2) - \varepsilon_k < f_{n_k}(c_2; w_k) = w_k \psi_{n_k}(c_2) < g(c_2) + \varepsilon_k. \tag{S9}$$

Since $g(c_1) - \varepsilon_k = (k - \frac{1}{k})g(c_2) > 0$, $g(c_2) - \varepsilon_k = (1 - \frac{1}{k})g(c_2) > 0$, and $g(c_2) + \varepsilon_k = (1 + \frac{1}{k})g(c_2) > 0$, we have

$$\frac{\psi_{n_k}(c_1)}{\psi_{n_k}(c_2)} = \frac{w_k \psi_{n_k}(c_1)}{w_k \psi_{n_k}(c_2)} > \frac{g(c_1) - \varepsilon_k}{g(c_2) + \varepsilon_k} = \frac{k^2 - 1}{k + 1}. \tag{S10}$$

Therefore, $\dfrac{\psi_{n_k}(c_1)}{\psi_{n_k}(c_2)}$ can take infinite number of values as k varies ($k > 1$). However, this is impossible because $(\psi_n(c_1), \psi_n(c_2))_{n \in \mathbb{N}}$ can only takes the finite number of values. In other words, we were wrong to assume the quantum feature framework defined by this setting has the UAP. We can conclude that the setting of the sequential scenario mentioned in the main script cannot provide the UAP on an infinite set $[a, b]$. The same statement can also be obtained if the input set $\mathcal{X}$ is a connected infinite subset of $\mathbb{R}^d$.

PROOF OF LEMMA 3

LEMMA 3

If real numbers $1, a_1, a_2, \ldots, a_M$ are linearly independent over the field $\mathcal{Q}$ of rational numbers [2], then $(\{na_1\}, \ldots, \{na_M\})_{n \in \mathcal{N}}$ is dense in $[0, 1)^M$, i.e., given $(t_1, t_2, \ldots, t_M) \in [0, 1)^M$, for every $\varepsilon > 0$, there exists $n \in \mathcal{N}$ such that $|\{na_i\} - t_i| < \varepsilon$ for $i = 1, 2, \ldots, M$.

Proof.— We prove the lemma from scratch. The proof for the special case when $M = 1$ (which is known as Wely's equidistribudion theorem) can be found in some analysis textbooks, for example, in Ref. [3].

Let $\boldsymbol{a} = (a_1, a_2, \ldots, a_M) \in \mathbb{R}^M$, $\boldsymbol{k} = (k_1, k_2, \ldots, k_M) \in \mathbb{Z}^M$, $\mathbf{0} = (0, 0, \ldots, 0) \in \mathbb{Z}^M$, and $\boldsymbol{k} \cdot \boldsymbol{a}$ denote the dot product between $\boldsymbol{k}$ and $\boldsymbol{a}$ as $\boldsymbol{k} \cdot \boldsymbol{a} = k_1 a_1 + k_2 a_2 + \ldots + k_M a_M$. From the linearly independent property over $\mathcal{Q}$ of $1, a_1, a_2, \ldots, a_M$, we have

$$\mathrm{e}^{2\pi i \boldsymbol{k} \cdot \boldsymbol{a}} = 1 \Leftrightarrow \boldsymbol{k} \cdot \boldsymbol{a} \in \mathbb{Z} \Leftrightarrow k_1 = k_2 = \ldots = k_M = 0. \tag{S11}$$

Therefore,

$$\frac{1}{M} \sum_{n=1}^{M} \mathrm{e}^{2\pi i n \boldsymbol{k} \cdot \boldsymbol{a}} = \begin{cases} \dfrac{\mathrm{e}^{2\pi i \boldsymbol{k} \cdot \boldsymbol{a}}}{M} \dfrac{1 - \mathrm{e}^{2\pi i M \boldsymbol{k} \cdot \boldsymbol{a}}}{1 - \mathrm{e}^{2\pi i \boldsymbol{k} \cdot \boldsymbol{a}}} \to 0 \quad (\text{as } M \to \infty), & \text{if } \boldsymbol{k} \neq \mathbf{0}. \\ 1, & \text{if } \boldsymbol{k} = \mathbf{0}. \end{cases} \tag{S12}$$

Then,

$$\lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} \mathrm{e}^{2\pi i n \boldsymbol{k} \cdot \boldsymbol{a}} = \int_{\mathbb{T}^M} \mathrm{e}^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}} \, d\boldsymbol{x}, \tag{S13}$$

where $\mathbb{T} = [0, 1), \mathbb{T}^M = [0, 1)^M$. Let consider a trigonometric polynomial $p : \mathbb{T}^M \to \mathbb{C}$,

$$p(\boldsymbol{x}) = \sum_{\boldsymbol{k} \in \mathcal{I}} \hat{p}_{\boldsymbol{k}} \, \mathrm{e}^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}}, \hat{p}_{\boldsymbol{k}} \in \mathbb{C}, \mathcal{I} \subset \mathbb{Z}^M, |\mathcal{I}| < \infty, \tag{S14}$$

supported on an arbitrary frequency index set $\mathcal{I}$ of finite cardinality. From (S13), we obtain

$$\lim_{M \to \infty} \frac{1}{M} \sum_{n=1}^{M} p(n\boldsymbol{a}) = \int_{\mathbb{T}^M} p(\boldsymbol{x}) d\boldsymbol{x}. \tag{S15}$$

We apply the Weierstrass's polynomial approximation theorem for the case of trigonometric polynomials [1]. For an arbitrary continuous real-valued function f defined on $\mathbb{T}^M$, there exists a sequence of trigonometric polynomial

$\{p_j(\boldsymbol{x})\}_{j\in\mathbb{N}}$ uniformly converging to $f(\boldsymbol{x})$. It means that for every $\varepsilon > 0$, there exists a $j_0 \in \mathbb{N}$ such that $\|f - p_j\|_\infty < \varepsilon$ for all $j \geq j_0$. Therefore, for $j \geq j_0$, we have

$$\left|\frac{1}{M}\sum_{n=1}^{M} f(n\boldsymbol{a}) - \int_{\mathbb{T}^M} f(\boldsymbol{x})d\boldsymbol{x}\right| \tag{S16}$$

$$= \left|\frac{1}{M}\sum_{n=1}^{M} f(n\boldsymbol{a}) - \frac{1}{M}\sum_{n=1}^{M} p_j(n\boldsymbol{a}) + \frac{1}{M}\sum_{n=1}^{M} p_j(n\boldsymbol{a}) - \int_{\mathbb{T}^M} p_j(\boldsymbol{x})d\boldsymbol{x} + \int_{\mathbb{T}^M} p_j(\boldsymbol{x})d\boldsymbol{x} - \int_{\mathbb{T}^M} f(\boldsymbol{x})d\boldsymbol{x}\right| \tag{S17}$$

$$\leq \left|\frac{1}{M}\sum_{n=1}^{M} f(n\boldsymbol{a}) - \frac{1}{M}\sum_{n=1}^{M} p_j(n\boldsymbol{a})\right| + \left|\frac{1}{M}\sum_{n=1}^{M} p_j(n\boldsymbol{a}) - \int_{\mathbb{T}^M} p_j(\boldsymbol{x})d\boldsymbol{x}\right| + \left|\int_{\mathbb{T}^M} p_j(\boldsymbol{x})d\boldsymbol{x} - \int_{\mathbb{T}^M} f(\boldsymbol{x})d\boldsymbol{x}\right| \tag{S18}$$

$$\leq \frac{1}{M}\sum_{n=1}^{M} ||f - p_j|| + \left|\frac{1}{M}\sum_{n=1}^{M} p_j(n\boldsymbol{a}) - \int_{\mathbb{T}^M} p_j(\boldsymbol{x})d\boldsymbol{x}\right| + \int_{\mathbb{T}^M} ||p_j - f||d\boldsymbol{x} \tag{S19}$$

$$\leq 2\epsilon + \left|\frac{1}{M}\sum_{n=1}^{M} p_j(n\boldsymbol{a}) - \int_{\mathbb{T}^M} p_j(\boldsymbol{x})d\boldsymbol{x}\right|. \tag{S20}$$

From Eq. (S15) and Eq. (S20), we have $\lim_{M\to\infty}\left|\frac{1}{M}\sum_{n=1}^{M} f(n\boldsymbol{a}) - \int_{\mathbb{T}^M} f(\boldsymbol{x})d\boldsymbol{x}\right| \leq 2\varepsilon$ for all $\varepsilon > 0$. Then, we have

$$\lim_{M\to\infty}\frac{1}{M}\sum_{n=1}^{M} f(n\boldsymbol{a}) = \int_{\mathbb{T}^M} f(\boldsymbol{x})d\boldsymbol{x}. \tag{S21}$$

Let consider a distance ρ on $\mathbb{T}^M$, and $B(\boldsymbol{y}, r) = \{\boldsymbol{x} \in \mathbb{T}^M \mid \rho(\boldsymbol{y}, \boldsymbol{x}) \leq r\}$ for each $\boldsymbol{y} \in \mathbb{T}^M$ and $r > 0$. We define the characteristic function of $B(\boldsymbol{y}, r)$ as

$$\chi_{B(\boldsymbol{y},r)}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \rho(\boldsymbol{y}, \boldsymbol{x}) \leq r \\ 0 & \text{otherwise.} \end{cases} \tag{S22}$$

For any $\epsilon > 0$, we define the following continuous real-valued functions on $\mathbb{T}^M$

$$f_+^{(\epsilon)}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \rho(\boldsymbol{y}, \boldsymbol{x}) \leq r \\ 1 - \dfrac{\rho(\boldsymbol{y}, \boldsymbol{x}) - r}{\epsilon} & \text{if } r < \rho(\boldsymbol{y}, \boldsymbol{x}) \leq r + \epsilon) \\ 0 & \text{otherwise,} \end{cases} \tag{S23}$$

$$f_-^{(\epsilon)}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \rho(\boldsymbol{y}, \boldsymbol{x}) \leq r - \epsilon \\ \dfrac{r - \rho(\boldsymbol{y}, \boldsymbol{x})}{\epsilon} & \text{if } r - \epsilon < \rho(\boldsymbol{y}, \boldsymbol{x}) \leq r \\ 0 & \text{otherwise.} \end{cases} \tag{S24}$$

Then,

$$\frac{1}{M}\sum_{n=1}^{M} f_-^{(\epsilon)}(n\boldsymbol{a}) \leq \frac{1}{M}\sum_{n=1}^{M} \chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) \leq \frac{1}{M}\sum_{n=1}^{M} f_+^{(\epsilon)}(n\boldsymbol{a}). \tag{S25}$$

Since $f_+^{(\epsilon)}, f_-^{(\epsilon)}$ are continuous, from Eq. (S21), we have

$$\lim_{M\to\infty}\frac{1}{M}\sum_{n=1}^{M} f_-^{(\epsilon)}(n\boldsymbol{a}) = \int_{\mathbb{T}^M} f_-^{(\epsilon)}(\boldsymbol{x})d\boldsymbol{x}, \tag{S26}$$

$$\lim_{M\to\infty}\frac{1}{M}\sum_{n=1}^{M} f_+^{(\epsilon)}(n\boldsymbol{a}) = \int_{\mathbb{T}^M} f_+^{(\epsilon)}(\boldsymbol{x})d\boldsymbol{x}. \tag{S27}$$

Therefore, from Eq. (S25) and Eqs. (S26)(S27), we obtain the following relations

$$\int_{\mathbb{T}^M} f_-^{(\epsilon)}(\boldsymbol{x})d\boldsymbol{x} \leq \liminf_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) \leq \limsup_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) \leq \int_{\mathbb{T}^M} f_+^{(\epsilon)}(\boldsymbol{x})d\boldsymbol{x}. \tag{S28}$$

We take $\epsilon \to 0$, then $\int_{\mathbb{T}^M} f_\pm^{(\epsilon)}(\boldsymbol{x})d\boldsymbol{x} \to \int_{\mathbb{T}^M} \chi_{B(\boldsymbol{y},r)}(\boldsymbol{x})d\boldsymbol{x}$, and

$$\liminf_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) = \limsup_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) = \int_{\mathbb{T}^M} \chi_{B(\boldsymbol{y},r)}(\boldsymbol{x})d\boldsymbol{x}. \tag{S29}$$

Therefore, $\lim_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) = \int_{\mathbb{T}^M} \chi_{B(\boldsymbol{y},r)}(\boldsymbol{x})d\boldsymbol{x}$. Since $\int_{\mathbb{T}^M} \chi_{B(\boldsymbol{y},r)}(\boldsymbol{x})d\boldsymbol{x} > 0$ with $r > 0$, then $\lim_{M\to\infty} \frac{1}{M}\sum_{n=1}^{M} \chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) > 0$. Therefore, with an arbitrary $\boldsymbol{y} \in \mathbb{T}^M$ and an arbitrary $r > 0$, there exists $n \in \mathbb{N}$ such that $\chi_{B(\boldsymbol{y},r)}(n\boldsymbol{a}) > 0$, or $n\boldsymbol{a} \in B(\boldsymbol{y}, r)$. This result proves that $\{n\boldsymbol{a} \in \mathbb{T}^M \mid n \in \mathbb{N}\}$ is dense in $\mathbb{T}^M$.

* goto.takahiro.2020@gmail.com
† k09tranhoan@gmail.com
‡ k_nakajima@mech.t.u-tokyo.ac.jp

[1] K. Yoshida, *Functional Analysis* (Springer-Verlag Berlin Heidelberg, 1980) pp. 9–10.
[2] Real numbers $b_1, b_2, \ldots, b_L$ are linearly independent over $\mathcal{Q}$ if the only integral solution to $\alpha_1 b_1 + \alpha_2 b_2 + \ldots + \alpha_L b_M = 0$ is the all zero $\alpha_1 = \alpha_2 = \ldots = \alpha_L = 0$.
[3] E. M. Stein and R. Shakarchi, *Fourier analysis: an introduction*, Vol. 1 (Princeton University Press, 2011) pp. 105–108.