Journal of Computational Physics 399 (2019) 108929

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Solving many-electron Schrödinger equation using deep neural networks

Jiequn Han [a], Linfeng Zhang [a,*], Weinan E [a,b,c]

[a] Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544, USA
[b] Department of Mathematics, Princeton University, Princeton, NJ 08544, USA
[c] Beijing Institute of Big Data Research, Beijing, 100871, PR China

ARTICLE INFO

Article history:
Received 3 December 2018
Received in revised form 13 August 2019
Accepted 1 September 2019
Available online 17 September 2019

Keywords:
Schrödinger equation
Variational Monte Carlo
Deep neural networks
Trial wave-function

ABSTRACT

We introduce a new family of trial wave-functions based on deep neural networks to solve the many-electron Schrödinger equation. The Pauli exclusion principle is dealt with explicitly to ensure that the trial wave-functions are physical. The optimal trial wave-function is obtained through variational Monte Carlo and the computational cost scales quadratically with the number of electrons. The algorithm does not make use of any prior knowledge such as atomic orbitals. Yet it is able to represent accurately the ground-states of the tested systems, including He, H_2, Be, B, LiH, and a chain of 10 hydrogen atoms. This opens up new possibilities for solving large-scale many-electron Schrödinger equation.

1. Introduction

An accurate quantum mechanical treatment of the interaction between many electrons and ions is the foundation for modeling physical, chemical, and biological systems. Theoretically, these systems are described by the many-electron Schrödinger equation, which consists of the kinetic and Coulomb interaction terms [1]. Solving the Schrödinger equation accurately for real physical systems has been prohibitively difficult due to the high dimensionality of the Hilbert space involved, the high degree of entanglement produced by the electron-electron and electron-ion interactions, and the notorious Pauli exclusion principle imposed on the wave-function, i.e., the wave-function has to change its sign when two identical electrons exchange places [2].

Developing efficient algorithms for this problem is among the most heroic endeavors in computational science and has achieved remarkable successes. An incomplete list of the major methodologies developed so far includes the Hartree-Fock (HF) based methods [3,4], configuration interaction (CI) based methods [5–8], coupled cluster (CC) based schemes [9–12], Monte Carlo-based approaches [13–20], and the more recently developed density matrix renormalization group (DMRG) theory [21–25] and density matrix embedding theory (DMET) [26–28]. We refer to Refs. [29–31] for a more detailed review of these advances.

Of particular interest to this work is the variational Monte Carlo (VMC) scheme which uses variational principle and Monte Carlo sampling to obtain the best parametrized trial wave-function [13–15,20]. Naturally the key component is the representation of the trial wave-functions. The most commonly used trial wave-functions typically consist of an anti-symmetric Slater determinant [32] multiplied by a symmetric Jastrow correlation factor [33]. There have been tremendous

* Corresponding author.
E-mail addresses: jiequnh@princeton.edu (J. Han), linfengz@princeton.edu (L. Zhang), weinan@math.princeton.edu (W. E).

https://doi.org/10.1016/j.jcp.2019.108929
0021-9991/

efforts on improving the nodal surface (a subspace on which the function value equals zero and across which it changes the sign) of the anti-symmetric part of the trial wave-function and the representability of the symmetric part [34–37].

With remarkable advances in many fields such as computer vision and speech recognition, the deep neural network (DNN) has shown great capacity in approximating high-dimensional functions (see, e.g., review [38] and the references therein). Furthermore, DNN has been successfully used in solving general high-dimensional partial differential equations [39–42] and certain quantum many-body problems for Bosonic and lattice systems [43–47]. However, there have been few attempts to solve the many-electron Schrödinger equations based on DNN, and this constitutes the main objective of this work.

To achieve this, we develop a general and efficient DNN representation for the many-electron wave-function satisfying the Pauli exclusion principle. The resulted trial wave-function can naturally fit into the framework of VMC to optimize the parameters in our model. As preliminary tests, we show that this DNN-based trial wave-function is able to produce reasonably well ground-state energies for some small systems, such as Be, B, LiH, and a chain of 10 hydrogen atoms (H_{10}). In addition, learning from scratch without any prior knowledge and without resorting to a reference of atomic bases, the DNN-based wave-function is able to reproduce the electronic structures of the tested systems. We call the methodology introduced here the Deep WaveFunction method, abbreviated DeepWF. This paper only reports our initial results. There is still a huge room for improvement.

2. Method

We consider a system of N electrons and M ions, under the Born-Oppenheimer approximation [48]. This system is described by the Hamiltonian

$$\hat{H} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{1}{|\boldsymbol{r}_i-\boldsymbol{r}_j|} - \sum_{I=1}^{M}\sum_{i=1}^{N}\frac{Z_I}{|\boldsymbol{r}_i-\boldsymbol{R}_I|} + \sum_{I=1}^{M}\sum_{J=I+1}^{M}\frac{Z_I Z_J}{|\boldsymbol{R}_I-\boldsymbol{R}_J|}, \tag{1}$$

where $\boldsymbol{r}=(\boldsymbol{r}_1,\dots,\boldsymbol{r}_N)$ and $\boldsymbol{R}=(\boldsymbol{R}_1,\dots,\boldsymbol{R}_M)$ are the coordinates of the electrons and the ions, respectively, and Z_I denotes the nuclear charge. Since $\hat{H}$ is spin-independent, it is valid to assume that the first $N_\uparrow$ electrons are of spin-up and the remaining $N_\downarrow = N - N_\uparrow$ electrons are of spin-down [20]. Accordingly, we can write the wave-function in a spin-independent form $\Psi(\boldsymbol{r};\boldsymbol{R})$. Let N_{tp} be the number of ion types in the system, then there are $N_{\text{tp}}+2$ types of particles in total, taking into account the electrons of spin-up and spin-down separately.

In this work we restrict our attention to the ground state of the system. The variational principle states that the wave-function associated with the ground state minimizes within the required symmetry the following energy functional

$$E[\Psi] = \frac{\int \Psi^*(\boldsymbol{r};\boldsymbol{R})\hat{H}\Psi(\boldsymbol{r};\boldsymbol{R})\,d\boldsymbol{r}}{\int \Psi^*(\boldsymbol{r};\boldsymbol{R})\Psi(\boldsymbol{r};\boldsymbol{R})\,d\boldsymbol{r}}.$$

We now discuss how to represent the wave-function $\Psi(\boldsymbol{r};\boldsymbol{R})$ with DNN. The trial wave-functions are assumed to be real-valued. The primary goal is to ensure that the represented wave-function satisfies the anti-symmetry property. To this end, we decompose the wave-function as

$$\Psi(\boldsymbol{r};\boldsymbol{R}) = S(\boldsymbol{r};\boldsymbol{R})A^\uparrow(\boldsymbol{r}^\uparrow)A^\downarrow(\boldsymbol{r}^\downarrow),$$

where $\boldsymbol{r}^\uparrow, \boldsymbol{r}^\downarrow$ denote the positions of spin-up and spin-down electrons, respectively. We require $S(\boldsymbol{r},\boldsymbol{R})$ to be symmetric and $A^\uparrow(\boldsymbol{r}^\uparrow)$, $A^\downarrow(\boldsymbol{r}^\downarrow)$ to be both anti-symmetric. As an analogy, one can view $S(\boldsymbol{r};\boldsymbol{R})$ as a Jastrow factor and $A^\uparrow(\boldsymbol{r}^\uparrow)$, $A^\downarrow(\boldsymbol{r}^\downarrow)$ as Slater determinant-like functions. In practice, $\log(\Psi^2(\boldsymbol{r};\boldsymbol{R}))$ is more convenient for implementing VMC. From this perspective, the introduced decomposition of wave-function becomes

$$\log(\Psi^2(\boldsymbol{r};\boldsymbol{R})) = \tilde{S}(\boldsymbol{r};\boldsymbol{R}) + \log(|A^\uparrow(\boldsymbol{r}^\uparrow)|^2) + \log(|A^\downarrow(\boldsymbol{r}^\downarrow)|^2), \tag{2}$$

where $\tilde{S}(\boldsymbol{r};\boldsymbol{R}) := \log(S^2(\boldsymbol{r};\boldsymbol{R}))$ is still a general symmetric function. On the other hand, although being symmetric, $\log(|A^\uparrow(\boldsymbol{r}^\uparrow)|^2)$ and $\log(|A^\downarrow(\boldsymbol{r}^\downarrow)|^2)$ have singularities on the nodal surface. Therefore we need to secure this property by representing $A^\uparrow(\boldsymbol{r}^\uparrow)$ and $A^\downarrow(\boldsymbol{r}^\downarrow)$ directly. In the following we discuss how to represent $A^\uparrow(\boldsymbol{r}^\uparrow)$, $A^\downarrow(\boldsymbol{r}^\downarrow)$, and $\tilde{S}(\boldsymbol{r};\boldsymbol{R})$ with DNN and how to learn the optimal parameters using VMC. See Fig. 1 for a schematic illustration.

The basic building block for the anti-symmetric function $A^\uparrow(\boldsymbol{r}^\uparrow)$ is an ansatz

$$\boldsymbol{a}^\uparrow(\boldsymbol{r}^\uparrow) = \prod_{1\le i<j\le N_\uparrow}\boldsymbol{a}_0^\uparrow(\boldsymbol{r}_i,\boldsymbol{r}_j), \tag{3}$$

where $\boldsymbol{a}_0^\uparrow$ is a general two-body anti-symmetric function with M_1-dimensional output, i.e., $\boldsymbol{a}_0^\uparrow(\boldsymbol{r}_i,\boldsymbol{r}_j) = -\boldsymbol{a}_0^\uparrow(\boldsymbol{r}_j,\boldsymbol{r}_i)$, and the products are component-wise. In (3) we have assumed $N_\uparrow \ge 2$. Such representation can be viewed as a generalization of the Laughlin wave-function [49], which describes well the anomalous quantum Hall effect. A related model is the so-called correlator product state (CPS) [37], which works well for lattice systems and has natural connections with some recently

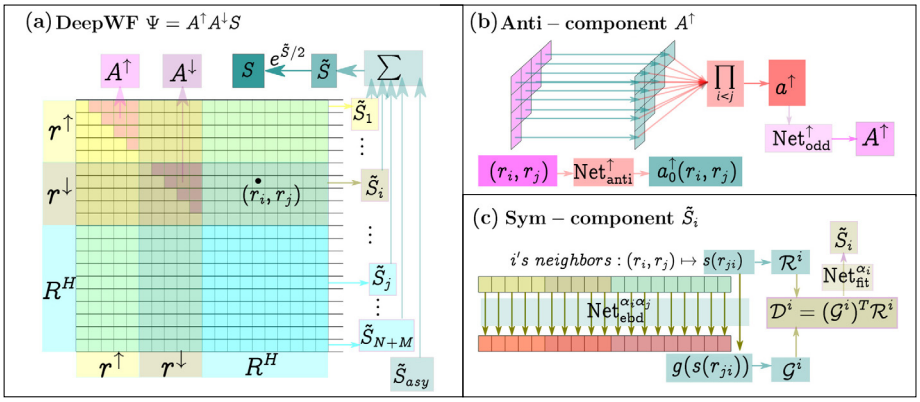

Fig. 1. (Color online.) Schematic plot of DeepWF, taking H_{10} as an example. $\boldsymbol{r}^{\uparrow}$ and $\boldsymbol{r}^{\downarrow}$ denote the positions of 5 electrons of spin-up and 5 of spin-down, respectively, and $\boldsymbol{R}^H$ denotes the positions of 10 hydrogen atoms. (a) The entire wave-function $\Psi(\boldsymbol{r};\boldsymbol{R}) = S(\boldsymbol{r};\boldsymbol{R})A^{\uparrow}(\boldsymbol{r}^{\uparrow})A^{\downarrow}(\boldsymbol{r}^{\downarrow})$, decomposed into symmetric function $S(\boldsymbol{r};\boldsymbol{R})$ and two anti-symmetric functions $A^{\uparrow}(\boldsymbol{r}^{\uparrow})$, $A^{\downarrow}(\boldsymbol{r}^{\downarrow})$, with respect to spin-up and spin-down electrons; (b) anti-symmetric function $A^{\uparrow}(\boldsymbol{r}^{\uparrow})$ for spin-up electrons; (c) symmetric function $\tilde{S}_i(\mathcal{R}^i)$ as the contribution of particle i in the symmetric function $\log(S^2(\boldsymbol{r};\boldsymbol{R}))$. See text for details.

proposed neural-network quantum states [50,51]. In practice, we let $\boldsymbol{a}_0^{\uparrow}(\boldsymbol{r}_i,\boldsymbol{r}_j) = \mathrm{Net}^{\uparrow}_{\mathrm{anti}}(\boldsymbol{r}_i,\boldsymbol{r}_j,r_{ji}) - \mathrm{Net}^{\uparrow}_{\mathrm{anti}}(\boldsymbol{r}_j,\boldsymbol{r}_i,r_{ji})$, where $\mathrm{Net}^{\uparrow}_{\mathrm{anti}}(\cdot)$ is represented by a DNN. By convention $r_{ji} = |\boldsymbol{r}_{ji}| = |\boldsymbol{r}_j - \boldsymbol{r}_i|$ denotes the Euclidean distance between particles i and j. Next $\boldsymbol{a}^{\uparrow}$ is fed into another DNN $\mathrm{Net}^{\uparrow}_{\mathrm{odd}}$, which outputs a scalar and satisfies $\mathrm{Net}^{\uparrow}_{\mathrm{odd}}(\boldsymbol{x}) = -\mathrm{Net}^{\uparrow}_{\mathrm{odd}}(-\boldsymbol{x})$. We can further adjust its scale and define $A^{\uparrow}(\boldsymbol{r}^{\uparrow})$ through $\log|A^{\uparrow}(\boldsymbol{r}^{\uparrow})| = f_{\mathrm{odd}}\log|\mathrm{Net}^{\uparrow}_{\mathrm{odd}}(\boldsymbol{a}^{\uparrow}(\boldsymbol{r}))|$ where f_{odd} is a positive scalar factor. The sign of $A^{\uparrow}(\boldsymbol{r}^{\uparrow})$ is taken as the same as $\mathrm{Net}^{\uparrow}_{\mathrm{odd}}(\boldsymbol{a}^{\uparrow}(\boldsymbol{r}^{\uparrow}))$ such that $A^{\uparrow}(\boldsymbol{r}^{\uparrow})$ inherits the anti-symmetry. The same construction introduced above are applied to spin-down electrons and yield $A^{\downarrow}(\boldsymbol{r}^{\downarrow})$.

While DNN-based anti-symmetric function has seldom been investigated in the literature, DNN-based symmetric function has gained wide attentions recently in modeling many-body potential energy surface [52–56], free energy surface [57–59], etc. For instance, the Deep Potential-Smooth Edition (DeepPot-SE) model is able to efficiently describe the interatomic potential energy of both finite and extended systems [55]. A crucial step there is to faithfully map the input atomic coordinates onto a symmetry-preserving feature space. Therefore, it shares the symmetric property of the Jastrow factor, and might become a more general Jastrow factor that is able to better capture the many-body correlations between the electrons and ions.

We extend the DeepPot-SE model to construct $\tilde{S}(\boldsymbol{r};\boldsymbol{R})$. For later convenience, we regard $\boldsymbol{R}$ as $(\boldsymbol{r}_{N+1},\ldots,\boldsymbol{r}_{N+M})$ and let $\bar{\boldsymbol{r}} := (\boldsymbol{r}_1,\cdots,\boldsymbol{r}_{N+M})$. For each particle i, we consider its neighboring environment in terms of the Cartesian coordinates:

$$\mathcal{R}^i = \{\boldsymbol{s}(\boldsymbol{r}_{1i})^T,\cdots,\boldsymbol{s}(\boldsymbol{r}_{ji})^T,\cdots\}^T \in \mathbb{R}^{(N+M-1)\times 4}, \tag{4}$$

where $\boldsymbol{s}(\boldsymbol{r}_{ji}) = (\boldsymbol{r}_{ji}/r_{ji}^2, 1/r_{ji}) \in \mathbb{R}^4$, $j \in \{1,\ldots,N+M\}\backslash\{i\}$, are extended relative coordinates. The motivation of introducing $\boldsymbol{s}(\boldsymbol{r}_{ji})$ is twofold. First, it is found that providing a direct distance dependency in the construction of the network is usually beneficial. Second, the factor $1/r_{ji}$ automatically reduces the weight of the particles j that are more distant from atom i. Given the local environment $\mathcal{R}^i$, we decompose $\tilde{S}(\boldsymbol{r};\boldsymbol{R})$ into the sum of contributions from each particle and an extra asymptotic term, i.e.,

$$\tilde{S}(\boldsymbol{r};\boldsymbol{R}) = \tilde{S}_{\mathrm{asy}}(\bar{\boldsymbol{r}}) + \sum_{i=1}^{N+M}\tilde{S}_i(\mathcal{R}^i). \tag{5}$$

We define $\tilde{S}_i$ in a way that guarantees the symmetric property. We first construct a series of embedding networks $\mathrm{Net}^{\alpha_i\alpha_j}_{\mathrm{ebd}}$, each mapping the extended relative coordinates $\boldsymbol{s}(\boldsymbol{r}_{ji})$ to M_2 components. Here α_i denotes the type of particle i, and the parameters of $\mathrm{Net}^{\alpha_i\alpha_j}_{\mathrm{ebd}}$ depend on the types of both particle i and its neighbor particle j. The embedding matrix $\mathcal{G}^i \in \mathbb{R}^{(N+M-1)\times M_2}$ and the encoded feature matrix $\mathcal{D}^i \in \mathbb{R}^{M_2\times 4}$ of particle i are then defined as

$$(\mathcal{G}^i)_{jk} = (\mathrm{Net}^{\alpha_i\alpha_j}_{\mathrm{ebd}}(\boldsymbol{s}(\boldsymbol{r}_{ji})))_k, \tag{6}$$

$$\mathcal{D}^i = (\mathcal{G}^i)^T\mathcal{R}^i. \tag{7}$$

By definition one can check that the feature matrix $\mathcal{D}^i$ is invariant under the permutation of particles in the same type, and its $M_2\times 4$ components are finally reshaped into a vector to serve as the input of a fitting network $\mathrm{Net}^{\alpha_i}_{\mathrm{fit}}$ and yield $\tilde{S}_i$.

The remaining term $\tilde{S}_{\mathrm{asy}}(\bar{\boldsymbol{r}})$ should also be symmetric and serve the additional purposes of making the trial wave-function close to some desired asymptotic properties. This term has the form

$$\tilde{S}_{\text{asy}}(\bar{\boldsymbol{r}}) = -\sum_{k=N+1}^{N+M} Z_{k-N} \sum_{i=1}^{N} \frac{r_{ik} + f_{\text{dec}} r_{ik}^2}{r_{ik} + 0.5} + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{r_{ij}}{r_{ij} + 1}, \tag{8}$$

where f_{dec} is a positive scalar as decay factor. The first purpose is to ensure that $\Psi(\boldsymbol{r}; \boldsymbol{R})$ is square integrable when f_{dec} is large enough such that $\Psi(\boldsymbol{r}; \boldsymbol{R})$ has exponential decay for the electrons that are far away. The second purpose is related to the so-called *cusp* conditions, meaning that the trial wave-function has the correct singular behavior when two particles are close to each other [60]. If we let $\log(\Psi^2(\boldsymbol{r}; \boldsymbol{R})) = \tilde{S}_{\text{asy}}(\bar{\boldsymbol{r}})$, it can be shown from (8) that for $1 \le i < j \le N,\ N+1 \le k \le N+M$,

$$\frac{1}{\Psi} \frac{\partial \Psi}{\partial r_{ik}}\bigg|_{r_{ik}=0} = -Z_{k-N}, \quad \frac{1}{\Psi} \frac{\partial \Psi}{\partial r_{ij}}\bigg|_{r_{ij}=0} = \frac{1}{2}. \tag{9}$$

This indicates $\Psi(\boldsymbol{r}; \boldsymbol{R})$ meets the cusp conditions of both the electron-nuclear pairs and unlike spin pairs (see e.g., [20,60] for more details). In numerical computation we find that adding $\tilde{S}_{\text{asy}}(\bar{\boldsymbol{r}})$ makes the training process much more stable, even though the cusp conditions are not strictly satisfied due to the influence of $\tilde{S}_i$ and $A(\boldsymbol{r})$ on the terms in (9).

We denote all the parameters in the DNNs ($\text{Net}^{\uparrow}_{\text{anti}}$, $\text{Net}^{\downarrow}_{\text{anti}}$, $\text{Net}^{\uparrow}_{\text{odd}}$, $\text{Net}^{\downarrow}_{\text{odd}}$, $\text{Net}^{\alpha_i\alpha_j}_{\text{ebd}}$, and $\text{Net}^{\alpha_i}_{\text{fit}}$), the two scalar factors (f_{odd} and f_{dec}) together as $\boldsymbol{\theta}$ and the associated trial wave-function as $\Psi_{\boldsymbol{\theta}}(\boldsymbol{r})$ (below for convenience we ignore the dependence on $\boldsymbol{R}$, the clamped ion positions). $\boldsymbol{\theta}$ are initialized randomly from a Gaussian distribution without any pre-training on pre-calculated wave-functions. We use VMC to optimize the trial wave-function. Specifically, we keep track of N_{wk} walkers to approximate the squared wave-function through an empirical distribution. The learning process consists of two phases, the sampling phase and the optimization phase, which we proceed with alternatively. In the sampling phase of step t, we run several steps of the Metropolis-Hasting algorithm to update the positions of the walkers, according to a target probability proportional to $\Psi^2_{\boldsymbol{\theta}_t}(\boldsymbol{r})$. In the optimization phase of step t, we aim to minimize the second moment of the local energy $E_{\text{loc}}(\boldsymbol{r})$ with respect to a fixed reference energy E_{ref}. Here the local energy is defined as $E_{\text{loc}}(\boldsymbol{r}) \coloneqq \hat{H}\Psi_{\boldsymbol{\theta}}(\boldsymbol{r})/\Psi_{\boldsymbol{\theta}}(\boldsymbol{r})$. Accordingly, the objective function has the explicit form

$$\Omega_{E_{\text{ref}}}(\boldsymbol{\theta}) \coloneqq \frac{\int \Psi^2_{\boldsymbol{\theta}}(\boldsymbol{r})(E_{\text{loc}}(\boldsymbol{r}) - E_{\text{ref}})^2 \, d\boldsymbol{r}}{\int \Psi^2_{\boldsymbol{\theta}}(\boldsymbol{r}) \, d\boldsymbol{r}}.$$

In practice, E_{ref} is a constant chosen according to the benchmark energies from quantum chemistry calculations. It is slightly lower than the benchmark energies in order to guarantee that the ideal minimum for the optimization process is the exact ground state energy. In numerical computation, in order to reduce the variance when evaluating the objective function, we use a technique called correlated sampling. Assuming the walkers' current positions $\{\boldsymbol{r}^{(1)}, \boldsymbol{r}^{(2)}, \ldots, \boldsymbol{r}^{(N_{\text{wk}})}\}$ approximate the distribution $\Psi^2_{\boldsymbol{\theta}_t}(\boldsymbol{r})$ well, we define a reweighting factor $w_i \coloneqq \Psi_{\boldsymbol{\theta}}(\boldsymbol{r}^{(i)})/\Psi_{\boldsymbol{\theta}_t}(\boldsymbol{r}^{(i)})$ and rewrite the objective function as

$$\begin{aligned}\Omega_{E_{\text{ref}}}(\boldsymbol{\theta}) &= \frac{\int \Psi^2_{\boldsymbol{\theta}_t}(\boldsymbol{r})(\Psi^2_{\boldsymbol{\theta}}(\boldsymbol{r})/\Psi^2_{\boldsymbol{\theta}_t}(\boldsymbol{r}))(E_{\text{loc}}(\boldsymbol{r}) - E_{\text{ref}})^2 \, d\boldsymbol{r}}{\int \Psi^2_{\boldsymbol{\theta}_t}(\boldsymbol{r})(\Psi^2_{\boldsymbol{\theta}}(\boldsymbol{r})/\Psi^2_{\boldsymbol{\theta}_t}(\boldsymbol{r})) \, d\boldsymbol{r}} \\ &\approx \frac{\sum_{i=1}^{N_{\text{wk}}} w_i^2 (E_{\text{loc}}(\boldsymbol{r}^{(i)}) - E_{\text{ref}})^2}{\sum_{i=1}^{N_{\text{wk}}} w_i^2}.\end{aligned}$$

Note here we take $\boldsymbol{\theta}_t$ as constants and view both w_i and $E_{\text{loc}}(\boldsymbol{r}^{(i)})$ as functions of the parameters $\boldsymbol{\theta}$ given the walker's position $\boldsymbol{r}^{(i)}$. The gradients $\nabla_{\boldsymbol{\theta}}\Omega_{E_{\text{ref}}}(\boldsymbol{\theta})$ are computed by the backpropagation algorithm and used to update the parameters through $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t - \eta \nabla_{\boldsymbol{\theta}}\Omega_{E_{\text{ref}}}(\boldsymbol{\theta}_t)$, with η being the learning rate. Motivated by the idea of stochastic gradient descent (SGD), we actually evaluate $\nabla_{\boldsymbol{\theta}}\Omega_{E_{\text{ref}}}(\boldsymbol{\theta})$ with a random batch of data $i \in \mathcal{B} \subseteq \{1, \ldots, N_{\text{wk}}\}$ and update $\boldsymbol{\theta}_t$ with a few steps.

3. Results and discussion

Using the algorithm described above, for a series of benchmark systems, we obtain the optimized parameters $\boldsymbol{\theta}^*$ and run further Monte Carlo simulations for several statistical properties. In Table 1, we present the ground state energies of H_2, He, LiH, Be, B, and a chain of 10 hydrogen atoms under open boundary conditions. The network structures and training schemes for these systems, as well as the estimated local energies during the process for the Be atom and LiH, are detailed in the appendix (Table A.1, Fig. A.1). In consideration of the spin symmetry, we set $N_{\uparrow} = N_{\downarrow}$ for all the systems with an even number of electrons. For B atom, we set $N_{\uparrow}$ to 3 and set $N_{\downarrow}$ to 2. Note that for H_2 and He we don't have the anti-symmetric part in the wave-function. Overall the ground-state energies obtained by DeepWF show great consistency with the benchmarks. However, as the number of electrons increases, the accuracy deteriorates.

In Fig. 2, we plot the potential energy curves of H_2 and H_{10} as a function of bond length. The result for H_2 is very close to the benchmark. For H_{10}, the DeepWF shows a good relative energy, with the correct prediction of local minimum, but is still a bit above the benchmark result.

Table 1
Ground-state energy of several systems obtained by DeepWF. The bond length of H_2, LiH, and H_{10} is 1.4, 1.62, and 1.8 *a.u.*, respectively. The benchmarks for H_2, He, LiH, Be, and B are taken from the Computational Chemistry Comparison and Benchmark (CCCBDB) DataBase [61], at the level of configuration interaction, singles and doubles (CISD) theory. The benchmark for H_{10} is taken from [31], at the level of multi-reference configuration with the Davidson correction (MRCI+Q) [7]. All the benchmark results are extrapolated to the complete basis set (CBS) limit. The relative difference (Rel. Diff) is reported. We notice that since we try to use a consistent and accurate theory for the benchmark, the corresponding energies may not be the lowest state-of-art values. For example, in Ref. [62], C.L. Pekeris et al. already obtained a value of -2.903724351 *a.u.* for the ground-state energy of He.

System	DeepWF [*a.u.*]	Benchmark [*a.u.*]	Rel. Diff
H_2	-1.1738	-1.1741	0.26%
He	-2.9036	-2.9029	-0.02%
LiH	-7.8732	-8.0243	1.88%
Be	-14.6141	-14.6190	0.03%
B	-24.2124	-24.6006	1.58%
H_{10}	-5.5685	-5.6655	1.71%

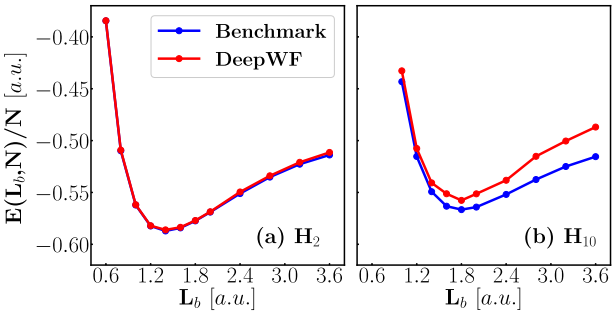

Fig. 2. Potential energy per atom $E(L_b, N)/N$ vs. bond length L_b for (a) H_2 ($N = 2$), and (b) H_{10} ($N = 10$). The benchmark result for H_2 is obtained using the PySCF package [63], at the level of full CI. The benchmark result for H_{10} is taken from [31], at the level of MRCI+Q [7]. Both the benchmark results are extrapolated to the CBS limit.

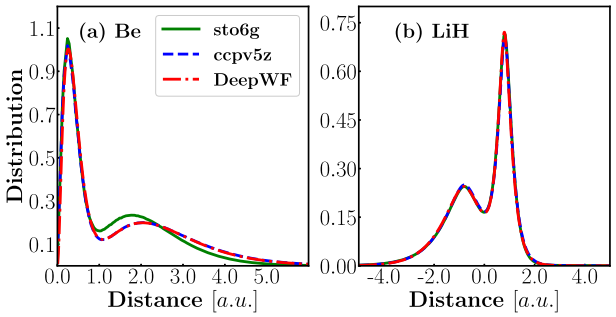

Fig. 3. Spatial distribution functions of electrons in different systems. (a) Radial distribution of 4 electrons in a Be atom; (b) axial distribution of 4 electrons in a LiH molecule. The results obtained using the PySCF package [63] with the sto6g and the ccpv5z bases are plotted for comparison.

Finally, Fig. 3 plots the spatial distributions of electrons in different systems. In the case of radial distribution function of electrons in a Be atom (Fig. 3 (a)), it is remarkable that the DeepWF learns from scratch the shell structure of the electrons and shows great consistency with the result obtained using the ccpv5z basis [64]. As a comparison, the sto6g basis [65] is not enough to describe the electronic dispersion. In the case of axial distribution function of electrons in a LiH molecule (Fig. 3 (b)), DeepWF and the other two methods show excellent agreement.

What is appealing to us is the simplicity of the proposed approach. Obviously there is a huge room for further improvement. In terms of the representation of DeepWF, the symmetric part is relatively general, considering the success of a similar version in representing the inter-atomic potential energy surface. However, the anti-symmetric ansatz, although appealing due to its quadratic scaling, might not be sufficient in representing the electronic correlations caused by the Pauli exclusion rule. How to impose the exact cusp conditions in the wavefunction regardless of the construction of neural networks is another direction for future investigation. Second, in terms of optimization, techniques for accelerating VMC through more efficient sampling (see e.g. [66,67]) can be directly adapted into our DeepWF method. Optimization method other than the

SGD-like correlated sampling method can also be employed. In any case, we hope that ideas presented here will add some ammunition to the heroic endeavor of attempting to solve the many-body Schrödinger equation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors acknowledge M. Motta for helpful discussions. This work is supported in part by Major Program of NNSFC under grant 91130005, ONR grant N00014-13-1-0338 and NSFC grant U1430237. We are grateful for the computing time provided by the High-performance Computing Platform of Peking University and the TIGRESS High Performance Computing Center at Princeton University.

Appendix A. Details of the training process

Table A.1
Hyperparameters related to network structure and training scheme for each system.

System	Net_{anti}[b]	$\text{Net}^{\alpha_i\alpha_j}_{\text{ebd}}$	$\text{Net}^{\alpha_i}_{\text{fit}}$	LR scheme[♮]	E_{ref}[♯]
H_2	-	(20, 40, 80)	(160, 80, 40)	(5e-4, 1e-8)	-1.3
He	-	(20, 40, 80)	(160, 80, 40)	(5e-4, 1e-8)	-3.0
LiH	(40, 40, 40)	(20, 40, 80)	(160, 80, 40)	(5e-4, 1e-8)	-10.0
Be	(40, 40, 40)	(20, 40, 80)	(160, 80, 40)	(5e-4, 1e-8)	-16.0
B	(40, 40, 40)	(20, 40, 80)	(160, 80, 40)	(5e-4, 1e-8)	-26.0
H_{10}	(40, 40, 40)	(20, 40, 80)	(160, 80, 40)	(1e-4, 1e-8)	-6.0

[b] The network is represented by the number of nodes in each hidden layer, from input to output. The activation function is always hyperbolic tangent, i.e., $\sigma(x) = \tanh(x)$. Net_{anti} is not used for two-electron systems H_2 and He. The parameters of $\text{Net}^{\uparrow}_{\text{anti}}$ and $\text{Net}^{\downarrow}_{\text{anti}}$ are shared except in the case of B, where the number of spin-up and spin-down electrons are different by one. $\text{Net}^{\uparrow}_{\text{odd}}$ and $\text{Net}^{\downarrow}_{\text{odd}}$ are chosen to be identity functions in this work, since we find no significant effect of network approximation to this term in the tested systems.

[♮] We use Adam stochastic gradient descent method [68] to optimize all the parameters. The first number denotes the starting learning rate and the second denotes the ending learning rate. The learning rate is a exponential decay function respect to the number of epochs, where the decay factor is determined by the starting/ending learning rates and total number of epochs. In all the tested systems, we call 24 batch stochastic gradient descent iterations as an epoch and allow each sample to be used multiple times in an epoch. The batch size and total number of epochs are always 256 and 5000, respectively.

[♯] The hyperparameters related to VMC sampling are the same for all the tested systems except the reference energy E_{ref}. We use $N_{\text{wk}} = 2048$ walkers and run 80 steps of Metropolis-Hasting sampling each time after one epoch iteration of parameters. The proposal distribution is Gaussian, with the standard deviation being adjusted on the fly such that the acceptance rate stays between 15% and 75%.

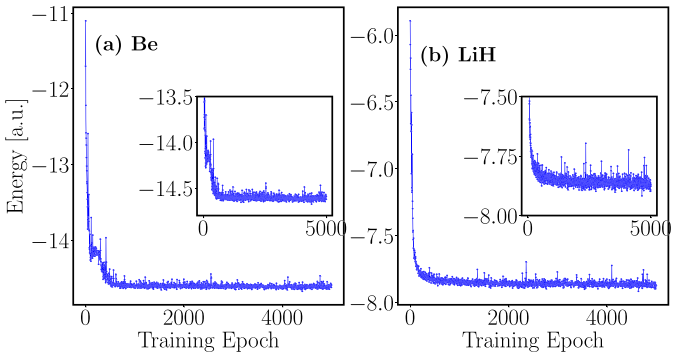

Fig. A.1. The estimated local energies during the training process for the Be atom and LiH.

References

[1] P.A.M. Dirac, Quantum mechanics of many-electron systems, Proc. R. Soc. A 123 (792) (1929) 714–733.
[2] W. Pauli, Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren, Z. Phys. 31 (1) (1925) 765–783.
[3] C. Roothaan, Self-consistent field theory for open shells of electronic systems, Rev. Mod. Phys. 32 (2) (1960) 179–185.
[4] J. Pople, R. Nesbet, Self-consistent orbitals for radicals, J. Chem. Phys. 22 (3) (1954) 571–572.

[5] J.A. Pople, M. Head-Gordon, K. Raghavachari, Quadratic configuration interaction. A general technique for determining electron correlation energies, J. Chem. Phys. 87 (10) (1987) 5968–5975.
[6] H.-J. Werner, P.J. Knowles, An efficient internally contracted multiconfiguration–reference configuration interaction method, J. Chem. Phys. 89 (9) (1988) 5803–5814.
[7] P.J. Knowles, H.-J. Werner, An efficient method for the evaluation of coupling coefficients in configuration interaction calculations, Chem. Phys. Lett. 145 (6) (1988) 514–522.
[8] T. Shiozaki, G. Knizia, H.-J. Werner, Explicitly correlated multireference configuration interaction: MRCI-F12, J. Chem. Phys. 134 (3) (2011) 034113.
[9] G.D. Purvis III, R.J. Bartlett, A full coupled-cluster singles and doubles model: the inclusion of disconnected triples, J. Chem. Phys. 76 (4) (1982) 1910–1918.
[10] J. Paldus, X. Li, A critical assessment of coupled cluster method in quantum chemistry, Adv. Chem. Phys. 110 (1999) 1–175.
[11] R.J. Bartlett, M. Musiał, Coupled-cluster theory in quantum chemistry, Rev. Mod. Phys. 79 (1) (2007) 291.
[12] I. Shavitt, R.J. Bartlett, Many-Body Methods in Chemistry and Physics: MBPT and Coupled-Cluster Theory, Cambridge University Press, 2009.
[13] W.L. McMillan, Ground state of liquid He 4, Phys. Rev. 138 (2A) (1965) A442.
[14] D. Ceperley, G. Chester, M. Kalos, Monte Carlo simulation of a many-fermion study, Phys. Rev. B 16 (7) (1977) 3081–3099.
[15] D. Bressanini, P.J. Reynolds, Between classical and quantum Monte Carlo methods: "Variational" QMC, Adv. Chem. Phys. 105 (1999) 37–64.
[16] R. Blankenbecler, D. Scalapino, R. Sugar, Monte Carlo calculations of coupled boson-fermion systems. I, Phys. Rev. D 24 (8) (1981) 2278.
[17] S. Zhang, J. Carlson, J. Gubernatis, Constrained path Monte Carlo method for fermion ground states, Phys. Rev. B 55 (12) (1997) 7464.
[18] S. Zhang, H. Krakauer, Quantum Monte Carlo method using phase-free random walks with Slater determinants, Phys. Rev. Lett. 90 (13) (2003) 136401.
[19] K. Van Houcke, E. Kozik, N. Prokof'ev, B. Svistunov, Diagrammatic Monte Carlo, Phys. Proc. 6 (2010) 95–105.
[20] W. Foulkes, L. Mitas, R. Needs, G. Rajagopal, Quantum Monte Carlo simulations of solids, Rev. Mod. Phys. 73 (1) (2001) 33.
[21] S.R. White, Density matrix formulation for quantum renormalization groups, Phys. Rev. Lett. 69 (19) (1992) 2863.
[22] S.R. White, R.L. Martin, Ab initio quantum chemistry using the density matrix renormalization group, J. Chem. Phys. 110 (9) (1999) 4127–4130.
[23] G.K.-L. Chan, M. Head-Gordon, Highly correlated calculations with a polynomial cost algorithm: a study of the density matrix renormalization group, J. Chem. Phys. 116 (11) (2002) 4462–4476.
[24] G.K.-L. Chan, A. Keselman, N. Nakatani, Z. Li, S.R. White, Matrix product operators, matrix product states, and ab initio density matrix renormalization group algorithms, J. Chem. Phys. 145 (1) (2016) 014102.
[25] E.M. Stoudenmire, S.R. White, Sliced basis density matrix renormalization group for electronic structure, Phys. Rev. Lett. 119 (4) (2017) 046401.
[26] G. Knizia, G.K.-L. Chan, Density matrix embedding: a simple alternative to dynamical mean-field theory, Phys. Rev. Lett. 109 (18) (2012) 186404.
[27] G. Knizia, G.K.-L. Chan, Density matrix embedding: a strong-coupling quantum embedding theory, J. Chem. Theory Comput. 9 (3) (2013) 1428–1432.
[28] S. Wouters, C.A. Jiménez-Hoyos, Q. Sun, G.K.-L. Chan, A practical guide to density matrix embedding theory in quantum chemistry, J. Chem. Theory Comput. 12 (6) (2016) 2706–2719.
[29] J. Grotendorst, Modern Methods and Algorithms of Quantum Chemistry, John von Neumann Institute for Computing, 2000.
[30] A. Szabo, N.S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory, Courier Corporation, 2012.
[31] M. Motta, D.M. Ceperley, G.K.-L. Chan, J.A. Gomez, E. Gull, S. Guo, C.A. Jiménez-Hoyos, T.N. Lan, J. Li, F. Ma, et al., Towards the solution of the many-electron problem in real materials: equation of state of the hydrogen chain with state-of-the-art many-body methods, Phys. Rev. X 7 (3) (2017) 031059.
[32] J.C. Slater, Note on Hartree's method, Phys. Rev. 35 (2) (1930) 210.
[33] R. Jastrow, Many-body problem with strong forces, Phys. Rev. 98 (5) (1955) 1479.
[34] C. Umrigar, K. Wilson, J. Wilkins, Optimized trial wave functions for quantum Monte Carlo calculations, Phys. Rev. Lett. 60 (17) (1988) 1719.
[35] C. Umrigar, J. Toulouse, C. Filippi, S. Sorella, R.G. Hennig, Alleviation of the fermion-sign problem by optimization of many-body wave functions, Phys. Rev. Lett. 98 (11) (2007) 110201.
[36] M. Casula, S. Sorella, Geminal wave functions with Jastrow correlation: a first application to atoms, J. Chem. Phys. 119 (13) (2003) 6500–6511.
[37] H.J. Changlani, J.M. Kinder, C.J. Umrigar, G.K.-L. Chan, Approximating strongly correlated wave functions with correlator product states, Phys. Rev. B 80 (24) (2009) 245116.
[38] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436.
[39] J. Han, A. Jentzen, W. E, Solving high-dimensional partial differential equations using deep learning, Proc. Natl. Acad. Sci. 115 (34) (2018) 8505–8510.
[40] W. E, J. Han, A. Jentzen, Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations, Commun. Math. Stat. 5 (4) (2017) 349–380.
[41] J. Berg, K. Nyström, A unified deep artificial neural network approach to partial differential equations in complex geometries, arXiv preprint arXiv: 1711.06464.
[42] Y. Khoo, J. Lu, L. Ying, Solving parametric PDE problems with artificial neural networks, arXiv preprint arXiv:1707.03351.
[43] H. Saito, Method to solve quantum few-body problems with artificial neural networks, J. Phys. Soc. Jpn. 87 (7) (2018) 074002.
[44] G. Carleo, M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science 355 (6325) (2017) 602–606.
[45] X. Gao, L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. 8 (1) (2017) 662.
[46] H. Saito, Solving the Bose–Hubbard model with machine learning, J. Phys. Soc. Jpn. 86 (9) (2017) 093001.
[47] Z. Cai, J. Liu, Approximating quantum many-body wave functions using artificial neural networks, Phys. Rev. B 97 (3) (2018) 035116.
[48] M. Born, R. Oppenheimer, Zur Quantentheorie der Molekeln, Ann. Phys. 389 (9) (1927) 1–31.
[49] R.B. Laughlin, Anomalous quantum Hall effect: an incompressible quantum fluid with fractionally charged excitations, Phys. Rev. Lett. 50 (18) (1983) 1395.
[50] I. Glasser, N. Pancotti, M. August, I.D. Rodriguez, J.I. Cirac, Neural-network quantum states, string-bond states, and chiral topological states, Phys. Rev. X 8 (1) (2018) 011006.
[51] S.R. Clark, Unifying neural-network quantum states and correlator product states via tensor networks, J. Phys. A, Math. Theor. 51 (13) (2018) 135301.
[52] J. Behler, M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, Phys. Rev. Lett. 98 (14) (2007) 146401.
[53] J. Han, L. Zhang, R. Car, W. E, Deep potential: a general representation of a many-body potential energy surface, Commun. Comput. Phys. 23 (3) (2018) 629–639.
[54] L. Zhang, J. Han, H. Wang, R. Car, W. E, Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics, Phys. Rev. Lett. 120 (2018) 143001.
[55] L. Zhang, J. Han, H. Wang, W.A. Saidi, R. Car, W. E, End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems, in: Advances of the Neural Information Processing Systems (NIPS), 2018.
[56] K. Schütt, P.-J. Kindermans, H.E.S. Felix, S. Chmiela, A. Tkatchenko, K.-R. Müller, Schnet: a continuous-filter convolutional neural network for modeling quantum interactions, in: Advances in Neural Information Processing Systems (NIPS), 2017, pp. 992–1002.
[57] L. Zhang, H. Wang, W. E, Reinforced dynamics for enhanced sampling in large atomic and molecular systems, J. Chem. Phys. 148 (12) (2018) 124113.
[58] L. Zhang, J. Han, H. Wang, R. Car, W. E, DeePCG: constructing coarse-grained models via deep neural networks, J. Chem. Phys. 149 (3) (2018) 034101.

[59] E. Schneider, L. Dai, R.Q. Topper, C. Drechsel-Grau, M.E. Tuckerman, Stochastic neural network approach for learning high-dimensional free energy surfaces, Phys. Rev. Lett. 119 (15) (2017) 150601.
[60] T. Kato, On the eigenfunctions of many-particle systems in quantum mechanics, Commun. Pure Appl. Math. 10 (2) (1957) 151–177.
[61] R.D. Johnson III, CCCBDB computational chemistry comparison and benchmark database, NIST Standard Reference Database Number 101.
[62] K. Frankowski, C.L. Pekeris, Logarithmic terms in the wave functions of the ground state of two-electron atoms, Phys. Rev. 146 (1966) 46–49.
[63] Q. Sun, T.C. Berkelbach, N.S. Blunt, G.H. Booth, S. Guo, Z. Li, J. Liu, J.D. McClain, E.R. Sayfutyarova, S. Sharma, et al., PySCF: the Python-based simulations of chemistry framework, Wiley Interdiscip. Rev. Comput. Mol. Sci. 8 (1) (2018) e1340.
[64] D.E. Woon, T.H. Dunning Jr., Gaussian basis sets for use in correlated molecular calculations. V. Core-valence basis sets for boron through neon, J. Chem. Phys. 103 (11) (1995) 4572–4585.
[65] W.J. Hehre, R.F. Stewart, J.A. Pople, Self-consistent molecular-orbital methods. I. Use of Gaussian expansions of Slater-type atomic orbitals, J. Chem. Phys. 51 (6) (1969) 2657–2664.
[66] R. Lee, G. Conduit, N. Nemec, P.L. Ríos, N. Drummond, Strategies for improving the efficiency of quantum Monte Carlo calculations, Phys. Rev. E 83 (6) (2011) 066706.
[67] M. Dewing, Improved efficiency with variational Monte Carlo using two level sampling, J. Chem. Phys. 113 (13) (2000) 5123–5125.
[68] D. Kingma, J. Ba, Adam: a method for stochastic optimization, in: Proceedings of the International Conference on Learning Representations (ICLR), 2015.