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

Anonymous Authors[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 $O(N^2)$. Furthermore, we formally prove that the pairwise construction built upon permutation-equivariant architectures can *universally* represent any antisymmetric function.

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 heavily 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.

Pfau et al. (2020) propose 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. 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 (Bradbury et al., 2020), 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, Inui et al. (2021) develop a determinant-free strategy to separately model the sign and magnitude of wave functions, which reduces the computation to $O(N^2)$. Nevertheless, they are applied to discrete lattice systems, and it is unspecified how it could be applied to VMC methods. It also explicitly defines the ordering of input that may constrain the model flexibility.

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

In this paper, we develop pairwise constructions to achieve *universal* antisymmetry in FermiNet with $O(N^2)$ computation. Our pairwise constructions are straightforward to implement, with minor modification on the FermiNet architecture, and can be applied to other network-based Ansatz like PauliNet (Hermann et al., 2020). We provide several instantiations of our pairwise constructions and demonstrate that our methods can also separately model the sign and magnitude of any antisymmetric function without explicitly defining input ordering.

2. Preliminary

We let $x = (x_1, \cdots, x_N) \in \mathbb{R}^{3N}$ represent the coordinates of an N-electron system. 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. *(Proof in 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 $\psi_{Fermi}(x) = \Psi(x)$.*

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, Inui et al. (2021) propose a determinant-free strategy, which separately models the sign and magnitude of wavefunctions. However, their strategy is based on explicit choice of ordering and designed for discrete lattice systems, which is not off-the-shelf for different VMC methods and could break the natural symmetry of identical particles (Luo and Clark, 2019; Pfau et al., 2020).

3.1. Achieving $O(N^2)$ via Pairwise Construction

In this section, 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, which can similarly induce universal antisymmetry. 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:

Theorem 1. *(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. □

To implement the function $F(x_i, x_j)$, we resort to the permutation-equivariant architecture designed in FermiNet (Pfau et al., 2020), 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 two functions $\phi_A(x_j; \{x_{\setminus j}\})$ and $\phi_B(x_j; \{x_{\setminus j}\})$. Given ϕ_A and ϕ_B, we can implement

$$F(x_i, x_j) = \phi_A(x_i; \{x_{\setminus i}\})\phi_B(x_j; \{x_{\setminus j}\}); \tag{4}$$

$$\mathcal{A} \circ F(x_i, x_j) = \begin{vmatrix} \phi_A(x_i; \{x_{\setminus i}\}) & \phi_A(x_j; \{x_{\setminus j}\}) \\ \phi_B(x_i; \{x_{\setminus i}\}) & \phi_B(x_j; \{x_{\setminus j}\}) \end{vmatrix}. \tag{5}$$

Under the instantiation of $\mathcal{A} \circ F$ in Eq. (5), we can prove that the model family of $\psi_{\text{pair}}(x)$ is universal to represent any antisymmetric function, as stated below:

Theorem 2. *(Universality) For any antisymmetric function $\Psi(x)$, there exist multi-electron functions ϕ_A and ϕ_B, such that $\forall x$, there is $\psi_{pair}(x) = \Psi(x)$.*

Proof. Following the similar proof routine of Appendix B in Pfau et al. (2020), we define a unique ordering on the

[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.

vectors $x_1, \cdots, x_N$, and let π be their sorting operation, i.e., $x_{\pi(1)} \leq_{\pi(2)} \leq \cdots \leq x_{\pi(N)}$. We denote $\sigma(\pi)$ as the sign of sorting π. Note that

$$\begin{aligned}\psi_{\text{pair}}(x) &= \prod_{i<j} \mathcal{A} \circ F(x_i, x_j) \\ &= \sigma(\pi) \cdot \prod_{i<j} \mathcal{A} \circ F(x_{\pi(i)}, x_{\pi(j)})\end{aligned} \tag{6}$$

holds due to the antisymmetry of $\psi_{\text{pair}}(x)$. To represent any antisymmetric function $\Psi(x) = \Psi(x_1, \cdots, x_N)$, we denote its sign function as $\text{sgn}(\Psi)$ and its magnitude function as $|\Psi|$. Below we separately prove the universality in the case of N being an even or odd number.

(i) N is an even number ($N \geq 2$). In this case, we define ϕ_B to satisfy that $\forall 1 \leq i < j \leq N$, there is

$$\phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\}) < \phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) \tag{7}$$

and $\phi_B(x_{\pi(1)}) = 0$. For notation compactness, we let

$$\Lambda_0 = \prod_{1<j} \phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\});$$
$$\Lambda_1 = \prod_{1<i<j} \left[\phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) - \phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\})\right].$$

It is easy to know that $\Lambda_0, \Lambda_1 > 0$. In particular, we can simply set $\phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\}) = i$ for $i = 2, \cdots, N$, and then the product of $\Lambda_0\Lambda_1$ becomes a constant as

$$\Lambda_0\Lambda_1 = N! \prod_{1<i<j} (j - i), \tag{8}$$

where ! is the factorial function. Based on the above construction of ϕ_B, we can define

$$\phi_A(x_i; \{x_{\setminus i}\}) = \begin{cases} \text{sgn}(\Psi) \cdot \left(\frac{|\Psi|}{\Lambda_0\Lambda_1}\right)^{\frac{1}{N-1}}\Big|_{x_\pi}, & i = \pi(1); \\ 1, & \text{otherwise,} \end{cases}$$

where $x_\pi = (x_{\pi(1)}, \cdots, x_{\pi(N)})$. Now we can compute the terms of $i = 1$ as

$$\begin{aligned}&\prod_{1<j} \mathcal{A} \circ F(x_{\pi(1)}, x_{\pi(j)}) \\ =&\left[\phi_A(x_{\pi(1)}; \{x_{\setminus\pi(1)}\})\right]^{N-1} \prod_{1<j} \phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) \\ =&\Lambda_0 \cdot \left[\phi_A(x_{\pi(1)}; \{x_{\setminus\pi(1)}\})\right]^{N-1};\end{aligned} \tag{9}$$

and the terms of $i > 1$ as

$$\begin{aligned}&\prod_{1<i<j} \mathcal{A} \circ F(x_{\pi(i)}, x_{\pi(j)}) \\ =&\prod_{1<i<j} \left[\phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) - \phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\})\right] \\ =&\Lambda_1.\end{aligned} \tag{10}$$

Thus by combining Eq. (9) and Eq. (10), we have

$$\begin{aligned}\psi_{\text{pair}}(x) &= \sigma(\pi) \cdot \prod_{i<j} \mathcal{A} \circ F(x_{\pi(i)}, x_{\pi(j)}) \\ &= \sigma(\pi) \cdot (\Lambda_0\Lambda_1) \cdot \left[\phi_A(x_{\pi(1)}; \{x_{\setminus\pi(1)}\})\right]^{N-1} \\ &= \sigma(\pi) \cdot (\text{sgn}(\Psi))^{N-1} \cdot |\Psi|\Big|_{x_\pi} \\ &= \sigma(\pi) \cdot \Psi(x_{\pi(1)}, \cdots, x_{\pi(N)}) \\ &= \Psi(x_1, \cdots, x_N).\end{aligned} \tag{11}$$

Specially when $N = 2$, there is no term in Eq. (10) because the solution set of $1 < i < j \leq 2$ is empty. In this case we can directly set $\Lambda_1 = 1$ and the conclusion in Eq. (11) is unchanged.

(ii) N is an odd number ($N \geq 3$). In this case, we define ϕ_B to satisfy that $\phi_B(x_{\pi(1)}; \{x_{\setminus\pi(1)}\}) = -\text{sgn}(\Psi)\big|_{x_\pi}$, $\phi_B(x_{\pi(2)}; \{x_{\setminus\pi(2)}\}) = 0$, and $\forall 3 \leq i < j \leq N$, there is

$$1 < \phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\}) < \phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}). \tag{12}$$

For notation compactness, we let

$$\Lambda_0 = \prod_{2<j} \left[\phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) + \text{sgn}(\Psi)\Big|_{x_\pi}\right];$$
$$\Lambda_1 = \prod_{2<j} \phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\});$$
$$\Lambda_2 = \prod_{2<i<j} \left[\phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) - \phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\})\right].$$

It is also easy to know that $\Lambda_0, \Lambda_1, \Lambda_2 > 0$. In particular, we can simply set $\phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\}) = i - \frac{\text{sgn}(\Psi)}{2}\Big|_{x_\pi}$ for $i = 3, \cdots, N$, and then the product of $\Lambda_0\Lambda_1\Lambda_2$ becomes a constant as

$$\Lambda_0\Lambda_1\Lambda_2 = \prod_{2<j} \left(j^2 - \frac{1}{4}\right) \cdot \prod_{2<i<j} (j - i). \tag{13}$$

Based on the construction of ϕ_B, we can define

$$\phi_A(x_i; \{x_{\setminus i}\}) = \begin{cases} \text{sgn}(\Psi) \cdot \left(\frac{|\Psi|}{\Lambda_0\Lambda_1\Lambda_2}\right)^{\frac{1}{N-1}}\Big|_{x_\pi}, & i = \pi(2); \\ 1, & \text{otherwise,} \end{cases}$$

where $x_\pi = (x_{\pi(1)}, \cdots, x_{\pi(N)})$. Now we can calculate the term of $i = 1, j = 2$ as

$$\mathcal{A} \circ F(x_{\pi(1)}, x_{\pi(2)}) = \text{sgn}(\Psi) \cdot \phi_A(x_{\pi(2)}; \{x_{\setminus\pi(2)}\}); \tag{14}$$

and the terms of $i = 1$ as

$$\begin{aligned}&\prod_{2<j} \mathcal{A} \circ F(x_{\pi(1)}, x_{\pi(j)}) \\ =&\prod_{2<j} \left[\phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) + \text{sgn}(\Psi)\right] \\ =&\Lambda_0;\end{aligned} \tag{15}$$

and the terms of $i = 2$ as

$$\begin{aligned}&\prod_{2<j} \mathcal{A} \circ F(x_{\pi(2)}, x_{\pi(j)}) \\ =&\left[\phi_A(x_{\pi(2)}; \{x_{\setminus\pi(2)}\})\right]^{N-2} \prod_{2<j} \phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) \\ =&\Lambda_1 \cdot \left[\phi_A(x_{\pi(2)}; \{x_{\setminus\pi(2)}\})\right]^{N-2};\end{aligned} \tag{16}$$

and the terms of $i > 2$ as

$$\begin{aligned}&\prod_{2<i<j} \mathcal{A} \circ F(x_{\pi(i)}, x_{\pi(j)}) \\ =&\prod_{2<i<j} \left[\phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) - \phi_B(x_{\pi(i)}; \{x_{\setminus\pi(i)}\})\right] \\ =&\Lambda_2.\end{aligned} \tag{17}$$

Thus by combining from Eq. (14) to Eq. (17), we have

$$\begin{aligned}&\psi_{\text{pair}}(x) \\ =&\sigma(\pi) \cdot \prod_{i<j} \mathcal{A} \circ F(x_{\pi(i)}, x_{\pi(j)}) \\ =&\sigma(\pi) \cdot \text{sgn}(\Psi) \cdot (\Lambda_0\Lambda_1\Lambda_2) \cdot \left[\phi_A(x_{\pi(2)}; \{x_{\setminus\pi(2)}\})\right]^{N-1} \\ =&\sigma(\pi) \cdot (\text{sgn}(\Psi))^N \cdot |\Psi|\Big|_{x_\pi} \\ =&\sigma(\pi) \cdot \Psi(x_{\pi(1)}, \cdots, x_{\pi(N)}) \\ =&\Psi(x_1, \cdots, x_N).\end{aligned}$$

Specially when $N = 3$, there is no term in Eq. (17) because the solution set of $2 < i < j \leq 3$ is empty. In this case we can directly set $\Lambda_2 = 1$ and the above conclusion is unchanged. □

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. (3) and Eq. (5)), 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 in $\psi_{\text{pair}}(x)$ and each term $\mathcal{A} \circ F(x_i, x_j)$ requires $O(1)$ extra computation on top of ϕ_A and ϕ_B. Therefore, totally $\psi_{\text{pair}}(x)$ still requires $O(N^2)$ computation.

Similar to the discussion in Pfau et al. (2020), the construction of ϕ_A, ϕ_B may not be everywhere continuous due to the implicitly selected π. Therefore, in practice we still use multiple $\psi^k_{\text{pair}}(x)$ to achieve high accuracy for $k = 1, \cdots, K$, and the ensemble wavefunction $\Psi_{\text{pair}}(x)$ is constructed as

$$\Psi_{\text{pair}}(x) = \sum_k \omega_k \psi^k_{\text{pair}}(x). \tag{18}$$

where ω_k is the weight of k-th component and $\Psi_{\text{pair}}(x)$ is the wavefunction we actually use in the experiments.

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

Atom	ψ_{pair}	Exact
Li	-7.4782	-7.47806032
Be	-14.6673	-14.66736
B	-24.5602	-24.65391
C	-37.3531	-37.8450
N	-53.1855	-54.5892

3.2. Separate Modeling of Sign and Magnitude

Inui et al. (2021) and Luo and Clark (2019) propose to separately model the sign and magnitude of wavefunctions. Below we show that this kind of separate modeling can be easily implemented by our pairwise construction. Specifically, in addition to ϕ_A and ϕ_B, we let $\phi_C(x) = \phi_C(x_1, \cdots, x_N)$ be a multi-electron symmetric function, whose symmetry can be implemented via ϕ_A like $\phi_C(x) = \sum_j \phi_A(x_j; \{x_{\setminus j}\})$ or $\phi_C(x) = \prod_j \phi_A(x_j; \{x_{\setminus j}\})$. Then we can instantiate another pairwise wavefunction Ansatz:

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

where the first term is responsible for modeling the sign, and the second term $|\phi_C(x)|$ is used to model the magnitude such that it does not contribute to the sign. The instantiation of $\psi'_{\text{pair}}(x)$ corresponds to Eq. (3) where

$$\begin{aligned}&\mathcal{A} \circ F(x_i, x_j) \\ =&\text{sgn}\left(\phi_B(x_j; \{x_{\setminus j}\}) - \phi_B(x_i; \{x_{\setminus i}\})\right) \cdot |\phi_C(x)|^{\frac{2}{N(N-1)}}.\end{aligned}$$

By the similar justification as in Theorem 1, we know that $\psi'_{\text{pair}}(x)$ is antisymmetric. Besides, we could also show that the model family of $\psi'_{\text{pair}}(x)$ is universal, as stated below:

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

Proof. Similar as the proof of Theorem 2, given a certain unique ordering π such that $x_{\pi(1)} \leq_{\pi(2)} \leq \cdots \leq x_{\pi(N)}$, we use $\sigma(\pi)$ denote the sorting sign of π. For any antisymmetric function $\Psi(x) = \Psi(x_1, \cdots, x_N)$, we denote its sign function as $\text{sgn}(\Psi)$. Then we let ϕ_B to satisfy that

$$\begin{aligned}&\phi_B(x_{\pi(1)}; \{x_{\setminus\pi(1)}\}) = -\text{sgn}(\Psi)\big|_{x_\pi}; \\ &\phi_B(x_{\pi(2)}; \{x_{\setminus\pi(2)}\}) = 0; \\ &\phi_B(x_{\pi(j)}; \{x_{\setminus\pi(j)}\}) = j, \forall 3 \leq j \leq N,\end{aligned} \tag{19}$$

where $x_\pi = (x_{\pi(1)}, \cdots, x_{\pi(N)})$. Given this, we have

$$\begin{aligned}
&\mathrm{sgn}\Big(\prod_{i<j}\big(\phi_B(x_{\pi(j)};\{x_{\setminus\pi(j)}\}) - \phi_B(x_{\pi(i)};\{x_{\setminus\pi(i)}\})\big)\Big)\\
=&\mathrm{sgn}\left(\mathrm{sgn}(\Psi)\big|_{x_\pi}\prod_{3\le j\le N} j\left(j+\mathrm{sgn}(\Psi)\big|_{x_\pi}\right)\prod_{3\le i<j\le N}(j-i)\right)\\
=&\mathrm{sgn}(\Psi)\big|_{x_\pi}.
\end{aligned} \tag{20}$$

We further let ϕ_C to satisfy that

$$|\phi_C(x)| = |\Psi(x_\pi)| = \mathrm{sgn}(\Psi)\big|_{x_\pi}\cdot\Psi(x_\pi). \tag{21}$$

Finally, by combing Eq. (20) and Eq. (21), there is

$$\begin{aligned}
&\psi'_{\text{pair}}(x)\\
=&\mathrm{sgn}\Big(\prod_{i<j}\big(\phi_B(x_j;\{x_{\setminus j}\}) - \phi_B(x_i;\{x_{\setminus i}\})\big)\Big)\cdot|\phi_C(x)|\\
=&\Big(\mathrm{sgn}(\Psi)\big|_{x_\pi}\cdot\sigma(\pi)\Big)\cdot\Big(\mathrm{sgn}(\Psi)\big|_{x_\pi}\cdot\Psi(x_\pi)\Big)\\
=&\sigma(\pi)\cdot\Psi(x_\pi)\\
=&\sigma(\pi)\cdot\sigma(\pi)\cdot\Psi(x)\\
=&\Psi(x).
\end{aligned}$$

Specially when $N = 2, 3$, the final equation in Eq. (20) remains unchanged and the universality still holds. □

According to the similar analysis as in Section 3.1, we know that the computation complexity of $\psi'_{\text{pair}}(x)$ is also $O(N^2)$. In the proof construction of Corollary 1, the ordering π of input is *implicitly* defined, which can be adaptive for different inputs x. Compared to the explicitly defined ordering of input as did in Inui et al. (2021), implicit ordering of input is more flexible, allows for interaction between $|\phi_C|$ and ϕ_B, and is easier to learn in practice.

Remark. Note that when ϕ_B is a single-electron function, the term $\mathrm{sgn}\left(\prod_{i<j}(\phi_B(x_j) - \phi_B(x_i))\right)$ is not enough for modeling the sign of any antisymmetric function. Specifically, we instantiate another pairwise Ansatz $\psi''_{\text{pair}}(x)$ as

$$\psi''_{\text{pair}}(x) = \mathrm{sgn}\Big(\prod_{i<j}(\phi_B(x_j) - \phi_B(x_i))\Big)\cdot|\phi_C(x)|. \tag{22}$$

In this case, the model family of $\psi''_{\text{pair}}(x)$ is *not* universal.

Corollary 2. *There exist an antisymmetric function $\Psi(x)$, such that $\psi''_{pair}(x)$ cannot represent $\Psi(x)$.*

Proof. In a two-electron system, we have three coordinate candidates x_1, x_2, x_3. The sign of, e.g., $\psi''_{\text{pair}}(x_1, x_2)$ is decided by $\mathrm{sgn}((\phi_B(x_1) - \phi_B(x_2)))$. Assuming that the antisymmetric function $\Psi(x)$ satisfies $\mathrm{sgn}(\Psi(x_1, x_2)) = -1$, $\mathrm{sgn}(\Psi(x_1, x_3)) = 1$, $\mathrm{sgn}(\Psi(x_2, x_3)) = -1$. Then if $\psi''_{\text{pair}}(x)$ can represent $\Psi(x)$, there should be $\phi_B(x_1) < \phi_B(x_2)$, $\phi_B(x_1) > \phi_B(x_3)$, $\phi_B(x_2) < \phi_B(x_3)$, which leads to a contradiction. □

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 of them has 256 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 the 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 $\Psi_{\text{pair}}(x)$. Here we do not apply Hartree-Fock initialization. Our computational cost is less than FermiNet from two aspects: first, we use only 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, for larger atoms, our methods empirically ask for permutation-equivariant functions with higher model capacity. This can be achieved via applying more powerful network architectures, and we leave this to future exploration.

4. Conclusion

In this paper, we develop 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).

Limitations. Although theoretically we demonstrate the universality of pairwise constructions, empirically we may require a large model capacity to fit the ground-truth ϕ_A, ϕ_B, especially on their discontinuous points. 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 in this case the neural network part dominates the computation.

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.

James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, and Skye Wanderman-Milne. Jax: composable transformations of python+ numpy programs, 2018. *URL http://github. com/google/jax*, 4:16, 2020.

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.

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.

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.

Di Luo and Bryan K Clark. Backflow transformations via neural networks for quantum many-body wave functions. *Physical review letters*, 122(22):226401, 2019.

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.

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.

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

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.