COMMUN. MATH. SCI.
Vol. 20, No. 5, pp. 1397–1408
© 2022 International Press

UNIVERSAL APPROXIMATION OF SYMMETRIC AND ANTI-SYMMETRIC FUNCTIONS*

JIEQUN HAN†, YINGZHOU LI‡, LIN LIN§, JIANFENG LU¶,
JIEFU ZHANG‖, AND LINFENG ZHANG**

Abstract. We consider universal approximations of symmetric and anti-symmetric functions, which are important for applications in quantum physics, as well as other scientific and engineering computations. We give constructive approximations with explicit bounds on the number of parameters with respect to the dimension and the target accuracy ϵ. While the approximation still suffers from the curse of dimensionality, to the best of our knowledge, these are the first results in the literature with explicit error bounds for functions with symmetry or anti-symmetry constraints.

Keywords. Universal approximation; Symmetric function; Anti-symmetric function; Neural Network; Vandermonde determinant; Quantum many-body problem.

AMS subject classifications. 41A25; 41A29; 41A63.

1. Introduction

We consider in this study universal approximations of symmetric and anti-symmetric functions. A function $f:(\mathbb{R}^d)^N\to\mathbb{R}$ is **(totally) symmetric** if

$$f(\mathbf{x}_{\sigma(1)},\ldots,\mathbf{x}_{\sigma(N)})=f(\mathbf{x}_1,\ldots,\mathbf{x}_N), \tag{1.1}$$

for any permutation $\sigma\in S(N)$, and elements $\mathbf{x}_1,\ldots,\mathbf{x}_N\in\mathbb{R}^d$. Similarly f is **(totally) anti-symmetric** if

$$f(\mathbf{x}_{\sigma(1)},\ldots,\mathbf{x}_{\sigma(N)})=(-1)^\sigma f(\mathbf{x}_1,\ldots,\mathbf{x}_N), \tag{1.2}$$

for any permutation, where $(-1)^\sigma$ is the signature of σ. Note that the permutation is only applied to the outer indices $1,\ldots,N$ (also referred to as the "particle" indices later), but not the Cartesian indices $1,\ldots,d$ for each $\mathbf{x}_i$. In other words, f is not totally symmetric/anti-symmetric when viewed as a function on $\mathbb{R}^{d\times N}$. More precisely, f in (1.2) lies in the anti-symmetric inner product of N copies of $L^2(\mathbb{R}^d)$. This is the relevant setup in many applications in scientific and engineering computation. A totally symmetric function is also called a permutation invariant function. A closely related concept is the permutation equivariant mapping, which is of the form $Y:(\mathbb{R}^d)^N\to(\mathbb{R}^{\tilde{d}})^N$ that satisfies

$$Y_i(\mathbf{x}_{\sigma(1)},\ldots,\mathbf{x}_{\sigma(N)})=Y_{\sigma(i)}(\mathbf{x}_1,\ldots,\mathbf{x}_N),\qquad i=1,\ldots,N \tag{1.3}$$

*Received: August 03, 2021; Accepted (in revised form): December 02, 2021. Communicated by Lexing Ying.

†Department of Mathematics, Princeton University, Princeton, NJ 08544, USA; Center for Computational Mathematics, Flatiron Institute, New York, NY 10010, USA (jiequnhan@gmail.com).

‡School of Mathematical Sciences, Fudan University, Shanghai, China 200433 (yingzhouli@fudan.edu.cn).

§Department of Mathematics, University of California, Berkeley, CA 94720, USA; Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA (linlin@math.berkeley.edu).

¶Department of Mathematics, Department of Physics, and Department of Chemistry, Duke University, Durham, NC 27708, USA (jianfeng@math.duke.edu).

‖Department of Mathematics, University of California, Berkeley, CA 94720, USA (jiefuzhang@berkeley.edu).

**Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA (linfengz@princeton.edu).

for any permutation $\sigma \in S(N)$, and $\mathbf{x}_1,\ldots,\mathbf{x}_N \in \mathbb{R}^d$. Here each component $Y_i \in \mathbb{R}^{\tilde{d}}$, and $\tilde{d}$ can be different from d.

Perhaps the most important example of totally symmetric and anti-symmetric functions is the wavefunction of identical particles in quantum mechanics. The indistinguishability of identical particles implies that their wavefunctions should be either totally symmetric or totally anti-symmetric upon exchanging the variables associated with any two of the particles, corresponding to two categories of particles: bosons and fermions. The former can share quantum states, giving rise to, e.g., the celebrated Bose-Einstein condensate; while the latter cannot share quantum states as described by the famous Pauli exclusion principle. Such exchange/permutation symmetry also arises from applications other than identical particles in quantum mechanics, mostly in the form of symmetric functions. For instance, in chemistry and materials science, the interatomic potential energy should be invariant under the permutation of the atoms of the same chemical species. Another example is in computer vision, where the classification of point clouds should not depend on the ordering of points.

The dimension of symmetric and anti-symmetric functions is usually large in practice because it is proportional to the number of considered elements. This means that in computation the notorious difficulty of "curse of dimensionality" is often encountered when dealing with such functions. Recent years have witnessed compelling success of neural networks in representing high-dimensional symmetric functions with great accuracy and efficiency, see, e.g., [3, 25, 28, 29] for interatomic potential energy, [20, 21] for 3D classification and segmentation of point sets, and [10, 30] for solutions of partial differential equations. For anti-symmetric functions, some recent work in the past few years [5, 11, 12, 15, 17, 19, 26] has shown exciting potential of solving the many-electron Schrödinger equation with neural networks. Within the framework of variational Monte Carlo (VMC), for some benchmark systems, the anti-symmetric wavefunction parameterized by neural networks can be on a par with the state-of-the-art wavefunctions constructed based on chemical or physical knowledge.

Despite the empirical success of neural networks approximating symmetric and anti-symmetric functions, theoretical understanding of these approximations is still limited. There are numerous results (see e.g. [2,6,13]) concerning the universal approximation of general continuous functions on compact domains. Nevertheless, if the target function is symmetric or anti-symmetric, it is much less investigated whether one can achieve the universal approximation with a class of functions with the same symmetry constraints. Explicitly guaranteeing the symmetric or anti-symmetric property of an ansatz is often mandatory. For example, for electronic systems, if the wavefunction is not constrained within the space of anti-symmetric functions, the resulting variational energy could be lower than the exact ground-state energy and would be no longer physically meaningful. From a machine learning perspective, symmetries can also significantly reduce the number of effective degrees of freedom, improve the efficiency of training, and enhance the generalizability of the model (see e.g. [8]). However, one needs to first make sure that the function class is still universal and sufficiently expressive.

The universal approximations of symmetric functions were partially studied in [27]. However, as will be illustrated in Section 2, the proof of [27] only holds for the case when $d=1$. Moreover, there is no error estimation provided for the proposed approximation. The more recent work [24] considered the universal approximation of permutation invariant functions and equivariant mappings for $d=1$ as well. The work [23] gives a generalization bound that is improved by a factor of $\sqrt{N!}$ with the introduction of the quotient feature space, but the result only applies for $d=1$. By respecting the permu-

tation symmetry, the resulting neural network involves much fewer parameters than the corresponding dense neural networks. The recent works [1, 7] on polynomial approximation of symmetric functions display a reduction of the curse of dimensionality when the target function has low-order many-body expansions or sparse polynomial approximates. The work [31] studies learning symmetric functions from a different perspective by treating symmetric functions (of any size) as functions over probability measures. But such a setting is not always applicable to practical problems. Similar to the case of symmetric functions, when $d=1$, any anti-symmetric polynomial can be factorized as the product of a Vandermonde determinant and a totally symmetric polynomial. Such a universal representation has been known since Cauchy [4]. However, to our knowledge, there is no such simple factorization for $d>1$. For $d>1$, the FermiNet [19] achieves good numerical results by using a linear combination of generalized Slater Determinants, and [14, Theorem 14] shows the universality of the FermiNet Ansatz again based on a global lexicographic ordering for $d>1$.

In this paper we aim to study the universal approximation of general symmetric and anti-symmetric functions for any $d\geq 1$. We now summarize the main results of the paper: First, for the symmetric function, we give two different proofs of the universality of the ansatz proposed in [27], both with explicit error bounds. The first proof is based on the Ryser formula [22] for permanents, and the second is based on the partition of the state space, as elaborated in Section 2 and Section 3, respectively. Moreover, we also show in Section 4 that, for the general anti-symmetric function with elements in any dimension, a simple ansatz combining Vandermonde determinants and the ansatz for symmetric functions is universal, with similar explicit error bounds as in the symmetric case. For readers' convenience, we summarize below in two theorems the ansatz we proved with the universal approximation for symmetric and anti-symmetric functions. The approximation rate only relies on a weak condition that the gradient is uniformly bounded.

We conclude in Section 5 with some practical considerations and future directions for further investigation. The proofs of Theorems 1.1 and 1.2 are given in Sections 3 and 4, respectively.

THEOREM 1.1 (Approximation to symmetric functions). *Let $f:\Omega^N\to\mathbb{R}$ be a continuously differentiable, totally symmetric function, where Ω is a compact subset of $\mathbb{R}^d$. Let $0<\epsilon<\|\nabla f\|_2\sqrt{Nd}N^{-\frac{1}{d}}$, here $\|\nabla f\|_2=\max_X\|\nabla f(X)\|_2$. Then there exist $g:\mathbb{R}^d\to\mathbb{R}^M$, $\phi:\mathbb{R}^M\to\mathbb{R}$, such that for any $X=(\mathbf{x}_1,\ldots,\mathbf{x}_N)\in\Omega^N$,*

$$\left|f(X)-\phi\left(\sum_{j=1}^N g(\mathbf{x}_j)\right)\right|\leq\epsilon,$$

where M, the number of feature variables, satisfies the bound

$$M\leq 2^N(\|\nabla f\|_2^2Nd)^{Nd/2}/(\epsilon^{Nd}N!). \tag{1.4}$$

THEOREM 1.2 (Approximation to anti-symmetric functions). *Let $f:\Omega^N\to\mathbb{R}$ be a continuously differentiable, totally anti-symmetric function, where Ω is a compact subset of $\mathbb{R}^d$. Then there exist K permutation equivariant mappings $Y^k:(\mathbb{R}^d)^N\to\mathbb{R}^N$, and permutation invariant functions $U^k:(\mathbb{R}^d)^N\to\mathbb{R}$, $k=1,\ldots,K$, such that for any $X=(\mathbf{x}_1,\ldots,\mathbf{x}_N)\in\Omega^N$,*

$$\left|f(X)-\sum_{k=1}^K U^k(X)\prod_{i<j}(y_i^k(X)-y_j^k(X))\right|<\epsilon,$$

where K satisfies the bound

$$K \leq (\|\nabla f\|_2^2 N d)^{Nd/2}/(\epsilon^{Nd} N!).$$

For each U^k, there exists $g^k : \mathbb{R}^d \to \mathbb{R}^m$, $\phi^k : \mathbb{R}^m \to \mathbb{R}$ with $m \leq 2^N$, such that for any $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N) \in \Omega^N$,

$$U^k(X) = \phi^k \left(\sum_{j=1}^{N} g^k(\mathbf{x}_j) \right).$$

We remark that for the bounds in both theorems above, $\|\nabla f\|_2$ may also scale with respect to N. For the examples listed below, we assume the domain for the function to be a hypercube $\Omega_N = [0,1]^N$. In the symmetric case, if f is in the simple form $f(X) = \sum_{i=1}^{N} g(\mathbf{x}_i)$, $\|\nabla f\|_2$ will scale as $\mathcal{O}(\sqrt{N})$. In the antisymmetic case, if f is a Slater determinant, the integral on the global domain $\int_{\Omega^N} \|\nabla f(X)\|_2^2 dX$ (which can be interpreted as the kinetic energy of the quantum system) is expected to scale as $\mathcal{O}(N)$. Hence $\|\nabla f(X)\|_2$ scales as $\mathcal{O}(\sqrt{N})$ at a given X on average. Both the simple examples above have $\mathcal{O}(\sqrt{N})$ scaling, but the exact scaling will depend on the choice of function f. We also remark that Theorem 1.2 only relies on a local lexicographic ordering procedure, instead of a global sorting operation.

2. Totally symmetric functions

Let $X = (\mathbf{x}_1, \ldots, \mathbf{x}_N)$, with $\mathbf{x}_i \in \Omega := [0,1]^d$. Consider a totally symmetric function $f(X)$. It is proved in [27] that when $d = 1$ (therefore $\Omega = [0,1]$), the following universal approximation representation holds

$$f(X) = \phi \left(\sum_{j=1}^{N} g(x_j) \right) \tag{2.1}$$

for continuous functions $g : \mathbb{R} \to \mathbb{R}^M$, and $\phi : \mathbb{R}^M \to \mathbb{R}$. For completeness, we briefly recall the proof.

Let $\mathcal{X} = \{(x_1, \ldots, x_N) \in \Omega^N = \mathbb{R}^N : x_1 \leq x_2 \leq \cdots \leq x_N\}$. Define the mapping $E : \mathcal{X} \to \mathcal{Z} \subset \mathbb{R}^{N+1}$, with each component function defined as

$$z_q = E_q(X) := \sum_{n=1}^{N} (x_n)^q, \quad q = 0, 1, \ldots, N.$$

The mapping E is a homeomorphism between $\mathcal{X}$ and its image in $\mathbb{R}^{N+1}$ [27]. Hence, if we let $g : \mathbb{R} \to \mathbb{R}^{N+1}, x \mapsto [1, x, x^2, \ldots, x^N]$ and $\phi : \mathbb{R}^{N+1} \to \mathbb{R}, Z \mapsto f(E^{-1}(Z))$, then we have

$$f(X) = \phi \left(\sum_{j=1}^{N} g(x_j) \right).$$

Here the number of feature variables is $M = N + 1$ by construction. The main difficulty associated with this construction is that the mapping E^{-1} can be arbitrarily complex to be approximated in practice. In fact the construction is similar in flavor to the Kolmogorov-Arnold representation theorem [16], which provides a universal representation for multivariable continuous functions, but without any *a priori* guarantee of the accuracy with respect to the number of parameters.

In order to generalize to the case $d>1$, the proof of [27, Theorem 9] in fact suggested an alternative proof for the case $d=1$ as follows. Using the Stone-Weierstrass theorem, a totally symmetric function can be approximated by a polynomial. After symmetrization, this polynomial becomes a totally symmetric polynomial. By the fundamental theorem of symmetric polynomials [18], any symmetric polynomial can be represented by a polynomial of elementary symmetric polynomials. In other words, for any symmetric polynomial P, we have

$$P(X)=Q(e_1(X),\dots,e_N(X)),$$

where Q is some polynomial, and the elementary polynomials e_k are defined as

$$e_k(X)=\sum_{1\le j_1<j_2<\cdots<j_k\le N} x_{j_1}x_{j_2}\cdots x_{j_k}.$$

Using the Newton-Girard formula, an elementary symmetric polynomial can be represented with power sums by

$$e_k(X)=\frac{1}{k!}\begin{vmatrix} E_1(X) & 1 & 0 & 0 & \cdots & 0 \\ E_2(X) & E_1(X) & 2 & 0 & \cdots & 0 \\ E_3(X) & E_2(X) & E_1(X) & 3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ E_{k-1}(X) & E_{k-2}(X) & E_{k-3}(X) & E_{k-4}(X) & \cdots & k-1 \\ E_k(X) & E_{k-1}(X) & E_{k-2}(X) & E_{k-3}(X) & \cdots & E_1(X) \end{vmatrix}. \tag{2.2}$$

Now let g be the same function defined in the previous proof, and define ϕ in terms of Q and the determinant computation. We can obtain a polynomial approximation in the form of

$$f(x)=P(X)+\epsilon=\phi\left(\sum_{j=1}^{N} g(x_j)\right)+\epsilon.$$

Here the error ϵ is due to the Stone-Weierstrass approximation. Letting $\epsilon\to 0$ we obtain the desired representation.

However, it is in fact not straightforward to extend the two proofs above to the case $d>1$. For the first proof, we can not define an ordered set $\mathcal{X}$, when each $\mathbf{x}_i\in\mathbb{R}^d$, to define the homeomorphism E. For the second proof, in the case $d>1$, a monomial (before symmetrization) takes the form

$$\prod_{i=1}^{N}\prod_{\alpha=1}^{d}\mathbf{x}_{i,\alpha}^{\gamma_{i,\alpha}}, \quad \gamma_{i,\alpha}\in\mathbb{N}.$$

Note that $f(X)$ is only symmetric with respect to the particle index i, but not the component index α. Hence the symmetrized monomial is *not* a totally symmetric function with respect to all variables. Therefore the fundamental theorem of symmetric polynomials does not apply.

Below we prove that the representation (2.1) indeed holds for any $d\ge 1$, and therefore we complete the proof of [27]. For technical reasons to be illustrated below, and without loss of generality, we shift the domain Ω and assume $\mathbf{x}_i\in\Omega:=[1,2]^d$. Following the Stone-Weierstrass theorem and after symmetrization, $f(X)$ can be approximated

by a symmetric polynomial. Every symmetric polynomial can be written as the linear combination of symmetrized monomials of the form

$$\sum_{\sigma\in S(N)}\prod_{i=1}^{N}\prod_{\alpha=1}^{d}\mathbf{x}_{\sigma(i),\alpha}^{\gamma_{i,\alpha}} := \sum_{\sigma\in S(N)}\prod_{i=1}^{N} f_i(\mathbf{x}_{\sigma(i)}) := \mathrm{perm}([f_i(\mathbf{x}_j)]).$$

Here $f_i(\mathbf{x}_j)=\prod_{\alpha=1}^{d}\mathbf{x}_{j,\alpha}^{\gamma_{i,\alpha}}$, $[f_i(\mathbf{x}_j)]$ denotes an $N\times N$ matrix whose entry in the i-th row and j-th column is $f_i(\mathbf{x}_j)$, and $\mathrm{perm}(A)$ stands for the permanent of square matrix A.

Following the Ryser formula [22] for representing a permanent (noting that permanent is invariant under transposition), we have

$$\begin{aligned}\mathrm{perm}([f_i(\mathbf{x}_j)]) &= (-1)^N \sum_{S\subseteq\{1,\dots,N\}} (-1)^{|S|}\prod_{i=1}^{N}\sum_{j\in S} f_j(\mathbf{x}_i)\\ &= (-1)^N \sum_{S\subseteq\{1,\dots,N\}} (-1)^{|S|} e^{\sum_{i=1}^{N}\log\left(\sum_{j\in S} f_j(\mathbf{x}_i)\right)}.\end{aligned}$$

Here we have used that $f_j(\mathbf{x})>0$ for all $\mathbf{x}\in\Omega$. Now we write down the approximation using a symmetric polynomial, which is a linear combination of L symmetrized monomials

$$\begin{aligned}f(X)-\epsilon = P(X) &= \sum_{l=1}^{L} c^{(l)}\mathrm{perm}([f_i^{(l)}(\mathbf{x}_j)])\\ &= (-1)^N\sum_{l=1}^{L} c^{(l)} \sum_{S\subseteq\{1,\dots,N\}} (-1)^{|S|} e^{\sum_{i=1}^{N}\log\left(\sum_{j\in S} f_j^{(l)}(\mathbf{x}_i)\right)}\end{aligned}$$

Define $g:\mathbb{R}^d\to\mathbb{R}^{L2^N}$ with each component function

$$g_S^{(l)}(\mathbf{x}) = \log\left(\sum_{j\in S} f_j^{(l)}(\mathbf{x})\right).$$

Then we define $\phi:\mathbb{R}^{L2^N}\to\mathbb{R}$ given by

$$\phi(Y) = (-1)^N\sum_{l=1}^{L} c^{(l)} \sum_{S\subseteq\{1,\dots,N\}} (-1)^{|S|} e^{Y_S^{(l)}},$$

where $Y_S^{(l)}$ is the S-th component of $g^{(l)}$. We now have an approximation of the target totally symmetric function in the desired form

$$f(X) = \phi\left(\sum_{j=1}^{N} g(\mathbf{x}_j)\right)+\epsilon,$$

and we finish the proof. Here the number of feature variables is $M=L2^N$, where L is the number of symmetrized monomials used in the approximation.

3. Totally symmetric function, revisited

In this section, we prove Theorem 1.1 for any $d \geq 1$. In particular, our proof is more explicit and does not rely on the Stone-Weierstrass theorem. The main idea is to partition the space into a lattice and use piecewise-constant functions to approximate the target permutation invariant function.

Again, without loss of generality, we assume $\mathbf{x}_i \in [0,1]^d := \Omega$. We then partition the domain Ω into a lattice $\mathbb{L}$ with grid size δ along each direction. Due to symmetry, we can assign a lexicographical order $\preceq$ to all lattice points $\mathbf{z}_i \in \mathbb{L}$. That is, $\mathbf{z}_1 \preceq \mathbf{z}_2$ if $\mathbf{z}_{1,\alpha} < \mathbf{z}_{2,\alpha}$ for the first α where $\mathbf{z}_{1,i}$ and $\mathbf{z}_{2,i}$ differs, $\alpha = 1,2,\cdots,d$. We define the tensor product of the N copies of the lattice $\mathbb{L}$ as $\mathbb{L}^N$, and a wedge of $\mathbb{L}^N$ is defined accordingly as

$$\bigwedge^N \mathbb{L} := \{Z = (\mathbf{z}_1, \ldots, \mathbf{z}_N) | \mathbf{z}_1 \preceq \mathbf{z}_2 \cdots \preceq \mathbf{z}_N\}.$$

For each $Z \in \bigwedge^N \mathbb{L}$, a corresponding union of boxes in Ω^N can be written as

$$B^{Z;\delta} = \bigcup_{\sigma \in S(N)} \{X \mid \mathbf{x}_i = \mathbf{z}_{\sigma(i)} + \delta \mathbf{u}_i, \quad \mathbf{u}_i \in [0,1]^d\}.$$

By construction, the piecewise-constant approximation to the target permutation invariant function is then

$$f(X) = \sum_{Z \in \bigwedge^N \mathbb{L}} f(Z) \mathbb{1}_{B^{Z;\delta}}(X) + \epsilon.$$

We assume that the derivative $\nabla_X f(X)$ is uniformly bounded for $X \in \Omega^N$ and denote the bound in 2-norm by $\|\nabla f\|_2$. The maximal distance between two points in a box is bounded by the length of the longest diagonal, which is $\delta\sqrt{Nd}$. Hence the approximation error satisfies $|\epsilon| \leq \|\nabla f\|_2 \delta \sqrt{Nd}$, which is obtained by applying the mean value theorem and the Cauchy-Schwarz inequality. Note that the indicator function $\mathbb{1}_{B^{Z;\delta}}(X)$ is permutation invariant and can be rewritten as

$$\begin{aligned}
\mathbb{1}_{B^{Z;\delta}}(X) &= \frac{1}{C_Z} \sum_{\sigma \in S(N)} \mathbb{1}_{\{X | \mathbf{x}_i = \mathbf{z}_{\sigma(i)} + \delta \mathbf{u}_i, \mathbf{u}_i \in [0,1]^d, 1 \leq i \leq N\}}(X) \\
&= \frac{1}{C_Z} \sum_{\sigma \in S(N)} \prod_{i=1}^N \mathbb{1}_{\{\mathbf{x}_i = \mathbf{z}_{\sigma(i)} + \delta \mathbf{u}_i, \mathbf{u}_i \in [0,1]^d\}}(\mathbf{x}_i) \\
&= \frac{1}{C_Z} \sum_{\sigma \in S(N)} \prod_{i=1}^N \mathbb{1}_{\{\mathbf{x}_{\sigma(i)} = \mathbf{z}_i + \delta \mathbf{u}_i, \mathbf{u}_i \in [0,1]^d\}}(\mathbf{x}_{\sigma(i)}) = \frac{1}{C_Z} \mathrm{perm}([f_i^Z(\mathbf{x}_j)]),
\end{aligned}$$

where $f_i^Z(\mathbf{x}) = \mathbb{1}_{\{\mathbf{x} | \mathbf{x} = \mathbf{z}_i + \delta \mathbf{u}, \mathbf{u} \in [0,1]^d\}}(\mathbf{x})$. The constant C_Z takes care of repetition that can happen depending on Z. When all elements in Z are distinct, the box that X lives in only corresponds to one permutation, so $C_Z = 1$ in this case. If say $\mathbf{z}_1 = \mathbf{z}_2$ and all other elements are distinct, then the box that X lives in has two corresponding permutations that differ by a swapping of the first two elements. In this case, $C_Z = 2$ will account for the arising repetition. Next we apply the Ryser formula to the permanent,

$$\mathbb{1}_{B^{Z;\delta}}(X) = \frac{1}{C_Z} \mathrm{perm}([f_i^Z(\mathbf{x}_j)]) = \frac{(-1)^N}{C_Z} \sum_{S \subseteq \{1,\ldots,N\}} (-1)^{|S|} e^{\sum_{i=1}^N \log\left(\sum_{j \in S} f_j^Z(\mathbf{x}_i)\right)}.$$

We can now define $g:\mathbb{R}^d \to \mathbb{R}^{|S(N)|\times|\bigwedge^N \mathbb{L}|}$ where each component function is given by

$$g_S^Z(\mathbf{x}) = \log\left(\sum_{j\in S} f_j^Z(\mathbf{x})\right)$$

and we define $\phi:\mathbb{R}^{|S(N)|\times|\bigwedge^N \mathbb{L}|} \to \mathbb{R}$ as

$$\phi(Y) = \sum_{Z\in\bigwedge^N \mathbb{L}} \frac{(-1)^N f(Z)}{C_Z} \sum_{S\subseteq\{1,\ldots,N\}} (-1)^{|S|} e^{Y_S^Z}. \tag{3.1}$$

Since f_j^Z's are indicator functions, we naturally have $\sum_{j\in S} f_j^Z(\mathbf{x}) \geq 0$. In the case when $\sum_{j\in S} f_j^Z(\mathbf{x}) = 0$, $g_S^Z(\mathbf{x}) = -\infty$. In this case, $e^{g_S^Z(\mathbf{x})} = 0$, and therefore its contribution to ϕ vanishes as desired. In summary, we arrive at the universal approximation

$$f(X) = \phi\left(\sum_{j=1}^N g(\mathbf{x}_j)\right) + \epsilon. \tag{3.2}$$

Due to the explicit tabulation strategy, the number of terms needed in the approximation (3.2) can be counted as follows. The number of points in $\bigwedge^N \mathbb{L}$ is $\mathcal{O}((\delta^{-Nd})/N!)$, where $N!$ comes from the lexicographic ordering. Note that formally as $N\to\infty$, $(\delta^{-Nd})/N! \sim (\delta^{-d}/N)^N$ can vanish for fixed δ. However, this means that the number of elements has exceeded the number of grid points in $\mathbb{L}$ and is unreasonable. So we should at least have $\delta \lesssim N^{-\frac{1}{d}}$. In order to obtain an ϵ-close approximation of $f(X)$, we require our error bound $\|\nabla f\|_2 \delta\sqrt{Nd} \sim \epsilon$. When $\epsilon \leq \|\nabla f\|_2 \sqrt{Nd} N^{-\frac{1}{d}}$, we can choose $\delta = \frac{\epsilon}{\|\nabla f\|_2}(Nd)^{-\frac{1}{2}}$ so that both $\delta \lesssim N^{-\frac{1}{d}}$ and $\|\nabla f\|_2 \delta\sqrt{Nd} \sim \epsilon$ are fulfilled, and the number of points in $\bigwedge^N \mathbb{L}$ becomes $(\|\nabla f\|_2^2 Nd)^{Nd/2}/(\epsilon^{Nd} N!)$ because $\delta^{-Nd} = (\|\nabla f\|_2^2 Nd)^{Nd/2}/\epsilon^{Nd}$. For each Z, the number of terms to be summed over in Equation (3.1) is $|S(N)| = 2^N$. Therefore in order to obtain an ϵ-approximation, the number of feature variables is given by Equation (1.4). This proves Theorem 1.1.

This is of course a very pessimistic bound, and we will discuss on the practical implications for designing neural network architectures in Section 5. We remark that one may expect that following the same tabulation strategy, we may also provide a quantitative bound for ϕ constructed by the homeomorphism mapping E^{-1} as discussed in Section 2. However, the difference is that our bound only relies on the smoothness of the original function f and hence the bound for δ. On the other hand, the mapping E^{-1} and hence ϕ can be arbitrarily pathological, and therefore it is not even clear how to obtain a double-exponential type of bound as discussed above. We also remark that if the indicator functions $\mathbb{1}_{B^{Z;\delta}}(X)$ and $\mathbb{1}_{\{\mathbf{x}|\mathbf{x}=\mathbf{z}_i+\delta\mathbf{u},\mathbf{u}\in[0,1]^d\}}(\mathbf{x})$ in the proof are replaced by proper smooth cutoff functions with respect to corresponding domains, the ansatz in Equation (3.2) can be continuous to accommodate the applications that require continuity. For example, an interatomic potential energy should be continuous to guarantee the total energy is conserved during molecular dynamics simulations.

4. Totally anti-symmetric functions

Now we consider an anti-symmetric function. Similar to the symmetric case, when $d=1$ and $f(X)$ is a polynomial of X and anti-symmetric, it is known that

$$f(x_1,\ldots,x_N) = U(X)\prod_{i<j}(x_i - x_j),$$

where $U(X)$ is a symmetric polynomial and the second term is a Vandermonde determinant. This was first proved by Cauchy [4], who of course also first introduced the concept of determinant in its modern sense.

Our aim is to generalize to $d>1$. Without loss of generality we again assume $\mathbf{x}_i \in [0,1]^d := \Omega$. The construction of the ansatz is parallel to the totally symmetric case. Recall the lattice $\mathbb{L}$, the wedge $\bigwedge^N \mathbb{L}$, and the corresponding union of boxes $B^{Z,\delta}$. In the totally symmetric case, we make a piece-wise constant approximation over the union of boxes. For the anti-symmetric case, we would need to insert an anti-symmetric factor (w.r.t. X):

$$f(X) = \sum_{Z \in \bigwedge^N \mathbb{L}} f(Z) \frac{\psi^Z(X)}{\psi^Z(Z)} \mathbb{1}_{B^{Z;\delta}}(X) + \epsilon, \tag{4.1}$$

where $\psi^Z(\cdot)$ is a totally anti-symmetric function which might be chosen depending on Z. Here the approximation error, similar as before, satisfies $|\epsilon| < \|\nabla f\|_2 \delta \sqrt{Nd}$. Motivated by the one-dimension result, we focus on constructions of ψ^Z given by the Vandermonde determinant. Given a permutation equivariant map $X \mapsto Y^Z : (\mathbb{R}^d)^N \to \mathbb{R}^N$, we consider

$$\psi^Z(X) = \prod_{i<j} (y_i(X) - y_j(X)). \tag{4.2}$$

It is clear that due to the permutation equivariance of Y^Z, the ψ^Z defined is anti-symmetric. Observe that due to anti-symmetry $f(Z)=0$ whenever $\mathbf{z}_i = \mathbf{z}_j$ for some $i \neq j \in [n]$. In particular, this means that we only need to consider ψ^Z for those Z for which all $\mathbf{z}_i$'s are different. As $\mathbf{z}_i$'s are distinct, for $X \in B^{Z;\delta}$, there exists a unique $\sigma \in S(N)$ such that

$$X \in \left\{ X' \,|\, \mathbf{x}'_i = \mathbf{z}_{\sigma(i)} + \delta \mathbf{u}_i, \quad \mathbf{u}_i \in [0,1]^d \right\}. \tag{4.3}$$

Denote this unique permutation as $\sigma_{Z,X}$, and we take the permutation equivariant map Y^Z such that

$$y_i^Z(X) = \sigma_{Z,X}(i). \tag{4.4}$$

Since $\sigma_{Z,X}$ gives the sorting of X according to Z, it is easy to see that the above Y^Z is equivariant, and

$$\psi^Z(X) = \prod_{i<j} (y_i(X) - y_j(X)) = (-1)^{\sigma_{Z,X}} \prod_{i<j} (i-j). \tag{4.5}$$

Note that $\psi^Z(Z) = \prod_{i<j}(i-j)$, so we arrive at

$$f(X) = \sum_{Z \in \bigwedge^N \mathbb{L}} U^Z(X) \prod_{i<j} (y_i(X) - y_j(X)) + \epsilon, \tag{4.6}$$

with

$$U^Z(X) = \frac{f(Z)}{\prod_{i<j}(i-j)} \mathbb{1}_{B^{Z;\delta}}(X).$$

In this construction, $U^Z(X)$ is a scaled version of the characteristic function, and can thus be treated similarly as in the totally symmetric case based on Ryser's formula. As

the construction depends on the same tabular strategy as the totally symmetric case, the number of terms involved in the sum over $\bigwedge^N \mathbb{L}$ is the same too, which we will not repeat. The total number of feature variables is the same as that in Eq. (1.4) due to the use of the indicator function. This proves Theorem 1.2.

We remark that the index-based permutation equivariant map Y^Z in (4.4) leads to discontinuity at certain points. We may modify this permutation equivariant map in order to make the whole approximation continuous so that it is more suitable for scientific applications such as representing the many-body wavefunctions in VMC. However, such a continuous approximation in the current tabular strategy would imply that the anti-symmetric function $\psi^Z(\cdot)$ has the nodes close to those of f, which seems to be a very strong assumption. We leave the error bounds of a continuous approximation to an anti-symmetric function to future works.

5. Practical considerations and discussion

In this paper we study the universal approximation for symmetric and anti-symmetric functions. Following the line of learning theory, there are many questions open. For instance, the impact of symmetry on the generalization error remains unclear. This requires in-depth understanding of the suitable function class for symmetric and anti-symmetric functions, such as some adapted Barron space [9]. Note that a recent work [23] investigates the approximation and generalization bound of permutation invariant deep neural networks in the general case $d \geq 1$, however with two limitations. The first is a rather strong assumption that the target function is Lipschitz with respect to the ℓ_∞ norm (but not the usual Euclidean norm). This can be a severe limitation as the dimension (both N and d) increases. Indeed, under the same Lipschitz assumption of [23], the number of feature variables M in our Theorem 1.1 can be improved accordingly to $\mathcal{O}\left(2^N/(\epsilon^{Nd}N!)\right)$. The second limitation is that the proposed ansatz in [23] introduces *sorting layers* to represent the sorting procedure at the first step. The sorting procedure will bring discontinuity, which leads to serious problems in some scientific applications, such as interatomic potential energy in molecular dynamics simulations.

For anti-symmetric function, the ansatz suitable for the practical computation is of interest since one main motivation for studying anti-symmetric function is to integrate neural network-based wavefunction into VMC. The Vandermonde determinant considered here is proved to provide a simple but universal ansatz. Its universality suggests us to consider the following trial wavefunction in VMC:

$$f(X) = \sum_{k=1}^{K} U^k(X) \prod_{i<j} (y_i^k(X) - y_j^k(X)),$$

where $U^k(X)$ is symmetric and $(y_1^k(X), \ldots, y_N^k(X))$ is a permutation equivariant map. The ansatz for each U^k and Y^k can be still quite flexible. Another more general yet more complicated ansatz is based on replacing the above Vandermonde determinant with the Slater determinant (see e.g. [12, 19])

$$\det[Y^k(X)] = \begin{vmatrix} y_1^k(X) & \ldots & y_1^k(X) \\ \vdots & & \vdots \\ y_N^k(X) & \ldots & y_N^k(X) \end{vmatrix}.$$

The Slater determinant is widely used in quantum chemistry. It is known to be universal under the complete basis set and, indeed, the basis set derived from the Hartree-Fock

approximation often provides a fairly good starting point for most of the modern quantum chemistry methods. However, the complexity of computing a Slater determinant is $\mathcal{O}(N^3)$, while it is only $\mathcal{O}(N^2)$ for a Vandermonde determinant. This may become a more severe issue when one calculates the local energy, which involves the evaluation of the Laplacian of the trial wavefunction. Therefore, it remains interesting to spend more effort on the Vandermonde ansatz and its variants, hoping to find good ones that strike a good balance between accuracy and efficiency. It would also be interesting to learn from the second quantized representation of quantum systems, which lifts the symmetry requirement of functions to that of linear operators, and leads to powerful representations such as matrix product states.

Acknowledgements. We thank the hospitality of the American Institute of Mathematics (AIM) for the workshop "Deep learning and partial differential equation" in October 2019, which led to this collaborative effort. The work of LL and JZ was supported in part by the Department of Energy under Grant No. DE-SC0017867, and No. DE-AC02-05CH11231. The work of YL and JL was also supported in part by the National Science Foundation via grants DMS-1454939 and ACI-1450280.

REFERENCES

[1] M. Bachmayr, G. Dusson, and C. Ortner, *Polynomial approximation of symmetric functions*, arXiv preprint, arXiv:2109.14771, 2021. 1

[2] A.R. Barron, *Universal approximation bounds for superpositions of a sigmoidal function*, IEEE Trans. Inf. Theory, 39(3):930–945, 1993. 1

[3] J. Behler and M. Parrinello, *Generalized neural-network representation of high-dimensional potential-energy surfaces*, Phys. Rev. Lett., 98(14):146401, 2007. 1

[4] A.-L. Cauchy, *Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs gales et de signes contraires par suite des transpositions opres entre les variables qu'elles renferment*, J. Ec. Polytech., 1815. 1, 4

[5] K. Choo, A. Mezzacapo, and G. Carleo, *Fermionic neural-network states for ab-initio electronic structure*, Nat. Commun., 11(1):1–7, 2020. 1

[6] G. Cybenko, *Approximation by superpositions of a sigmoidal function*, Math. Control Signals Syst., 2(4):303–314, 1989. 1

[7] G. Dusson, M. Bachmayr, G. Csanyi, R. Drautz, S. Etter, C. van der Oord, and C. Ortner, *Atomic cluster expansion: Completeness, efficiency and stability*, J. Comput. Phys., 454:110946, 2022. 1

[8] W. E, J. Han, and L. Zhang, *Machine-learning-assisted modeling*, Phys. Today, 74(7):36–41, 2021. 1

[9] W. E, C. Ma, and L. Wu, *Barron spaces and the compositional function spaces for neural network models*, arXiv preprint, arXiv:1906.08039, 2019. 5

[10] M. Germain, M. Laurière, H. Pham, and X. Warin, *DeepSets and their derivative networks for solving symmetric PDEs*, J. Sci. Comput., 91:63, 2022. 1

[11] J. Han, L. Zhang, and W. E, *Solving many-electron Schrödinger equation using deep neural networks*, J. Comput. Phys., 399:108929, 2019. 1

[12] J. Hermann, Z. Schätzle, and F. Noé, *Deep-neural-network solution of the electronic Schrödinger equation*, Nat. Chem., 12(10):891–897, 2020. 1, 5

[13] K. Hornik, M. Stinchcombe, and H. White, *Multilayer feedforward networks are universal approximators*, Neural Netw., 2(5):359–366, 1989. 1

[14] M. Hutter, *On representing (anti)symmetric functions*, arXiv preprint, arXiv:2007.15298, 2020. 1

[15] S. Klus, P. Gelß, F. Nüske, and F. Noé, *Symmetric and antisymmetric kernels for machine learning problems in quantum physics and chemistry*, Mach. Learn.: Sci. Technol., 2(4):045016, 2021. 1

[16] A.N. Kolmogorov, *On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition*, Dokl. Akad. Nauk, 114:953–956, 1957. 2

[17] D. Luo and B.K. Clark, *Backflow transformations via neural networks for quantum many-body wave functions*, Phys. Rev. Lett., 122(22):226401, 2019. 1

[18] I.G. Macdonald, *Symmetric Functions and Hall Polynomials*, Oxford Univ. New York, 1979. 2

[19] D. Pfau, J.S. Spencer, A.G.D.G. Matthews, and W.M.C. Foulkes, *Ab initio solution of the many-electron Schrödinger equation with deep neural networks*, Phys. Rev. Res., 2(3):033429, 2020. 1, 5

[20] C.R. Qi, H. Su, K. Mo, and L.J. Guibas, *Pointnet: Deep learning on point sets for 3D classification and segmentation*, Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 77–85, 2017. 1

[21] C.R. Qi, L. Yi, H. Su, and L.J. Guibas, *Pointnet++: Deep hierarchical feature learning on point sets in a metric space*, Adv. Neur. Inf. Process. Syst., 5105–5114, 2017. 1

[22] H.J. Ryser, *Combinatorial Mathematics*, The Carus Mathematical Monographs, Mathematical Association of America, 14, 1963. 1, 2

[23] A. Sannai and M. Imaizumi, *Improved generalization bound of permutation invariant deep neural networks*, arXiv preprint, arXiv:1910.06552, 2019. 1, 5

[24] A. Sannai, Y. Takai, and M. Cordonnier, *Universal approximations of permutation invariant/equivariant functions by deep neural networks*, arXiv preprint, arXiv:1903.01939, 2019. 1

[25] K. Schütt, P.-J. Kindermans, H.E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, *SchNet: A continuous-filter convolutional neural network for modeling quantum interactions*, Adv. Neur. Inf. Process. Syst., 992–1002, 2017. 1

[26] J. Stokes, J.R. Moreno, E.A. Pnevmatikakis, and G. Carleo, *Phases of two-dimensional spinless lattice fermions with first-quantized deep neural-network quantum states*, Phys. Rev. B, 102:205122, 2020. 1

[27] M. Zaheer, S. Kottur, S. Ravanbakhsh, B. Poczos, R.R. Salakhutdinov, and A.J. Smola, *Deep sets*, Adv. Neur. Inf. Process. Syst., 3394–3404, 2017. 1, 2, 2, 2

[28] L. Zhang, J. Han, H. Wang, R. Car, and W. E, *Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics*, Phys. Rev. Lett., 120(14):143001, 2018. 1

[29] L. Zhang, J. Han, H. Wang, W. Saidi, R. Car, and W. E, *End-to-end symmetry preserving interatomic potential energy model for finite and extended systems*, Adv. Neur. Inf. Process. Syst., 4441–4451, 2018. 1

[30] X.-H. Zhou, J. Han, and H. Xiao, *Frame-independent vector-cloud neural network for nonlocal constitutive modeling on arbitrary grids*, Comput. Methods Appl. Mech. Eng., 388:114211, 2022. 1

[31] A. Zweig and J. Bruna, *A functional perspective on learning symmetric functions with neural networks*, Proc. Int. Conf. Mach. Learn., 139:13023–13032, 2021. 1