$O(N^2)$ Universal Antisymmetry in Fermionic Neural Networks

Tianyu Pang [1] **Shuicheng Yan** [1] **Min Lin** [1]

Abstract

Fermionic neural network (FermiNet) is a recently proposed wavefunction Ansatz, which is used in variational Monte Carlo (VMC) methods to solve the many-electron Schrödinger equation. FermiNet proposes permutation-equivariant architectures, on which a Slater determinant is applied to induce antisymmetry. FermiNet is proved to have *universal* approximation capability with a single determinant, namely, it suffices to represent any antisymmetric function given sufficient parameters. However, the asymptotic computational bottleneck comes from the Slater determinant, which scales with $O(N^3)$ for N electrons. In this paper, we substitute the Slater determinant with a pairwise antisymmetry construction, which is easy to implement and can reduce the computational cost to $O(N^2)$. We formally prove that the pairwise construction built upon permutation-equivariant architectures can *universally* represent any antisymmetric function. Besides, this universality can be achieved via *continuous* approximators when we aim to represent ground-state wavefunctions.

1. Introduction

How to efficiently solve fermionic quantum many-body systems has been a crucial problem in quantum physics and chemistry, as well as in a large number of applications in material science and drug discovery (Keimer and Moore, 2017; Heifetz, 2020). On the one hand, density functional theory (DFT) (Hohenberg and Kohn, 1964; Becke, 1992) is a widely used computational method in these fields due to its computational efficiency. There are also efforts trying to approximate DFT via deep neural networks (Gilmer et al., 2017; Schütt et al., 2019). However, due to the simplified assumptions on the wavefunction, i.e. limiting the wavefunction to be Slater determinant constructed out of orthogonal single particle wavefunctions, DFT relies on the exchange and correlation functional. Only approximate forms of the functionals (such as local-density approximations (LDA) and generalized gradient approximations (GGA)) are used in practice, leading to a potentially low prediction accuracy (Zhao et al., 1999; Iikura et al., 2001).

On the other hand, quantum Monte Carlo (QMC) is a family of Monte Carlo methods that provide more accurate approximations. Especially, variational Monte Carlo (VMC) involves a trial wavefunction to minimize the energy. The accuracy of VMC relies on how close the trial wavefunction is to the ground state. Optimizing the parameters of the wavefunction Ansatz only depends on the data sampled by Markov chain Monte Carlo (MCMC) methods (Hastings, 1970), and thus does not need supervised data.

Carleo and Troyer (2017) introduce neural networks to solve the quantum many-body problem by constructing neural quantum states as the Ansatz. Pfau et al. (2020) develop Fermionic Neural Network (FermiNet) as a wavefunction Ansatz to be used in VMC. The key insight from FermiNet is that the single-electron wavefunctions can be replaced with a permutation-equivariant neural network to capture the correlation of particles, while the Slater determinant is merely an antisymmetrizer. The model family of FermiNet is proved to be universal for representing any antisymmetric function even with a single Slater determinant. Compared to FermiNet, Hermann et al. (2020) design PauliNet which encodes built-in physical constraints to facilitate optimization. Along with this routine, graph neural networks (GNNs) and transformers are involved to model new Ansatz with equivariance (Gao and Günnemann, 2021; Atz et al., 2021; Thölke and Fabritiis, 2022).

A follow up work Spencer et al. (2020) implements FermiNet in JAX, which improves six fold compared to the original TensorFlow code. It shows the potential of applying recent advancements in both software and hardware to quantum chemistry problems. However, the antisymmetry of FermiNet is still implemented via the Slater determinant, which requires $O(N^3)$ computation for N-electron systems. The cost of computing the Slater determinant (or Pfaffian) will asymptotically overwhelm the other parts of the model. It limits the scalability of FermiNet and other variants that uses the Slater determinant.

To this end, several efforts have been devoted to determinant-

[1]Sea AI Lab, Singapore. Correspondence to: Tianyu Pang <tianyupang@sea.com>, Min Lin <linmin@sea.com>.

2nd AI4Science Workshop at the 39th International Conference on Machine Learning (ICML), 2022. Copyright 2022 by the author(s).

free strategies for separately modelling the sign and amplitude of wave functions (Torlai et al., 2018; Choo et al., 2020; Szabó and Castelnovo, 2020; Stokes et al., 2020; Inui et al., 2021). Nevertheless, they are applied to spin lattice systems or molecules in second quantization, as discussed in Schätzle et al. (2021). In real-space cases, Han et al. (2019) design a pairwise $O(N^2)$ Ansatz, but we demonstrate that their Ansatz requires discontinuous approximators to represent ground-state wavefunctions (detailed in Section 3.1).

In this paper, we develop $O(N^2)$ pairwise Ansatz based on the permutation-equivariant multi-electron functions used in FermiNet. Our pairwise Ansatz is straightforward to implement, with minor modification on the FermiNet architecture, and can be applied to other network-based architectures like PauliNet (Hermann et al., 2020). We formally prove that our pairwise Ansatz can *universal* represent any ground-state wavefunction with *continuous* approximators, which can be modelled via finite-capacity neural networks.

2. Preliminary

We let $x = (x_1, \cdots, x_N) \in \mathbb{R}^{dN}$ denote the d-dimensional coordinates of an N-electron system, where $x_i \in \mathbb{R}^d$ for $i = 1, \cdots, N$. Given a set of single-electron orbitals $\{\phi_1, \cdots, \phi_N\}$, we can construct an antisymmetric wavefunction Ansatz via the Slater determinant as

$$\psi_{\text{single}}(x) = \begin{vmatrix} \phi_1(x_1) & \cdots & \phi_1(x_N) \\ \vdots & & \vdots \\ \phi_N(x_1) & \cdots & \phi_N(x_N) \end{vmatrix} = \det[\phi_i(x_j)]. \tag{1}$$

It is known that only a small subset of all possible antisymmetric functions can be written as the form of $\psi_{\text{single}}(x)$, i.e., the model family of $\psi_{\text{single}}(x)$ is *not* universal.

To this end, FermiNet (Pfau et al., 2020) replaces the single-electron orbitals $\phi_i(x_j)$ with the multi-electron functions $\phi_i(x_j; \{x_{\setminus j}\})$ modeled by neural networks, where $\{x_{\setminus j}\}$ denotes the set of all electrons except x_j. By architecture design, the functions $\phi_i(x_j; \{x_{\setminus j}\})$ are invariant to the permutation of elements in $\{x_{\setminus j}\}$. The wavefunction Ansatz of FermiNet can be written as

$$\psi_{\text{Fermi}}(x) = \det\left[\phi_i(x_j; \{x_{\setminus j}\})\right]. \tag{2}$$

The model family of $\psi_{\text{Fermi}}(x)$ is proved to be universal to represent any antisymmetric function, as stated below:

Lemma 1. *(Universality by Pfau et al. (2020)) For any antisymmetric function $\Psi(x)$, there exist multi-electron functions $\phi_1, \cdots \phi_N$, such that $\forall x$, there is $\psi_{Fermi}(x) = \Psi(x)$.*

In practice, FermiNet uses multiple $\psi^k_{\text{Fermi}}(x)$ to achieve high accuracy for $k = 1, \cdots, K$, and the ensemble wavefunction $\Psi_{\text{Fermi}}(x)$ is constructed as $\Psi_{\text{Fermi}}(x) = \sum_k \omega_k \psi^k_{\text{Fermi}}(x)$ where ω_k is the weight of k-th component.

3. $O(N^2)$ Universal Antisymmetry

Evaluating the Slater determinant operator in Eq. (2) requires $O(N^3)$ computation, which asymptotically limits the scalability of FermiNet and most of the other VMC methods. Given this, we develop an easy-to-implement $O(N^2)$ substitute for the Slater determinant to build antisymmetric functions on top of a permutation-equivariant function. Technically, we require a function $F(x_i, x_j; \{x_{\setminus\{i,j\}}\})$ that is invariant to any permutation on the set $\{x_{\setminus\{i,j\}}\}$. Then we construct the **pairwise wavefunction Ansatz** as

$$\psi_{\text{pair}}(x) = \prod_{1 \le i < j \le N} \mathcal{A} \circ F(x_i, x_j; \{x_{\setminus\{i,j\}}\}), \tag{3}$$

where $\mathcal{A}$ is an antisymmetrizer such that $\mathcal{A} \circ F(x_i, x_j) = -\mathcal{A} \circ F(x_j, x_i)$ holds.[1] Now we demonstrate that $\psi_{\text{pair}}(x)$ is also antisymmetric:

Lemma 2. *(Antisymmetry) $\psi_{pair}(x)$ is antisymmetric under the permutation of any two elements x_m and x_n in x.*

Proof. Assuming that $m < n$ without loss of generality, after we permute x_m and x_n, the term $\mathcal{A} \circ F(x_m, x_i)$ changes to $\mathcal{A} \circ F(x_i, x_m)$ and the term $\mathcal{A} \circ F(x_i, x_n)$ changes to $\mathcal{A} \circ F(x_n, x_i)$ for $i \in (m, n)$. Besides, the term $\mathcal{A} \circ F(x_m, x_n)$ changes to $\mathcal{A} \circ F(x_n, x_m)$. Note that each term change contributes to a sign change (i.e., multiplying -1). So the total number of sign changes is $1 + 2 \times (n - m - 1)$, which is always an odd number and thus $\psi_{\text{pair}}(x)$ is antisymmetric. □

A simple implementation of $\mathcal{A} \circ F$ is $\mathcal{A} \circ F(x_i, x_j) = F(x_i, x_j) - F(x_j, x_i)$, which requires $O(1)$ extra computation on top of computing F. Since there are $C_N^2 = N \times (N-1)/2$ terms in $\psi_{\text{pair}}(x)$, totally it requires $O(N^2)$ extra computation to induce antisymmetry. In the following, we analyse how different choices of $\mathcal{A} \circ F$ influence the *universality* and *continuity* of $\psi_{\text{pair}}(x)$.

3.1. Instantiation from Han et al. (2019)

Han et al. (2019) developed an $O(N^2)$ scaling Ansatz $\psi_{\text{Han}}(x)$ formulated as

$$\psi_{\text{Han}}(x) = \phi_C(x) \cdot \prod_{1 \le i < j \le N} \left(\phi_B(x_j, x_i) - \phi_B(x_i, x_j)\right). \tag{4}$$

Here $\phi_B(x_j, x_i)$ is a two-electron function, and $\phi_C(x) = \phi_C(x_1, \cdots, x_N)$ is a multi-electron symmetric function. The form of $\psi_{\text{Han}}(x)$ can be regarded as a special instantiation of $\psi_{\text{pair}}(x)$, where

$$\mathcal{A} \circ F(x_i, x_j) = (\phi_B(x_j, x_i) - \phi_B(x_i, x_j)) \cdot \phi_C(x)^{\frac{2}{N(N-1)}}.$$

[1] For notation compactness, here we omit the dependence of $F(x_i, x_j)$ on other elements in $\{x_{\setminus\{i,j\}}\}$ without ambiguity.

We can show that for ground-state wavefunctions $\Psi^g(x)$ of fermionic systems,[2] we *cannot* find continuous constructions of both ϕ_B and ϕ_C to satisfy $\psi_{\text{Han}}(x) = \Psi^g(x)$, which makes it non-trivial for neural networks to approximate ϕ_B and ϕ_C. To prove this, we resort to the tiling property of ground-state antisymmetric wavefunctions, as stated below:

Lemma 3. *(Tiling property by Ceperley (1991)) We define $\mathcal{N}(\Psi^g) = \{x|\Psi^g(x) = 0\}$ to be the node set of ground-state Ψ^g, and $\Omega(\Psi^g, x)$ be the nodal cell around the point x. Then the tiling property claims that $\forall x \notin \mathcal{N}(\Psi^g)$,*

$$\bigcup_{\pi} \Omega(\Psi^g, x_\pi) = \mathbb{R}^{dN} \setminus \mathcal{N}(\Psi^g), \tag{5}$$

where x_π is the input x permuted by π and the union $\bigcup_\pi$ is taken over all possible permutations.

According to the tiling property of ground-state wavefunctions, we can further prove the following:

Theorem 1. *(Discontinuity) There exist ground-state wavefunctions $\Psi^g(x)$, such that if there is $\forall x$, $\psi_{Han}(x) = \Psi^g(x)$, then either ϕ_B or ϕ_C must be discontinuous on $\mathbb{R}^{dN}$.*

Proof. For notation compactness, we denote

$$A(x; \phi_B) = \prod_{i<j} \left(\phi_B(x_j, x_i) - \phi_B(x_i, x_j)\right), \tag{6}$$

and thus $\psi_{\text{Han}}(x) = A(x; \phi_B) \cdot \phi_C(x)$. We first prove that there exist ground-state $\Psi^g(x)$, such that for any ϕ_B, the sign of $A(x; \phi_B)$ *cannot* keep aligned with the sign of $\Psi^g(x)$. Specifically, we consider a three-electron system, where we have four coordinate candidates x_1, x_2, x_3, x_4. For simplicity, we write $A_{i,j,k} = A(x_i, x_j, x_k; \phi_B)$ and $a_{i,j} = \phi_B(x_j, x_i) - \phi_B(x_i, x_j)$, where $A_{i,j,k} = a_{i,j} \cdot a_{i,k} \cdot a_{j,k}$. Now we can derive that

$$\begin{aligned} &A_{1,2,3} \cdot A_{1,2,4} \cdot A_{1,3,4} \cdot A_{2,3,4} \\ =&a_{1,2}^2 \cdot a_{1,3}^2 \cdot a_{1,4}^2 \cdot a_{2,3}^2 \cdot a_{2,4}^2 \cdot a_{3,4}^2 \geq 0, \end{aligned}$$

which holds for any x_1, x_2, x_3, x_4. Thus, given any $\Psi^g(x)$ satisfying that $\text{sgn}(\Psi^g_{1,2,3} \cdot \Psi^g_{1,2,4} \cdot \Psi^g_{1,3,4} \cdot \Psi^g_{2,3,4}) = -1$,[3] there must be $\text{sgn}(A(x; \phi_B)) \neq \text{sgn}(\Psi^g(x))$ for any ϕ_B. Since $\text{sgn}((-\Psi^g_{1,2,3}) \cdot (-\Psi^g_{1,2,4}) \cdot (-\Psi^g_{1,3,4}) \cdot (-\Psi^g_{2,3,4})) = -1$, there is also $\text{sgn}(A(x; \phi_B)) \neq \text{sgn}(-\Psi^g(x))$.

Let Ψ^g be any one of the ground-state functions whose sign cannot be represented via $A(x; \phi_B)$, i.e., there exist $x' \notin \mathcal{N}(\Psi^g)$ such that $\text{sgn}(A(x'; \phi_B)) \neq \text{sgn}(\Psi^g(x'))$. Now we prove that if ϕ_B is continuous, there must be

$$\mathcal{N}(A(x; \phi_B)) \not\subset \mathcal{N}(\Psi^g). \tag{7}$$

[2] We assume that the ground-state $\Psi^g(x)$ is continuous on $\mathbb{R}^{dN}$.

[3] Intuitively, we can select x_1 and x_2 to be very close, and select x_3 and x_4 to satisfy $\text{sgn}(\Psi^g_{1,2,3} \cdot \Psi^g_{1,2,4}) = -1$. Then due to the continuity of Ψ^g, there is $\text{sgn}(\Psi^g_{1,3,4} \cdot \Psi^g_{2,3,4}) = 1$.

Specifically, we know that $A(x; \phi_B)$ is continuous and $\text{sgn}(A(x'; \phi_B)) = \text{sgn}(-\Psi^g(x'))$. Assume, to the contrary, that $\mathcal{N}(A(x; \phi_B)) \subset \mathcal{N}(\Psi^g)$. Then according to the tiling property, we have the nodal cell $\Omega(\Psi^g, x') \subset \mathbb{R}^{dN} \setminus \mathcal{N}(A(x; \phi_B))$, and due to the continuity of $A(x; \phi_B)$, we know that the signs of both $A(x; \phi_B)$ and Ψ^g are unchanged in $\Omega(\Psi^g, x')$. Thus, $\forall x \in \Omega(\Psi^g, x')$, $\text{sgn}(A(x; \phi_B)) = \text{sgn}(-\Psi^g(x))$. Since both $A(x; \phi_B)$ and $\Psi^g(x)$ are antisymmetric, given any permutation π, there is $\forall x \in \Omega(\Psi^g, x'_\pi)$, $\text{sgn}(A(x; \phi_B)) = \text{sgn}(-\Psi^g(x))$. Taking in the tiling property of Eq. (5), we achieve that $\forall x \notin \mathcal{N}(\Psi^g)$, there is $\text{sgn}(A(x; \phi_B)) = \text{sgn}(-\Psi^g(x))$, which lead to contradiction with our three-electron counter-example above.

Now we have proved that if ϕ_B is continuous, there must be $\mathcal{N}(A(x; \phi_B)) \not\subset \mathcal{N}(\Psi^g)$, i.e., there exist an input x'' such that $A(x''; \phi_B) = 0$ and $\Psi^g(x'') \neq 0$. Assume that $\psi_{\text{Han}}(x) = A(x; \phi_B) \cdot \phi_C(x) = \Psi^g(x)$ for $\forall x$, then for any finite value of $\phi_C(x'')$, we always have $\psi_{\text{Han}}(x'') = A(x''; \phi_B) \cdot \phi_C(x'') \neq \Psi^g(x')$, which leads to a contradiction. Thus, either ϕ_B or ϕ_C is discontinuous on $\mathbb{R}^{dN}$. □

3.2. Achieving Universality with Continuous ϕ_B

In the above section, we show that there exist ground-state functions $\Psi^g(x)$ that $\psi_{\text{Han}}(x)$ have to represent via discontinuous ϕ_B or ϕ_C, which makes finite-capacity neural networks difficult to approximate them, as also discussed in Pfau et al. (2020). Intuitively, this discontinuity comes from the two-electron dependence of ϕ_B, i.e., ϕ_B cannot model the relation among multiple electrons.

To this end, we resort to the permutation-equivariant architecture designed in FermiNet, which includes N multi-electron functions $\phi_1(x_j; \{x_{\setminus j}\}), \cdots, \phi_N(x_j; \{x_{\setminus j}\})$ that are invariant to any permutation in the set $\{x_{\setminus j}\}$, as described in Section 2. In contrast, here we only need one function $\phi_B(x_j; \{x_{\setminus j}\})$, and we instantiate $\psi'_{\text{pair}}(x)$ as

$$\psi'_{\text{pair}}(x) = \prod_{1 \leq i < j \leq N} \left(\phi_B(x_j; \{x_{\setminus j}\}) - \phi_B(x_i; \{x_{\setminus i}\})\right). \tag{8}$$

Now we prove that the model family of $\psi'_{\text{pair}}(x)$ is *universal* to represent any ground-state wavefunction Ψ^g using *continuous* ϕ_B, as stated below:

Theorem 2. *(Continuous universality) For any ground-state wavefunction $\Psi^g(x)$, there exist a continuous multi-electron function ϕ_B, such that $\forall x$, $\psi'_{pair}(x) = \Psi^g(x)$.*

Proof. For any ground-state wavefunction $\Psi^g(x)$, according to its tiling property, we can select one of the nodal cell Ω^*, such that $\forall x \in \Omega^*$, there is $\Psi^g(x) > 0$. Besides, there exist a permutation rule π^* to map the points from other nodal cells into the selected Ω^*, i.e., $\forall x \in \mathbb{R}^{dN} \setminus \mathcal{N}(\Psi^g)$, there is $x_{\pi^*} \in \Omega^*$ and $\Psi^g(x_{\pi^*}) > 0$, where x_{π^*} is the input

x permuted by π^*. We use $\sigma(\pi^*)$ denote the sorting sign of π^* and $\pi^*(x,j)$ denote the index of x_j after permutation. Then we let ϕ_B to satisfy that $\forall x \in \mathbb{R}^{dN} \setminus \mathcal{N}(\Psi^g)$, there is

$$\phi_B(x_j;\{x_{\setminus j}\}) = \pi^*(x,j)\cdot\left(\frac{\Psi^g(x_{\pi^*})}{\prod_{i<j}(j-i)}\right)^{\frac{2}{N(N-1)}}; \quad (9)$$

and $\forall x \in \mathcal{N}(\Psi^g)$, there is $\phi_B(x_j;\{x_{\setminus j}\}) = 0$, where $j = 1,\cdots,N$. Note that the tiling property also claims that if two points x and x' come from the same nodal cell (i.e., $\Omega(\Psi^g,x) = \Omega(\Psi^g,x')$), then π^* executes the same permutation operation on them (i.e., $\pi^*(x,j) = \pi^*(x',j)$ for any index j). Thus, the function $\pi^*(x,j)$ is constant inside any nodal cell, and due to the continuity of Ψ^g, we know that ϕ_B is continuous on $\mathbb{R}^{dN} \setminus \mathcal{N}(\Psi^g)$. Furthermore, when the input x approaches to the node set $\mathcal{N}(\Psi^g)$, the value of $\Psi^g(x_{\pi^*})$ will consistently approach to zero. So when $x \to \mathcal{N}(\Psi^g)$, there is $\phi_B(x_j;\{x_{\setminus j}\}) \to 0$ for any index j, by which we can conclude that ϕ_B is continuous on the entire space of $\mathbb{R}^{dN}$. Using the construction of continuous ϕ_B in Eq. (9), we have

$$\begin{aligned}\psi'_{\text{pair}}(x) =& \prod_{i<j}\left(\phi_B(x_j;\{x_{\setminus j}\}) - \phi_B(x_i;\{x_{\setminus i}\})\right)\\ =&\sigma(\pi^*)\cdot\prod_{i<j}(j-i)\cdot\left(\frac{\Psi(x_{\pi^*})}{\prod_{i<j}(j-i)}\right)^{\frac{2}{N(N-1)}}\\ =&\sigma(\pi^*)\cdot\Psi(x_{\pi^*}) = \sigma(\pi^*)\cdot\sigma(\pi^*)\cdot\Psi(x) = \Psi(x),\end{aligned}$$

which holds for $\forall x$. □

As to the antisymmetric function $\Psi(x)$ that is not ground-state, there still exist ϕ_B (but may not be continuous) making $\psi'_{\text{pair}}(x)$ represent Ψ, as stated below:

Corollary 1. *(Universality) For any antisymmetric function $\Psi(x)$, there exist a multi-electron function ϕ_B, such that $\forall x$, $\psi'_{pair}(x) = \Psi(x)$.*

Proof. Similar to the proof routine of Appendix B in Pfau et al. (2020), given a certain permutation π, we use $\sigma(\pi)$ to denote the sorting sign of π. For notation compactness, we define Λ as

$$\Lambda = \prod_{3\le j\le N} j\left(j + \text{sgn}(\Psi(x_\pi))\right)\prod_{3\le i<j\le N}(j-i),$$

where $\Psi(x)$ is the antisymmetric function that we aim to represent. It is easy to verify that $\Lambda > 0$. Then we could let ϕ_B satisfy that

$$\begin{aligned}&\phi_B(x_{\pi_1};\{x_{\pi_{\setminus 1}}\}) = -\text{sgn}(\Psi(x_\pi))\cdot\left(\frac{|\Psi(x_\pi)|}{\Lambda}\right)^{\frac{2}{N(N-1)}};\\ &\phi_B(x_{\pi_2};\{x_{\pi_{\setminus 2}}\}) = 0;\\ &\phi_B(x_{\pi_j};\{x_{\pi_{\setminus j}}\}) = j\cdot\left(\frac{|\Psi(x_\pi)|}{\Lambda}\right)^{\frac{2}{N(N-1)}}, \forall 3\le j\le N,\end{aligned}$$

where x_{π_j} is the j-th element of x_π. It is easy to verify that

$$\begin{aligned}\psi'_{\text{pair}}(x) =& \prod_{i<j}\left(\phi_B(x_j;\{x_{\setminus j}\}) - \phi_B(x_i;\{x_{\setminus i}\})\right)\\ =&\sigma(\pi)\cdot\Lambda\cdot\text{sgn}(\Psi(x_\pi))\cdot\frac{|\Psi(x_\pi)|}{\Lambda}\\ =&\sigma(\pi)\cdot\Psi(x_\pi) = \sigma(\pi)\cdot\sigma(\pi)\cdot\Psi(x) = \Psi(x),\end{aligned}$$

which holds for $\forall x$. □

Connection with $\psi_{\text{Fermi}}(x)$. The form of $\psi'_{\text{pair}}(x)$ can be written as the determinant of a Vandermonde matrix

$$\psi'_{\text{pair}}(x) = \begin{vmatrix} 1 & \cdots & 1\\ \phi_B(x_1;\{x_{\setminus 1}\}) & \cdots & \phi_B(x_N;\{x_{\setminus N}\})\\ \phi_B^2(x_1;\{x_{\setminus 1}\}) & \cdots & \phi_B^2(x_N;\{x_{\setminus N}\})\\ \vdots & \vdots & \vdots\\ \phi_B^{N-1}(x_1;\{x_{\setminus 1}\}) & \cdots & \phi_B^{N-1}(x_N;\{x_{\setminus N}\})\end{vmatrix},$$

which can be regarded as a special case of $\psi_{\text{Fermi}}(x)$ where $\phi_i = \phi_B^{i-1}$ (but $\psi'_{\text{pair}}(x)$ only requires $O(N^2)$ computation). Therefore, the model family of $\psi_{\text{Fermi}}(x)$ is strictly more powerful than $\psi'_{\text{pair}}(x)$, and consequently the continuous universality of representing ground-state wavefunctions (claimed in Theorem 2) also holds for $\psi_{\text{Fermi}}(x)$.

Extension of $\psi'_{\text{pair}}(x)$. We can extend $\psi'_{\text{pair}}(x)$ to involve more permutation-equivariant functions, e.g., we construct

$$\psi''_{\text{pair}}(x) = \prod_{1\le i<j\le N}\begin{vmatrix}\phi_A(x_i;\{x_{\setminus i}\}) & \phi_A(x_j;\{x_{\setminus j}\})\\ \phi_B(x_i;\{x_{\setminus i}\}) & \phi_B(x_j;\{x_{\setminus j}\})\end{vmatrix}. \quad (10)$$

Specially, if we trivially choose $\forall x$, $\phi_A(x_j;\{x_{\setminus j}\}) = 1$ be a constant function, then $\psi''_{\text{pair}}(x)$ degenerates to $\psi'_{\text{pair}}(x)$. Therefore, the model family of $\psi''_{\text{pair}}(x)$ is also strictly more powerful than $\psi'_{\text{pair}}(x)$, and the similar conclusions in Theorem 2 and Corollary 1 also hold for $\psi''_{\text{pair}}(x)$.

Computational complexity of implementing $\psi_{\text{pair}}(x)$. As analyzed in the Section IV B of Pfau et al. (2020), the complexity of implementing the multi-electron orbitals $\phi_1,\cdots,\phi_N$ is $O(N^2)$ for single-atom systems and fixed numbers of hidden units. After substituting the Slater determinant in FermiNet with our pairwise constructions (as described in Eq. (8) and Eq. (10)), we know that the cost of computing ϕ_A and ϕ_B is no more than $O(N^2)$. Besides, there are $C_N^2 = N\times(N-1)/2$ terms and each term $\mathcal{A}\circ F(x_i,x_j)$ requires $O(1)$ extra computation on top of ϕ_B (and ϕ_A). Therefore, totally $\psi'_{\text{pair}}(x)$ and $\psi''_{\text{pair}}(x)$ still require $O(N^2)$ computation.

3.3. Empirical Evaluation

In the experiments, we follow the main training settings in Pfau et al. (2020), as briefly recapped below. The FermiNet architecture includes four layers, each having 256

Table 1. Ground state energy. The values of 'Exact' column come from Chakravorty et al. (1993). The values of 'ψ'_{pair}' and 'ψ''_{pair}' are averaged on the last 1,000 iterations with sampling stride of 10.

Atom	ψ''_{pair}	ψ'_{pair}	Exact
Li	-7.4782	-7.4781	-7.47806032
Be	-14.6673	-14.6664	-14.66736
B	-24.5602	-24.4475	-24.65391
C	-37.3531	-37.2785	-37.8450
N	-53.1855	-53.0626	-54.5892

hidden units for the one-electron stream and 32 hidden units for the two-electron stream. The non-linear activation function is Tanh. To empirically estimate the gradient, 2,048 samples drawn from $\Psi_{\text{Fermi}}(x)$ are used in each training batch, with ensemble number $K = 16$. These samples are updated via the random-walk MCMC algorithm, with 10 iterations and a move width of 0.02. The optimizer is a modified version of KFAC (Martens and Grosse, 2015), with the initial learning rate of 10^{-4}.

In Table 1, we show some initial results using our pairwise constructions, where we substitute $\Psi_{\text{Fermi}}(x)$ with $\Psi'_{\text{pair}}(x)$ and $\Psi''_{\text{pair}}(x)$, respectively. Here we do not apply Hartree-Fock initialization. Our computational cost is less than FermiNet from two aspects: first, we use only one or two permutation-equivariant functions ϕ_A and ϕ_B for different atoms, while the number of functions $\phi_1, \cdots, \phi_N$ used in FermiNet scales with N; second, we do not involve determinant operators and only require $O(N^2)$ computation. However, as seen in the results, our methods empirically ask for permutation-equivariant functions with higher model capacity for larger atoms. This can be achieved via applying more powerful network architectures, and we leave this to future exploration.

4. Conclusion

This paper develops a simple way toward $O(N^2)$ universal antisymmetry in FermiNet by pairwise constructions. Our methods can also seamlessly adapt to other determinant-based Ansatz like PauliNet (Hermann et al., 2020). The universality of our pairwise Ansatz is established on the employment of multi-electron neural networks, which are proved to be universal approximators (Hornik et al., 1989). As to ground-state wavefunctions, we prove the existence of continuous construction, which facilitates practical learning via finite-capacity networks.

Although theoretically, we demonstrate the (continuous) universality of pairwise constructions, empirically, we may require a large model capacity to fit the ground-truth ϕ_A, ϕ_B. This limitation is reflected in our initial experiments, as seen in Table 1. Nevertheless, the model architecture used in FermiNet is still quite shallow from the aspect of deep learning, and many model designs could be applied to obtain much more powerful multi-electron functions. Another limitation is that there is no clear advantage compared to the Slater determinant when N is small, because the neural network part dominates the computation in this case.

References

Kenneth Atz, Francesca Grisoni, and Gisbert Schneider. Geometric deep learning on molecular representations. *Nature Machine Intelligence*, pages 1–10, 2021.

Axel D Becke. Density-functional thermochemistry. i. the effect of the exchange-only gradient correction. *The Journal of chemical physics*, 96(3):2155–2160, 1992.

Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, 2017.

David M Ceperley. Fermion nodes. *Journal of statistical physics*, 63(5):1237–1267, 1991.

Subhas J Chakravorty, Steven R Gwaltney, Ernest R Davidson, Farid A Parpia, and Charlotte Froese p Fischer. Ground-state correlation energies for atomic ions with 3 to 18 electrons. *Physical Review A*, 47(5):3649, 1993.

Kenny Choo, Antonio Mezzacapo, and Giuseppe Carleo. Fermionic neural-network states for ab-initio electronic structure. *Nature communications*, 11(1):1–7, 2020.

Nicholas Gao and Stephan Günnemann. Ab-initio potential energy surfaces by pairing gnns with neural wave functions. *arXiv preprint arXiv:2110.05064*, 2021.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International Conference on Machine Learning (ICML)*, 2017.

Jiequn Han, Linfeng Zhang, and E Weinan. Solving many-electron schrödinger equation using deep neural networks. *Journal of Computational Physics*, 399:108929, 2019.

W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

Alexander Heifetz. *Quantum mechanics in drug discovery*. Springer, 2020.

Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic schrödinger equation. *Nature Chemistry*, 12(10):891–897, 2020.

Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

Hisayoshi Iikura, Takao Tsuneda, Takeshi Yanai, and Kimihiko Hirao. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *The Journal of Chemical Physics*, 115(8):3540–3544, 2001.

Koji Inui, Yasuyuki Kato, and Yukitoshi Motome. Determinant-free fermionic wave function using feed-forward neural networks. *Physical Review Research*, 3 (4):043126, 2021.

B Keimer and JE Moore. The physics of quantum materials. *Nature Physics*, 13(11):1045–1055, 2017.

James Martens and Roger Grosse. Optimizing neural networks with kronecker-factored approximate curvature. In *International Conference on Machine Learning (ICML)*. PMLR, 2015.

David Pfau, James S Spencer, Alexander GDG Matthews, and W Matthew C Foulkes. Ab initio solution of the many-electron schrödinger equation with deep neural networks. *Physical Review Research*, 2(3):033429, 2020.

Zeno Schätzle, Jan Hermann, and Frank Noé. Convergence to the fixed-node limit in deep variational monte carlo. *The Journal of Chemical Physics*, 154(12):124108, 2021.

Kristof T Schütt, Michael Gastegger, Alexandre Tkatchenko, K-R Müller, and Reinhard J Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature communications*, 10(1):1–10, 2019.

James S. Spencer, David Pfau, Aleksandar Botev, and W. M.C. Foulkes. Better, faster fermionic neural networks, 2020. URL https://arxiv.org/abs/2011.07125.

James Stokes, Javier Robledo Moreno, Eftychios A Pnevmatikakis, and Giuseppe Carleo. Phases of two-dimensional spinless lattice fermions with first-quantized deep neural-network quantum states. *Physical Review B*, 102(20):205122, 2020.

Attila Szabó and Claudio Castelnovo. Neural network wave functions and the sign problem. *Physical Review Research*, 2(3):033075, 2020.

Philipp Thölke and Gianni De Fabritiis. Equivariant transformers for neural network based molecular potentials. In *International Conference on Learning Representations (ICLR)*, 2022.

Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. *Nature Physics*, 14(5):447–450, 2018.

GL Zhao, D Bagayoko, and TD Williams. Local-density-approximation prediction of electronic properties of gan, si, c, and ruo 2. *Physical Review B*, 60(3):1563, 1999.