Journal of Machine Learning **ISSN**: 2790-2048(e), 2790-203X(p)

Ab-initio Study of Interacting Fermions at Finite Temperature with Neural Canonical Transformation

Hao Xie * [1,2], Linfeng Zhang † [3,4], and Lei Wang ‡ [1,5]

[1]Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China.
[2]University of Chinese Academy of Sciences, Beijing 100049, China.
[3]AI for Science Institute, Beijing 100080, China
[4]DP Technology, Beijing 100080, China
[5]Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China.

Abstract. We present a variational density matrix approach to the thermal properties of interacting fermions in the continuum. The variational density matrix is parametrized by a permutation equivariant many-body unitary transformation together with a discrete probabilistic model. The unitary transformation is implemented as a quantum counterpart of neural canonical transformation, which incorporates correlation effects via a flow of fermion coordinates. As the first application, we study electrons in a two-dimensional quantum dot with an interaction-induced crossover from Fermi liquid to Wigner molecule. The present approach provides accurate results in the low-temperature regime, where conventional quantum Monte Carlo methods face severe difficulties due to the fermion sign problem. The approach is general and flexible for further extensions, thus holds the promise to deliver new physical results on strongly correlated fermions in the context of ultracold quantum gases, condensed matter, and warm dense matter physics.

Keywords:
Interacting fermions,
Thermodynamics,
Variational free energy,
Normalizing flows.

Article Info.:
Volume: 1
Number: 1
Pages: 38- 59
Date: March/2022
doi.org/10.4208/jml.220113

Article History:
Received: 22/11/2021
Accepted: 19/3/2022

Communicated by:
Weinan E

1 Introduction

Consider an interacting quantum system of N fermions in a d-dimensional continuous space with the generic Hamiltonian

$$H = -\frac{1}{2}\nabla^2 + V(\boldsymbol{x}), \tag{1.1}$$

where $\nabla^2 = \sum_{i=1}^N \nabla_i^2$ and $V(\boldsymbol{x}) = \sum_{i=1}^N v^{(1)}(\boldsymbol{r}_i) + \sum_{i<j}^N v^{(2)}(\boldsymbol{r}_i - \boldsymbol{r}_j)$ consists of one- and two-body potentials. We use the short-hand notation $\boldsymbol{x} \equiv (\boldsymbol{r}_1, \boldsymbol{r}_2, \cdots, \boldsymbol{r}_N) \in \mathbb{R}^{dN}$ to collectively denote all fermion coordinates. We also assume appropriate natural units so that constants like fermion mass and the Planck constant can be omitted. We would like to study thermodynamic properties of the system encoded in the partition function $Z = \mathrm{Tr} e^{-\beta H}$ at

*xiehao18@iphy.ac.cn.
†linfeng.zhang.zlf@gmail.com.
‡Corresponding author. wanglei@iphy.ac.cn.

inverse temperature $\beta = 1/k_B T$, which are relevant to a broad range of problems including ultracold Fermi gases [1], condensed matter [2], and warm dense matter [3].

Unfortunately, accurate ab-initio study of Eq. (1.1) at finite temperature is generally difficult. As a typical workhorse, quantum Monte Carlo methods suffer from the notorious fermion sign problem at low temperature [4–7]. There are extensions of ground-state quantum chemistry methods to finite temperature, e.g. [8]. On the other hand, a fundamental principle to solve quantum systems at finite temperature is to minimize the variational free energy

$$F = \frac{1}{\beta}\mathrm{Tr}(\rho \ln \rho) + \mathrm{Tr}(\rho H) \tag{1.2}$$

with respect to a variational density matrix ρ. The two terms in Eq. (1.2) correspond to the entropy and energy of the system, respectively. It can be shown that $F \geqslant -\frac{1}{\beta}\ln Z$ is a variational upper bound of the true free energy, where the equality holds only when ρ coincides with the exact density matrix $\frac{1}{Z}e^{-\beta H}$ of the system [9].

There are a number of physical constraints on the variational density matrix ρ. Besides basic properties like Hermitian ($\rho^\dagger = \rho$), positive definiteness ($\rho \succ 0$) and normalization ($\mathrm{Tr}\rho = 1$), it should also be antisymmetric $\langle \mathcal{P}\boldsymbol{x}|\rho|\boldsymbol{x}'\rangle = (-1)^{\mathcal{P}}\langle \boldsymbol{x}|\rho|\boldsymbol{x}'\rangle$ with respect to permutations $\mathcal{P}\boldsymbol{x} \equiv (\boldsymbol{r}_{\mathcal{P}1}, \boldsymbol{r}_{\mathcal{P}2}, \cdots, \boldsymbol{r}_{\mathcal{P}N})$ of the fermion coordinates. The challenge is then to devise a tractable computational scheme to perform optimization of ρ within such a constrained space. In practice, the entropy term in Eq. (1.2) often turns out to be difficult to compute [10,11]. As a result, most of the previous variational density matrix studies resort to an alternative imaginary-time evolution approach [10, 12–14], which is more appropriate at high temperatures.

In the low-temperature regime, the variational density matrix can be reasonably represented by a truncated [1] set of low-energy many-body basis states $|\Psi_n\rangle$:

$$\rho = \sum_n \mu_n(\boldsymbol{\phi})|\Psi_n(\boldsymbol{\theta})\rangle\langle\Psi_n(\boldsymbol{\theta})|, \tag{1.3}$$

where $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are variational parameters. A discrete probabilistic model μ_n, which satisfies $0 < \mu_n < 1$ and $\sum_n \mu_n = 1$, is used to parametrize the Boltzmann distribution of the basis states $|\Psi_n\rangle$. On the other hand, we choose to construct $|\Psi_n\rangle$ by applying a unitary transformation to a set of reference basis states $|\Phi_n\rangle$, e.g., the non-interacting Slater determinants: $|\Psi_n(\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\Phi_n\rangle$. To make such an ansatz powerful enough, the unitary transformation should have *many-body* nature, so that particle correlations can be effectively incorporated into the reference state. In addition, it should also preserve permutation antisymmetry of the reference wavefunction, which we will refer to as the *equivariance property*. Overall, the modeling of the variational density matrix in the present approach is illustrated in Fig. 1.1.

Parametrizing and optimizing a rich family of equivariant many-body unitary transformations U turn out to be a fairly nontrivial task. In this paper, we present an elegant solution to this problem by constructing U as unitary representation of the canonical

[1] Truncation is necessary for systems in the continuum with infinite-dimensional Hilbert space. Given N fermions and M available single-particle orbitals, the summation in Eq. (1.3) involves $\binom{M}{N}$ terms, which is exponentially large. Nevertheless, such summation appearing in relevant physical quantities can be estimated via Monta Carlo sampling and will not cause big troubles in practice. See also discussions in Section 4.

Figure 1.1: Architecture of the variational density matrix representation (1.3) of the present approach. A discrete probabilistic model μ_n parametrizes the Boltzmann distribution of a many-body basis set $|\Psi_n\rangle$. We construct $|\Psi_n\rangle$ by applying a parametrized unitary transformation to a reference basis $|\Phi_n\rangle$. The unitary transformation corresponds to a permutation equivariant many-body coordinate transformation implemented as a flow of fermion coordinates.

transformation of phase space variables in classical mechanics, extending the previous work on neural canonical transformations [15] from classical to quantum domain. The resulting approach naturally generalizes the ground-state variational Monte Carlo (VMC) method [16, 17] to finite temperature and is not hindered by the fermion sign problem. Moreover, based on Born's probabilistic interpretation of wavefunctions, the equivariant unitary transformation turns out to be intimately related to equivariant *normalizing flow* [18, 19], an important class of generative model developed well within the deep-learning community. This way, one can leverage the latest technical advances in probabilistic modeling to efficiently tackle the thermodynamics of strongly correlated fermions in a fully ab-initio way.

It is worth mentioning a variety of related works to put the present contribution into a broader perspective. First, there have been various wavefunction ansatzes for *ground-state* VMC calculation of fermions, from the traditional Slater-Jastrow [20], backflow [21–24] to more recent attempts based on neural networks [25–33]. However, unitary transformations are not considered in these ansatzes, since only a single wavefunction, instead of a whole basis, is needed in this situation. Second, there have been quantum algorithms for thermal properties of model Hamiltonians [34–36], which rely on quantum circuits to construct the unitary transformation. However, they still demand advances in quantum technologies to be practically useful. Third, variational free energy studies of statistical mechanics and field theory problems [15, 37–42] can be regarded as the classical counterparts of the present approach. Last but not least, the so-called quantum flow approach [43] also performs a learnable unitary transformation to a single-particle basis. In the many-particle settings considered here, one has to additionally deal with the permutation antisymmetry by imposing equivariance property into the coordinate transformation carried out by the normalizing flows [44–46]. In this way, normalizing flows have also precisely addressed the open problem envisioned in [47]: "The full use of the (coordinate) transform to compute from first principles requires adequate approximation to the Jacobian and the inverse transformation."

2 Theory

The method of constructing unitary transformations in this work is based on the fact that one can establish a one-to-one correspondence between the group of unitary transformations in quantum mechanics and the group of canonical transformations of phase space variables $(\boldsymbol{x}, \boldsymbol{p})$ in classical mechanics [48–51]. One can develop some basic understanding of the idea by inspecting the infinitesimal structure of these two groups. In classical mechanics, one can use an arbitrary generating function $G(\boldsymbol{x}, \boldsymbol{p})$ to define a continuous family of canonical transformations via the symplectic evolution $\frac{d\boldsymbol{x}}{d\lambda} = \frac{\partial G}{\partial \boldsymbol{p}}$, $\frac{d\boldsymbol{p}}{d\lambda} = -\frac{\partial G}{\partial \boldsymbol{x}}$, where λ denotes a continuous parameter. When the canonical transformation is quantized, the generating function is converted to a Hermitian operator $\hat{G}$ [2], and the corresponding unitary transformation then takes the form $U_\lambda = e^{-i\hat{G}\lambda}$. See Appendix A for more details.

An important class of canonical transformation is the so-called *point transformations*, in which the new generalized coordinates depend solely on the old coordinates, not on the old momenta. The generating function of point transformation reads $G = \boldsymbol{u}(\boldsymbol{x}) \cdot \boldsymbol{p}$, where $\boldsymbol{u} : (\boldsymbol{r}_1, \cdots, \boldsymbol{r}_N) \mapsto (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_N)$ is a function in the dN-dimensional coordinate space. The equation of motion followed by the transformed coordinates then takes the form

$$\frac{d\boldsymbol{x}}{d\lambda} = \boldsymbol{u}(\boldsymbol{x}). \tag{2.1}$$

$\boldsymbol{u}$ can be intuitively viewed as a vector field that guides all the particles to continuously flow in the coordinate space as the parameter λ increases [3]. To see this, one can consider the examples of spatial translation and rotation, which can be generated by the total momentum $G = \boldsymbol{e} \cdot \sum_{i=1}^{N} \boldsymbol{p}_i$ or angular momentum $G = \boldsymbol{n} \cdot \sum_{i=1}^{N} (\boldsymbol{r}_i \times \boldsymbol{p}_i)$ along certain directions. The corresponding vector fields are $\boldsymbol{u}_i = \boldsymbol{e}$ and $\boldsymbol{u}_i = \boldsymbol{n} \times \boldsymbol{r}_i$ respectively, as illustrated in the left and center panel of Fig. 2.1. For systems with such spatial symmetries, these transformations would leave the Hamiltonian unaltered. We also note that the vector fields associated with these two examples are "separable", i.e., the vector field experienced by one particle is independent of the positions of any other particles, thus completely ignores the interactions among them. For practical variational calculations, we need to seek for vector fields $\boldsymbol{u}$ that can effectively introduce particle correlations, as illustrated in the right panel of Fig. 2.1.

To study the unitary transformation induced by a general point transformation, we simply employ an anticommutator to make the quantized generator Hermitian [52]: $\hat{G} = \frac{1}{2}\{\boldsymbol{u}(\hat{\boldsymbol{x}}), \hat{\boldsymbol{p}}\}$. Given a set of basis states $\Phi_n(\boldsymbol{x})$, the transformed basis wavefunction $\Psi_n(\boldsymbol{x}, \lambda) \equiv (U_\lambda \Phi_n)(\boldsymbol{x})$ then reads

$$\Psi_n(\boldsymbol{x}, \lambda) = \langle \boldsymbol{x} | e^{-\frac{i}{2}\{\boldsymbol{u}(\hat{\boldsymbol{x}}), \hat{\boldsymbol{p}}\}\lambda} | \Phi_n \rangle. \tag{2.2}$$

To obtain a physical interpretation, it is instructive to differentiate both sides of Eq. (2.2)

[2] Note that we have put a hat $\hat{}$ on operators somewhere in this section to avoid possible confusions.

[3] Most generally, the vector field $\boldsymbol{u}(\boldsymbol{x}, \lambda)$ can also depend on the continuous parameter λ. The resulting symplectic evolution followed by the particles is then "time-inhomogeneous".

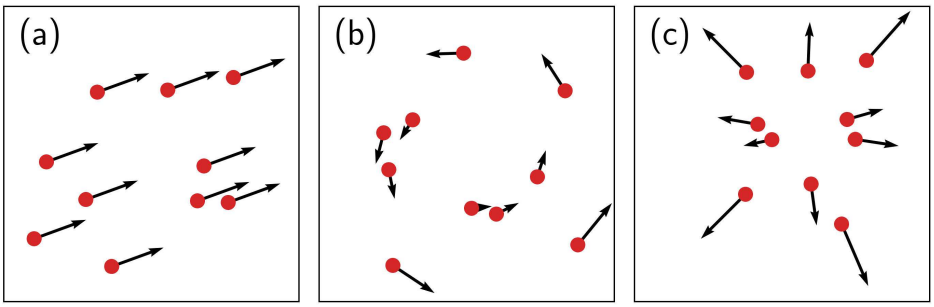

Figure 2.1: Schematic plot of three different vector fields $\boldsymbol{u}$ experienced by the particles. (a) Spatial translation. (b) Spatial rotation. (c) A "many-body" vector field involving interdependence among the particles. Evolution under such a vector field will introduce correlation effects.

with respect to λ to get

$$\frac{\partial}{\partial\lambda}|\Psi_n(\boldsymbol{x},\lambda)|^2 + \nabla\cdot\Big(|\Psi_n(\boldsymbol{x},\lambda)|^2\boldsymbol{u}(\boldsymbol{x})\Big) = 0, \tag{2.3}$$

which has the form of a continuity equation of probability density. See Appendix A for the derivation details. Pictorially, starting from a family of orthonormal reference states such as Slater determinants, the probability mass of each many-body wavefunction will undergo a continuous evolution guided by the many-body vector field $\boldsymbol{u}$. During this process, the particles constantly repel or attract each other and correlation effects are gradually cumulated. Crucially, these states remain orthonormal thanks to the unitary nature of the transformation. We thus obtain a correlated many-body basis in the end of the evolution, which can be used to build up the variational density matrix ansatz Eq. (1.3). Note the particles follow deterministic advection in Eq. (2.3) rather than random diffusion. In practice, one should integrate the differential equation for a finite amount of time instead of seeking for steady state solutions as in diffusion-based approaches [53–55].

Transforming probability density continuously in the coordinate space is precisely the idea of continuous normalizing flow [39, 54, 56, 57]. Specifically, the probability $p_n(\boldsymbol{z}) \equiv |\Phi_n(\boldsymbol{z})|^2$ associated with the reference state is known as the *base* distribution, while the *model* distribution $q_n(\boldsymbol{x}) \equiv |\Psi_n(\boldsymbol{x})|^2$ is obtained from the base by applying a learnable diffeomorphism $\boldsymbol{f} : \boldsymbol{z} \mapsto \boldsymbol{x}$ in the dN-dimensional coordinate space. Note we have omitted λ in the notation to avoid cluttering. In the present settings, one builds the diffeomorphism $\boldsymbol{f}$ by integrating the ordinary differential equation (ODE) (2.1). By making use of the change-of-variable formula, the model probability $q_n(\boldsymbol{x})$ can be written as

$$q_n(\boldsymbol{x}) = p_n(\boldsymbol{f}^{-1}(\boldsymbol{x}))\left|\det\left(\frac{\partial \boldsymbol{f}^{-1}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)\right|. \tag{2.4}$$

Taking the square root of both sides yields a more explicit expression for the basis wave-

function [4]:

$$\Psi_n(\boldsymbol{x}) = \Phi_n(\boldsymbol{f}^{-1}(\boldsymbol{x})) \left| \det \left(\frac{\partial \boldsymbol{f}^{-1}(\boldsymbol{x})}{\partial \boldsymbol{x}} \right) \right|^{\frac{1}{2}}. \tag{2.5}$$

Albeit not so evident at the first sight, Eqs. (2.5) and (2.2) are completely equivalent, which can be rigorously proved by more formally establishing the unitary representation of point transformations; see Appendix A. In practice, the diffeomorphism $\boldsymbol{f}$ can be constructed by composing a sequence of point transformations, which is similar to the iterative backflow approach in ground-state variational calculations [58, 59]. However, an important difference of Eq. (2.5) from the ground-state backflow wavefunction ansatzes is the presence of a Jacobian determinant factor. This factor is crucial to guarantee orthonormality of the basis states, which is an essential ingredient for the present finite-temperature approach.

Finally, as a many-fermion wavefunction, $\Psi_n(\boldsymbol{x})$ should satisfy the permutation antisymmetry property. Since this property holds already for the base $\Phi_n(\boldsymbol{z})$, the only requirement is the unitary transformation appearing in Eq. (2.2) being *permutation equivariant*: that is, it should commute with the particle permutation operator. This can be achieved simply by requiring the many-body vector field $\boldsymbol{u}$ to be equivariant too, which means that the permutation of particle positions will result in the same permutation of the vector fields they experience:

$$\boldsymbol{u}(\mathcal{P}\boldsymbol{x}) = \mathcal{P}\boldsymbol{u}(\boldsymbol{x}). \tag{2.6}$$

Intuitively, the indistinguishability of the particles is maintained throughout the continuous flow in the coordinate space, since one cannot label them by using the vector fields they experience at any time. The probability density $q_n(\boldsymbol{x}) = |\Psi_n(\boldsymbol{x})|^2$ associated with the transformed wavefunction, on the other hand, is *invariant* under particle permutations [44–46]. Another notable feature is that $q_n(\boldsymbol{x})$ inherits nodal lines from the fermionic reference state $p_n(\boldsymbol{z}) = |\Phi_n(\boldsymbol{z})|^2$, up to a deformation induced by the flow transformation.

To parametrize the permutation equivariant vector field $\boldsymbol{u}$, one can leverage many recent advances in natural language processing [60], molecular simulation [44, 61, 62], and point set modeling [45, 46]. Moreover, permutation equivariant functions have also been used in various ground-state VMC calculations [24, 27, 28]. Consequently, one can naturally port these efforts into the present framework almost without any modifications: just use the permutation equivariant layer as the vector field $\boldsymbol{u}$ to drive the flow.

3 Implementation

Substitution of the density matrix ansatz (1.3) into Eq. (1.2) yields the following estimator of the variational free energy:

$$F = \mathop{\mathbb{E}}_{n\sim\mu_n} \left[\frac{1}{\beta} \ln \mu_n + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim q_n(\boldsymbol{x})} \left[E_n^{\mathrm{loc}}(\boldsymbol{x}) \right] \right]. \tag{3.1}$$

[4] Such a sloppy "derivation" is fairly intuitive, yet not satisfying enough for mathematical rigor. In particular, Eq. (2.5) implicitly assumes the wavefunction ansatz $\Psi_n(\boldsymbol{x})$ has exactly the same *phase* as $\Phi_n(\boldsymbol{z})$, up to a spatial deformation brought by the transformation $\boldsymbol{z} \to \boldsymbol{x}$. Fortunately, this is indeed the case. One easy way to see this is from Eq. (2.2): note $\hat{\boldsymbol{p}} = -i\nabla$, thus the exponential operator acting on the base wavefunction Φ_n is actually *real valued*.

DOI https://doi.org/10.4208/jml.220113 | Generated on 2025-04-12 03:46:28

Notice the entropy term depends solely on the state occupation probability μ_n and can be easily computed, which is a direct consequence of orthonormality of the basis states (2.5). The second term consists of the local energy associated with each basis state:

$$\begin{aligned} E_n^{\mathrm{loc}}(\boldsymbol{x}) &\equiv \frac{H\Psi_n(\boldsymbol{x})}{\Psi_n(\boldsymbol{x})} \\ &= -\frac{1}{4}\nabla^2 \ln q_n(\boldsymbol{x}) - \frac{1}{8}\left(\nabla \ln q_n(\boldsymbol{x})\right)^2 + V(\boldsymbol{x}). \end{aligned} \tag{3.2}$$

In Eq. (3.1) the two-fold expectations correspond to classical thermal average of the Boltzmann distribution and quantum expectation according to the Born rule of wavefunction amplitudes. In the limit $\beta \to \infty$, only the energy term survives and one naturally restores the ground-state VMC method.

The gradients of Eq. (3.1) with respect to the parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$, which appear in the classical and quantum distributions μ_n and $q_n(\boldsymbol{x})$ respectively, have the following forms:

$$\nabla_{\boldsymbol{\phi}} F = \mathop{\mathbb{E}}_{n\sim\mu_n}\left[\left(\frac{1}{\beta}\ln\mu_n + \mathop{\mathbb{E}}_{\boldsymbol{x}\sim q_n(\boldsymbol{x})}\left[E_n^{\mathrm{loc}}(\boldsymbol{x})\right]\right)\nabla_{\boldsymbol{\phi}}\ln\mu_n\right], \tag{3.3a}$$

$$\nabla_{\boldsymbol{\theta}} F = \mathop{\mathbb{E}}_{n\sim\mu_n}\mathop{\mathbb{E}}_{\boldsymbol{x}\sim q_n(\boldsymbol{x})}\left[E_n^{\mathrm{loc}}(\boldsymbol{x})\nabla_{\boldsymbol{\theta}}\ln q_n(\boldsymbol{x})\right]. \tag{3.3b}$$

For both estimators we employ the control variate method [36, 38, 63] to further reduce their variances. An important observation is that only the non-negative probability density $q_n(\boldsymbol{x}) = |\Psi_n(\boldsymbol{x})|^2$ associated with the wavefunction is involved in the calculation. This is a satisfying feature of working directly in the continuum rather than on a finite basis set or lattice [29, 64]: one can deal with the quantum many-body problem completely within the framework of probabilistic modeling. Nevertheless, the sign structure of the fermion wavefunction $\Psi_n(\boldsymbol{x})$ is still important and relevant for the calculation of off-diagonal physical observables such as correlation function and momentum distribution.

In practice, Eqs. (3.1) and (3.3) are estimated by sampling a batch of pairs $(n, \boldsymbol{x})$ from the joint distribution $\mu_n q_n(\boldsymbol{x})$ following the ancestral sampling strategy. In particular, to sample coordinates $\boldsymbol{x}$, one can start from samples $\boldsymbol{z}$ from the prior distribution $p_n(\boldsymbol{z})$ (e.g., via Markov chain Monte Carlo) and evolve them according to the ODE (2.1). The log-likelihood $\ln q_n(\boldsymbol{x})$ appearing in the local energy (3.2) and gradient estimators (3.3) is evaluated by integrating Eq. (2.1) jointly with the following ODE [39, 57]:

$$\frac{d\ln q_n}{d\lambda} = -\nabla\cdot\boldsymbol{u}(\boldsymbol{x}). \tag{3.4}$$

To understand this, one can rewrite the continuity equation (2.3) in the form

$$\left(\frac{\partial}{\partial\lambda} + \boldsymbol{u}(\boldsymbol{x})\cdot\nabla\right)\ln q_n(\boldsymbol{x},\lambda) = -\nabla\cdot\boldsymbol{u}(\boldsymbol{x}) \tag{3.5}$$

and note that $\frac{d}{d\lambda} \equiv \frac{\partial}{\partial\lambda} + \boldsymbol{u}(\boldsymbol{x})\cdot\nabla$ is the material derivative associated with the sample $\boldsymbol{x}$. Furthermore, the gradient and Laplacian operations appearing in (3.2) and (3.3) can

be accurately and efficiently computed by differentiating through the ODE integration using automatic differentiation [65], where the adjoint method with constant memory cost turns out to be useful [57]. Our code implementation based on PyTorch [66] is publicly available [67].

4 Application: Electrons in two-dimensional quantum dot

We demonstrate the capability of the present approach by studying electrons in a two-dimensional quantum dot. The one- and two-body potentials take the form of a harmonic trap and repulsive Coulomb interaction, respectively:

$$v^{(1)}(\boldsymbol{r}) = \frac{1}{2}\boldsymbol{r}^2,\ v^{(2)}(\boldsymbol{r}-\boldsymbol{r}') = \frac{\kappa}{|\boldsymbol{r}-\boldsymbol{r}'|}, \tag{4.1}$$

where $\kappa > 0$ is the interaction strength. Despite its simplicity, this model shows rich phenomena due to the interplay of interaction and temperature effects. In particular, as κ increases, the Fermi liquid picture based on the concept of quasiparticles would eventually break down. The resulting phase is usually characterized as a *Wigner molecule* [2], where the kinetic motion of electrons is largely frozen, and the spatial density distribution would typically exhibit a shell structure. There have been a large number of numerical studies focusing on its ground-state [70–72] and finite-temperature properties [73–76]. However, there have been no reliable method that work for the entire interaction range at low temperature. Thus, such a system offers an ideal playground for the present method.

We consider the spin polarized case. The base wavefunctions $\Phi_n(\boldsymbol{z})$ are chosen to be Slater determinants of single-electron orbitals obtained simply by eliminating the two-body term $v^{(2)}$. Such Slater determinants constitute an exponentially large set of basis of the many-body Hilbert space. Focusing on low-temperature properties of the system, we carry out a truncation of the basis by including only those within an energy cutoff E_{cut} relative to the non-interacting ground state. In the considered parameter region, we found that $E_{\mathrm{cut}} \leqslant 4$ is sufficient to capture most of the finite-temperature effects. Since the corresponding number of basis states is no more than 2000, we choose to adopt a simple parametrization of the state probabilities $\mu_n(\boldsymbol{\phi}) = \frac{e^{\phi_n}}{\sum_m e^{\phi_m}}$ based on the softmax function. Nevertheless, we note that this is not a limiting factor of the present approach because one can capture exponentially large number of basis states by utilizing more sophisticated discrete probabilistic models [38, 77–79].

The wavefunction ansatz $\Psi_n(\boldsymbol{x})$ is generated from the base $\Phi_n(\boldsymbol{z})$ by the continuous flow guided by a many-body vector field $\boldsymbol{u}$. We take $\boldsymbol{u}$ to be of the backflow form [21–24] for simplicity and clear physical interpretation:

$$\boldsymbol{u}_i = \xi(|\boldsymbol{r}_i|)\boldsymbol{r}_i + \sum_{j\neq i}^{N} \eta(|\boldsymbol{r}_i - \boldsymbol{r}_j|)(\boldsymbol{r}_i - \boldsymbol{r}_j). \tag{4.2}$$

The many-body nature and permutation equivariance of this vector field can be easily confirmed by inspection. The scalar functions ξ and η can be referred to as the one- and

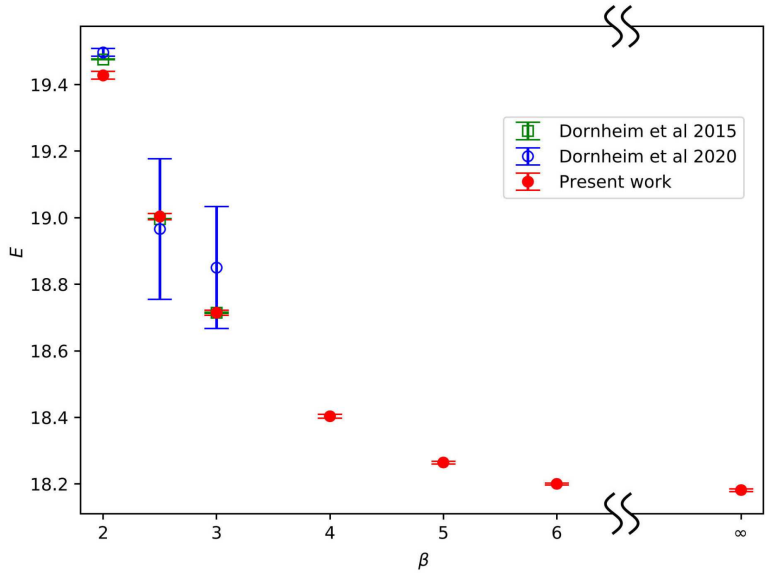

Figure 4.1: Energy E versus inverse temperature β for 6 spin-polarized electrons in a two-dimensional quantum dot with $\kappa = 0.5$. The green and blue points are benchmark data from two different variants of PIMC [68, 69], while the red points are results of the present approach, including the zero-temperature limit.

two-body "backflow potential", which capture the "mean field" and electron correlation effects, respectively. Note for a given distance r, $\eta(r) > 0$ stands for a repulsive interaction between two electrons, and vice versa; similarly for $\xi(r)$. We parametrize the potentials by two independent neural networks with single hidden layer. Initially, the backflow potentials are set to zero and μ_n to Boltzmann distribution of the non-interacting base states $\Phi_n(z)$. The optimization is performed on a batch of 8000 samples using the Adam stochastic gradient descent algorithm [80] for 3000 iteration steps.

As the first benchmark, Fig. 4.1 shows the temperature dependence of the energy for a system of $N = 6$ spin-polarized electrons with $\kappa = 0.5$. For such a weak interaction, the standard path integral Monte Carlo (PIMC) method is severely hindered by the fermion sign problem, since the electrons are largely delocalized and subject to exchange effects. Consequently, it provides reliable results only at relatively high temperatures $\beta \lesssim 1.5$ [6]. Variants of PIMC with alleviated fermion sign problem can access slightly lower temperatures [68, 69], where our results agree nicely with the benchmark data as shown in the figure. The slight discrepancy at $\beta = 2$ is likely due to insufficiently large E_{cut} in our calculation. On the other hand, the present approach can easily reach even lower temperatures, including the zero-temperature limit $\beta \to \infty$. We note that alternative Monte Carlo methods based on expansions in the Fock space [81,82] can work more favorably for such weak interactions, but will again suffer from the fermion sign problem in the strong coupling region.

Overall, the present approach serves as a valuable complement of conventional quantum Monte Carlo methods for studying the thermal properties of fermion sys-

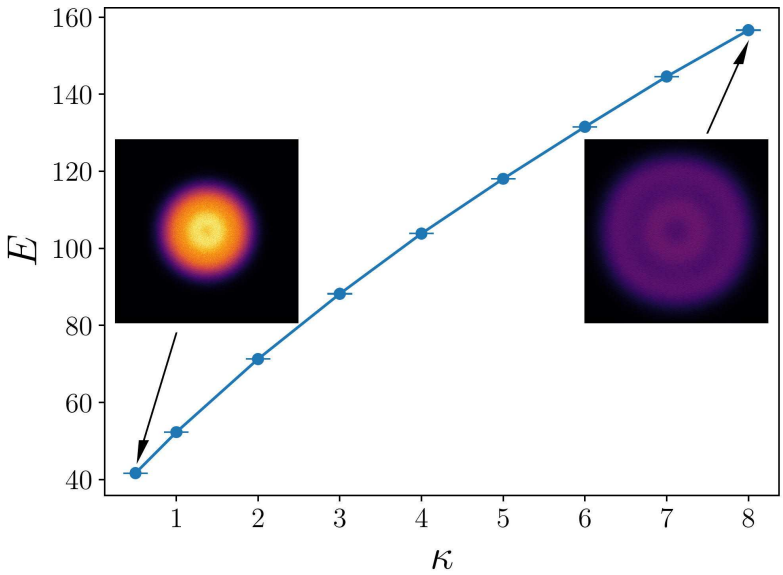

Figure 4.2: Energy E versus interaction strength κ for 10 spin-polarized electrons in a two-dimensional quantum dot with $\beta = 6$. The inset shows the electron density profiles for $\kappa = 0.5$ and 8.

tems, especially for low temperature, large particle number and intermediate interaction strength [83]. To demonstrate this, we perform systematic calculations of $N = 10$ spin-polarized electrons at $\beta = 6$ for a wide range of κ from 0.5 to 8. For more benchmark data, see Appendix B. Fig. 4.2 shows the energy dependence on κ, together with electron density profiles at the end points $\kappa = 0.5$ and 8. Notice the density centers around the origin of the trap in the weak coupling regime. On the other hand, stronger repulsive interaction smears out the electron cloud and induces a shell structure, indicating the emergence of the Wigner molecule phase. The observed spatial configuration consisting of two shells for the present parameter settings also agrees with the analysis in the classical limit $\kappa \to \infty$ [2], where quantum fluctuations arising from the kinetic term in Eq. (1.1) are ignored. Reaching this result in the strong interaction regime where the density profile is qualitatively different from the weak coupling case is a stringent test to the present method.

To obtain the density profiles as shown in Fig. 4.2, one starts from the density of non-interacting reference state consisting of a large number of electron coordinate samples $\boldsymbol{z}$, then evolves them according to the continuous flow specified by the ODE (2.1) towards the final spatial distribution of $\boldsymbol{x}$, as described previously in Section 3. The initial and final values of the continuous parameter λ are conventionally chosen to be 0 and 1, respectively, which are treated as fixed hyperparameters of the model. The many-body vector field $\boldsymbol{u}$ governing such an evolution process is determined by the backflow potentials ξ and η, which are shown in Fig. 4.3. Notice the interactions arising from one- and two-body backflow potentials are both repulsive, which can be viewed as the manifestation of electron repulsion at the level of mean field and two-body correlations, respectively. The overall evolution of the electrons is, nevertheless, jointly determined by the two po-

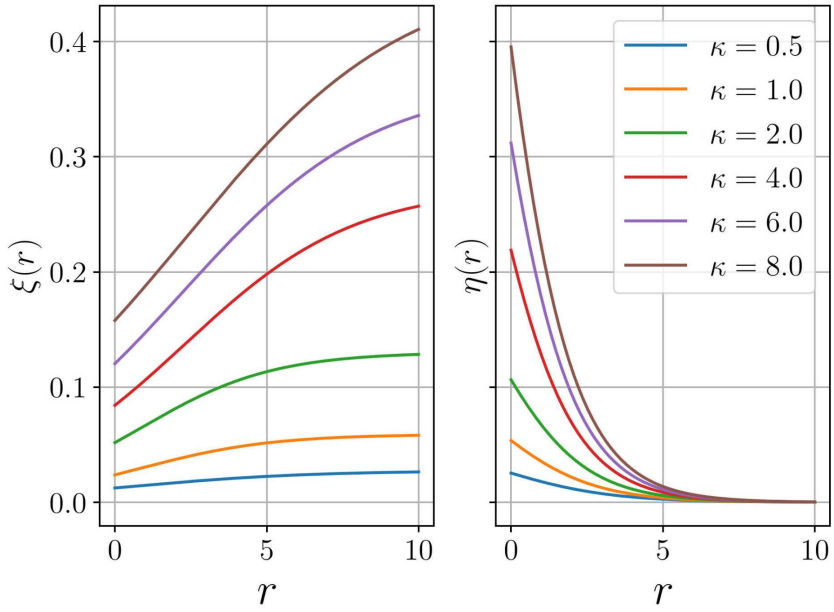

Figure 4.3: The optimized one- and two-body backflow potentials ξ and η, respectively, as functions of the distance r, for various values of interaction strength κ in a quantum dot of $N = 10$ spin-polarized electrons with $\beta = 6$.

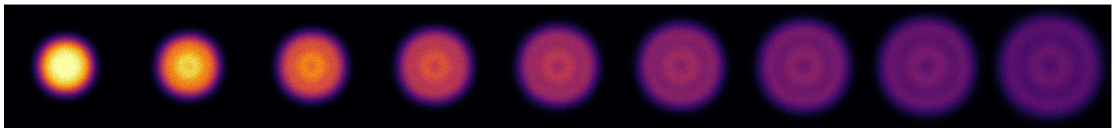

Figure 4.4: Several equally-spaced snapshots of the electron density along the continuous flow of coordinates at $\lambda = 0, 1/8, 2/8, \cdots, 1$. The system parameters are $N = 10, \beta = 6$ and $\kappa = 8$. Starting from the non-interacting reference state, the electron correlation effects can be gradually introduced, and one can finally reproduce the shell structure characteristic of a Wigner molecule in the strong interaction regime.

tentials together. Fig. 4.3 also shows that in the strong interaction regime, the backflow potentials deviate largely from the values of zero in the non-interacting case. To visualize how such strong potentials affect the evolution of electron coordinates, Fig. 4.4 shows several density snapshots along the continuous flow at $\lambda = 0, 1/8, 2/8, \cdots, 1$ for $\kappa = 8$. The cumulation of electron correlations and onset of the shell structure is clear.

5 Discussions

In essence, this work belongs to the large family of canonical transformation approaches for quantum many-body systems, except that the transformation is directly carried out upon particle coordinates instead of a many-body Fock space within the formulation of

second quantization [84–86]. Although the framework of neural canonical transformation is general, we have technically restricted to the subgroup of point transformations to make the calculation tractable, which corresponds to the specific choice $\hat{G} = \frac{1}{2}\{\boldsymbol{u}(\hat{\boldsymbol{x}}), \hat{\boldsymbol{p}}\}$ of the generator. The generalization to arbitrary generators, such as the Hamiltonian H in Eq. (1.1), seems fairly nontrivial in practice. In fact, the tractability of such a general calculation implies that it is possible to accurately simulate the real-time evolution $e^{-iHt}|\Phi\rangle$ of any many-body systems. In this perspective, the present approach can also be understood as a short-time variational approximation of adiabatic time evolution towards the thermal equilibrium.

One can take a different view of the limit of the present approach by inspecting the basis wavefunction representation Eq. (2.5). Compared to ground-state variational ansatzes, the coordinate transformation $\boldsymbol{f}$ in this work is implemented by continuous normalizing flow and subject to further limitations due to its invertibility [87–90]. Moreover, the continuous flow can only deform the nodal surface of reference states without changing its topology [24, 91, 92]. One example is the number of nodal cells of the ground-state wavefunction, which is conjectured to be always two for spatial dimensions d higher than one [92–95]. Similar conjecture has also been proposed for thermal density matrices at low temperature [95]. More thorough characterization of fermion nodes like possible topological obstructions is still lacking and worth further study [96,97]. In practice, one may remedy these issues by improving expressibility of the reference states $\Phi_n(\boldsymbol{z})$. For example, one can use more physically plausible reference states than the Slater determinants [98,99], or introduce additional parameters into the reference state which are pretrained or trained jointly with the flow transformation.

Although we have made use of continuous normalizing flow in this work, it should be possible to use other classes of permutation equivariant normalizing flows [18]. Some examples are the partitioned flow [62] and the invertible residual network [100], which can be more efficient than the present ODE-based implementation. One can also directly carry out Monte Carlo sampling of the electron coordinates other than transforming samples of the reference states. The local energy Eq. (3.2) resembles the score matching loss function [101] in training generative models, which is known to be expensive to compute. In light of this, advances in efficient score matching training might be beneficial to further reduce the computational efforts when scaling up to larger systems [102].

Building on these technical improvements in the implementation, a promising future direction is to scale up to larger particle number N and study other correlated fermion systems of fundamental importance, such as the homogeneous electron gas [103] and dense hydrogen [10]. Moreover, similar to what was shown in [36], one can also obtain information about the low-lying excited states of these systems as a byproduct of the thermodynamic calculation.

Acknowledgments

We thank Yuan Wan, Zi-Long Li, Hao Wu, Zi Cai, Jun Wang, Xiang Chen, and Vincent Moens for useful discussions. We thank Tobias Dornheim for providing the reference data shown in Fig. 4.1. This project is supported by the Strategic Priority Research Program of

the Chinese Academy of Sciences under Grant No. XDB30000000 and National Natural Science Foundation of China under Grant No. T2121001.

A Unitary representation of canonical transformations

We elaborate on the one-to-one correspondence between the group of unitary transformations in quantum mechanics and the group of canonical transformations in classical mechanics. Special emphasis will be placed on the subgroup of point transformations, which is the focus of the present work.

A.1 Basic formulation

In classical mechanics, a canonical transformation is a smooth bijection from the original set of phase space variables $(\boldsymbol{x}, \boldsymbol{p})$ to a new one $(\boldsymbol{X}(\boldsymbol{x}, \boldsymbol{p}), \boldsymbol{P}(\boldsymbol{x}, \boldsymbol{p}))$ satisfying the so-called symplectic condition, which is equivalent to saying that all Poisson brackets among the new variables are preserved [104]. To study its implication in the realm of quantum mechanics, we should convert the new (as well as old) variables into Hermitian operators $(\boldsymbol{X}, \boldsymbol{P}) \to (\hat{\boldsymbol{X}}, \hat{\boldsymbol{P}})$ following certain quantization procedure. As a result, $(\hat{\boldsymbol{X}}, \hat{\boldsymbol{P}})$ should satisfy the usual commutation relations of coordinate and momenta as $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{p}})$. This largely motivates us to reasonably expect the existence of a unitary transformation U that connects the two set of operators together:

$$\hat{\boldsymbol{X}} = U^\dagger \hat{\boldsymbol{x}} U, \quad \hat{\boldsymbol{P}} = U^\dagger \hat{\boldsymbol{p}} U. \tag{A.1}$$

U can thus be viewed as the unitary representation of the given canonical transformation.

To further clarify the nature of U, it is instructive to consider the eigenstate $|\boldsymbol{x})$ of the new coordinate operators $\hat{\boldsymbol{X}}$ defined as

$$\hat{\boldsymbol{X}}|\boldsymbol{x}) = \boldsymbol{x}|\boldsymbol{x}). \tag{A.2}$$

The essential point is that $|\boldsymbol{x})$ constitutes a *different* coordinate basis from the old ones $\hat{\boldsymbol{x}}|\boldsymbol{x}\rangle = \boldsymbol{x}|\boldsymbol{x}\rangle$, which is why a slightly different Dirac notation has been used in Eq. (A.2). By making use of these two basis, the unitary transformation U can then be formally defined as [48, 49]

$$U|\boldsymbol{x}) = |\boldsymbol{x}\rangle. \tag{A.3}$$

It is then straightforward to verify the operator transformation relations Eq. (A.1).

The transformation behavior of U upon the wavefunction $\Psi(\boldsymbol{x})$ of any given quantum state reads

$$\begin{aligned}(U\Psi)(\boldsymbol{x}) &= \langle \boldsymbol{x}|U|\Psi\rangle \\ &= (\boldsymbol{x}|\Psi\rangle = \int d\boldsymbol{x}'(\boldsymbol{x}|\boldsymbol{x}')\Psi(\boldsymbol{x}').\end{aligned} \tag{A.4}$$

The essential ingredient in evaluating this expression is the state overlap $\langle \boldsymbol{x}'|\boldsymbol{x})$, which can be shown to be uniquely determined (up to a phase factor) by the following set of

equations [49, 50]:

$$\hat{\boldsymbol{X}}\left(\boldsymbol{x}', -i\frac{\partial}{\partial \boldsymbol{x}'}\right)\langle \boldsymbol{x}'|\boldsymbol{x}\rangle = \boldsymbol{x}\langle \boldsymbol{x}'|\boldsymbol{x}\rangle, \tag{A.5a}$$

$$\hat{\boldsymbol{P}}\left(\boldsymbol{x}', -i\frac{\partial}{\partial \boldsymbol{x}'}\right)\langle \boldsymbol{x}'|\boldsymbol{x}\rangle = i\frac{\partial}{\partial \boldsymbol{x}}\langle \boldsymbol{x}'|\boldsymbol{x}\rangle. \tag{A.5b}$$

Unfortunately, for a general canonical transformation, the computation procedure outlined in Eqs. (A.4) and (A.5) above can be fairly difficult. We will thus restrict ourselves in the subgroup of point transformations, which can be constructed by the type-2 generating function $F_2(\boldsymbol{x}, \boldsymbol{P}) = \boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{P}$ via the following implicit relations [104]:

$$\boldsymbol{X} = \frac{\partial F_2}{\partial \boldsymbol{P}}, \quad \boldsymbol{p} = \frac{\partial F_2}{\partial \boldsymbol{x}}. \tag{A.6}$$

The resulting transformation formula can be obtained by some simple manipulations:

$$\boldsymbol{X} = \boldsymbol{f}(\boldsymbol{x}), \quad \boldsymbol{P} = \left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right)^{-T} \boldsymbol{p}. \tag{A.7}$$

In other words, the new coordinates depend only on the old coordinates through a bijective map $\boldsymbol{f}$, while the momenta are transformed in a covariant way so as to preserve the Poisson brackets. When such a point transformation is quantized, the corresponding expression for the state overlap $\langle \boldsymbol{x}'|\boldsymbol{x}\rangle$ turns out to be simple:

$$\langle \boldsymbol{x}'|\boldsymbol{x}\rangle = \delta\left(\boldsymbol{x}' - \boldsymbol{f}^{-1}(\boldsymbol{x})\right)\left|\det\left(\frac{\partial \boldsymbol{f}^{-1}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)\right|^{\frac{1}{2}}. \tag{A.8}$$

The correctness of this result can of course be verified by plugging it into Eq. (A.5). One can, however, get the intuitive feeling by noting that $\hat{\boldsymbol{X}} \equiv \boldsymbol{f}(\hat{\boldsymbol{x}})$ shares the same set of eigenstates as $\hat{\boldsymbol{x}}$, i.e., the Dirac delta function, while the additional Jacobian determinant factor is present to make the integration measure in the completeness relation $\int d\boldsymbol{x}|\boldsymbol{x})(\boldsymbol{x}| = 1$ as expected. By substituting (A.8) into Eq. (A.4), we finally obtain the transformed wavefunction through a point transformation as follows:

$$(U\Psi)(\boldsymbol{x}) = \Psi\left(\boldsymbol{f}^{-1}(\boldsymbol{x})\right)\left|\det\left(\frac{\partial \boldsymbol{f}^{-1}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)\right|^{\frac{1}{2}}. \tag{A.9}$$

This result is essentially Eq. (2.5) in the main text, where it is intuitively obtained from the perspective of normalizing flow.

A.2 Infinitesimal unitary transformations

The characteristics of the unitary representation of canonical transformations formulated above can be more clearly revealed by studying the infinitesimal behavior in the vicinity of identity transformation, which is a common practice in physics and of great theoretical

importance. An infinitesimal canonical transformation can be constructed by the type-2 generating function $F_2(\boldsymbol{x}, \boldsymbol{P}) = \boldsymbol{x} \cdot \boldsymbol{P} + d\lambda\, G(\boldsymbol{x}, \boldsymbol{P})$, where the two terms correspond to identity transformation and infinitesimal perturbation, respectively. Substituting this expression into Eq. (A.6) and retaining only lowest-order contributions, one could then obtain a continuous family $(\boldsymbol{x}(\lambda), \boldsymbol{p}(\lambda))$ of canonical transformations specified by the Hamilton's equations of motion:

$$\frac{d\boldsymbol{x}}{d\lambda} = \frac{\partial G}{\partial \boldsymbol{p}}, \quad \frac{d\boldsymbol{p}}{d\lambda} = -\frac{\partial G}{\partial \boldsymbol{x}}. \tag{A.10}$$

The function $G(\boldsymbol{x}, \boldsymbol{p})$ is usually also called the generating function. When such a canonical transformation is quantized, Eq. (A.10) is naturally replaced by the Heisenberg equations of motion. This observation is essential: in light of Eq. (A.1), the corresponding unitary transformation can be immediately recognized as

$$U = e^{-i\hat{G}\lambda}, \tag{A.11}$$

where the generator $\hat{G}$ should be ensured to be Hermitian by the quantization procedure.

Eq. (A.11) is clearly a more explicit and meaningful characterization of the unitary transformation than the formal definition (A.3). In particular, the state overlap appearing in Eq. (A.4) corresponds precisely to the propagator $(\boldsymbol{x}|\boldsymbol{x}') = \langle \boldsymbol{x}|e^{-i\hat{G}\lambda}|\boldsymbol{x}'\rangle$. Such a propagator is difficult to evaluate in general cases, so we again concentrate only on point transformations as in the previous section. Within the formulation presented here, a point transformation corresponds to the choice $G(\boldsymbol{x}, \boldsymbol{p}) = \boldsymbol{u}(\boldsymbol{x}) \cdot \boldsymbol{p}$, which leads to the type-2 generating function

$$F_2(\boldsymbol{x}, \boldsymbol{P}) = (\boldsymbol{x} + d\lambda\, \boldsymbol{u}(\boldsymbol{x})) \cdot \boldsymbol{P}. \tag{A.12}$$

To obtain the corresponding unitary representation, we employ a simple operator symmetrization $\hat{G} = \frac{1}{2}\{\boldsymbol{u}(\hat{\boldsymbol{x}}), \hat{\boldsymbol{p}}\}$ to make the generator Hermitian, as mentioned in the main text. Comparing Eq. (A.12) with the form $F_2(\boldsymbol{x}, \boldsymbol{P}) = \boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{P}$ discussed in the previous section, one can readily reach the conclusion that the transformed wavefunction $(U\Psi)(\boldsymbol{x}) = \langle \boldsymbol{x}|e^{-\frac{i}{2}\{\boldsymbol{u}(\hat{\boldsymbol{x}}), \hat{\boldsymbol{p}}\}\lambda}|\Psi\rangle$ can be equivalently written in the form of Eq. (A.9), where the coordinate bijection $\boldsymbol{f}$ is specified by the ODE

$$\frac{d\boldsymbol{x}}{d\lambda} = \boldsymbol{u}(\boldsymbol{x}). \tag{A.13}$$

We thus rigorously show the equivalence of the two basis wavefunction expressions Eqs. (2.2) and (2.5), which lies at the core of the finite-temperature approach in this work.

Finally, we give a few guidelines for the derivation of the continuity equation (2.3) from Eq. (2.2) in the main text for readers' convenience. Note that $\hat{\boldsymbol{p}} = -i\nabla$, we have

$$\begin{aligned}
\frac{\partial}{\partial\lambda}\Psi_n(\boldsymbol{x}, \lambda) &= -\frac{i}{2}\left\{\boldsymbol{u}(\hat{\boldsymbol{x}}), \hat{\boldsymbol{p}}\right\}\Psi_n(\boldsymbol{x}, \lambda) \\
&= -\frac{1}{2}\Big[\boldsymbol{u}(\boldsymbol{x}) \cdot \nabla\Psi_n(\boldsymbol{x}, \lambda) + \nabla \cdot (\boldsymbol{u}(\boldsymbol{x})\Psi_n(\boldsymbol{x}, \lambda))\Big] \\
&= -\boldsymbol{u}(\boldsymbol{x}) \cdot \nabla\Psi_n(\boldsymbol{x}, \lambda) - \frac{1}{2}\Psi_n(\boldsymbol{x}, \lambda)\nabla \cdot \boldsymbol{u}(\boldsymbol{x}).
\end{aligned} \tag{A.14}$$

To obtain Eq. (2.3), simply multiply Eq. (A.14) by $\Psi_n^*(\boldsymbol{x}, \lambda)$ and add the resulting equation to its own complex conjugate.

B Some more benchmark data for 2D quantum dot

The following table summarizes our results for the energy of a two-dimensional quantum dot at $\beta = 10$, for various electron number N and interaction strength κ. PIMC results from [74] are also listed when available. All data correspond to the fully spin-polarized case. Our finite-temperature calculations indicate that the entropy is negligible for a temperature as low as $\beta = 10$, so our energy results can be treated as variational. We anticipate these results (as well as those presented in the main text) can be further improved by adopting better model architecture and optimization schemes. We also note the results reported in Figure 4.8 of [83] for $\beta = 10$, $N = 3$, $\kappa = 2$ show that the data in [74] may be subject to slight systematic errors.

N	κ	This work	[74]
3	2	8.331(3)	8.37(1)
3	4	11.070(4)	11.05(1)
3	6	13.495(6)	13.43(1)
3	8	15.653(7)	15.59(1)
4	2	14.336(4)	14.30(5)
4	4	19.517(7)	19.42(1)
4	6	24.060(9)	23.790(12)
4	8	28.178(12)	27.823(11)
6	0.5	18.179(4)	–
6	1	22.003(6)	–
6	1.5	25.600(8)	–
6	2	28.994(9)	–
6	3	35.241(10)	–
6	4	41.012(11)	–
6	5	46.385(13)	–
6	6	51.448(13)	–
6	7	56.270(16)	–
6	8	60.837(15)	60.42(2)

References

[1] Stefano Giorgini, Lev P. Pitaevskii, and Sandro Stringari. Theory of ultracold atomic fermi gases. *Rev. Mod. Phys.*, 80:1215–1274, Oct 2008.

[2] Stephanie M. Reimann and Matti Manninen. Electronic structure of quantum dots. *Rev. Mod. Phys.*, 74:1283–1342, Nov 2002.

[3] Tobias Dornheim, Simon Groth, and Michael Bonitz. The uniform electron gas at warm dense matter conditions. *Physics Reports*, 744:1–86, 2018.

[4] E. Y. Loh, J. E. Gubernatis, R. T. Scalettar, S. R. White, D. J. Scalapino, and R. L. Sugar. Sign problem in the numerical simulation of many-electron systems. *Phys. Rev. B*, 41:9301–9307, May 1990.

[5] DM Ceperley. Path integral Monte Carlo methods for fermions. In K. Binder and G. Ciccotti, editors, *Monte Carlo and Molecular Dynamics of Condensed Matter Systems*. Editrice Compositori, Bologna, Italy, 1996.

[6] T. Dornheim. Fermion sign problem in path integral Monte Carlo simulations: Quantum dots, ultracold atoms, and warm dense matter. *Phys. Rev. E*, 100:023307, Aug 2019.

[7] Joonho Lee, Miguel A. Morales, and Fionn D. Malone. A phaseless auxiliary-field quantum Monte Carlo perspective on the uniform electron gas at finite temperatures: Issues, observations, and benchmark study. *Journal of Chemical Physics*, 154(6):064109, 2021.

[8] Matthew R. Hermes and So Hirata. Finite-temperature coupled-cluster, many-body perturbation, and restricted and unrestricted Hartree-Fock study on one-dimensional solids: Luttinger liquids, Peierls transitions, and spin- and charge-density waves. *Journal of Chemical Physics*, 143(10):102818, 2015.

[9] Albrecht Huber. *Variational Principles in Quantum Statistical Mechanics*, pages 364–392. Springer US, Boston, MA, 1968.

[10] Burkhard Militzer and E. L. Pollock. Variational density matrix method for warm, condensed matter: Application to dense hydrogen. *Phys. Rev. E*, 61:3470–3482, Apr 2000.

[11] Fionn D. Malone, N. S. Blunt, Ethan W. Brown, D. K. K. Lee, J. S. Spencer, W. M. C. Foulkes, and James J. Shepherd. Accurate exchange-correlation energies for the warm dense electron gas. *Phys. Rev. Lett.*, 117:115701, Sep 2016.

[12] Tapta Kanchan Roy and M. Durga Prasad. Development of a new variational approach for thermal density matrices. *The Journal of Chemical Physics*, 134(21):214110, 2011.

[13] Kensaku Takai, Kota Ido, Takahiro Misawa, Youhei Yamaji, and Masatoshi Imada. Finite-temperature variational Monte Carlo method for strongly correlated electron systems. *Journal of the Physical Society of Japan*, 85(3):034601, 2016.

[14] Jahan Claes and Bryan K. Clark. Finite-temperature properties of strongly correlated systems via variational Monte Carlo. *Phys. Rev. B*, 95:205109, May 2017.

[15] Shuo-Hui Li, Chen-Xiao Dong, Linfeng Zhang, and Lei Wang. Neural canonical transformation with symplectic flows. *Phys. Rev. X*, 10:021020, Apr 2020.

[16] W. L. McMillan. Ground state of liquid he^4. *Phys. Rev.*, 138:A442–A451, Apr 1965.

[17] D. Ceperley, G. V. Chester, and M. H. Kalos. Monte Carlo simulation of a many-fermion study. *Phys. Rev. B*, 16:3081–3099, Oct 1977.

[18] George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021.

[19] I. Kobyzev, S. Prince, and M. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.

[20] Robert Jastrow. Many-body problem with strong forces. *Phys. Rev.*, 98:1479–1484, Jun 1955.

[21] R. P. Feynman and Michael Cohen. Energy spectrum of the excitations in liquid helium. *Phys. Rev.*, 102:1189–1204, Jun 1956.

[22] K. E. Schmidt, Michael A. Lee, M. H. Kalos, and G. V. Chester. Structure of the ground state of a fermion fluid. *Phys. Rev. Lett.*, 47:807–810, Sep 1981.

[23] Yongkyung Kwon, D. M. Ceperley, and Richard M. Martin. Effects of three-body and backflow correlations in the two-dimensional electron gas. *Phys. Rev. B*, 48:12037–12046, Oct 1993.

[24] P. López Ríos, A. Ma, N. D. Drummond, M. D. Towler, and R. J. Needs. Inhomogeneous backflow transformations in quantum Monte Carlo calculations. *Phys. Rev. E*, 74:066701, Dec 2006.

[25] Di Luo and Bryan K. Clark. Backflow transformations via neural networks for quantum many-body wave functions. *Phys. Rev. Lett.*, 122:226401, Jun 2019.

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

[27] David Pfau, James S. Spencer, Alexander G. D. G. Matthews, and W. M. C. Foulkes. Ab initio solution of the many-electron Schrödinger equation with deep neural networks. *Phys. Rev. Research*, 2:033429, Sep 2020.

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

[29] Kenny Choo, Antonio Mezzacapo, and Giuseppe Carleo. Fermionic neural-network states for ab-initio electronic structure. *Nature Communications*, 11(1):2368, May 2020.

[30] 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. *Phys. Rev. B*, 102:205122, Nov 2020.

[31] James S. Spencer, David Pfau, Aleksandar Botev, and W. M. C. Foulkes. Better, Faster Fermionic Neural Networks. *arXiv e-prints*, page arXiv:2011.07125, November 2020.

[32] Max Wilson, Nicholas Gao, Filip Wudarski, Eleanor Rieffel, and Norm M. Tubman. Simulations of state-of-the-art fermionic neural network wave functions with diffusion Monte Carlo. *arXiv e-prints*, page arXiv:2103.12570, March 2021.

[33] Corey Adams, Giuseppe Carleo, Alessandro Lovato, and Noemi Rocco. Variational Monte Carlo calculations of $a \leq 4$ nuclei with an artificial neural-network correlator ansatz. *Phys. Rev. Lett.*, 127:022502, Jul 2021.

[34] John Martyn and Brian Swingle. Product spectrum ansatz and the simplicity of thermal states. *Phys. Rev. A*, 100:032107, Sep 2019.

[35] Guillaume Verdon, Jacob Marks, Sasha Nanda, Stefan Leichenauer, and Jack Hidary. Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm. *arXiv e-prints*, page arXiv:1910.02071, 2019.

[36] Jin-Guo Liu, Liang Mao, Pan Zhang, and Lei Wang. Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits. *Machine Learning: Science and Technology*, 2(2):025011, Feb 2021.

[37] Shuo-Hui Li and Lei Wang. Neural network renormalization group. *Phys. Rev. Lett.*, 121:260601, Dec 2018.

[38] Dian Wu, Lei Wang, and Pan Zhang. Solving statistical mechanics using variational autoregressive networks. *Phys. Rev. Lett.*, 122:080602, Feb 2019.

[39] Linfeng Zhang, Weinan E, and Lei Wang. Monge-ampère flow for generative modeling. *arXiv e-prints*, page arXiv:1809.10188, 2018.

[40] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457), 2019.

[41] M. S. Albergo, G. Kanwar, and P. E. Shanahan. Flow-based generative models for markov chain Monte Carlo in lattice field theory. *Phys. Rev. D*, 100:034515, Aug 2019.

[42] Gurtej Kanwar, Michael S. Albergo, Denis Boyda, Kyle Cranmer, Daniel C. Hackett, Sébastien Racanière, Danilo Jimenez Rezende, and Phiala E. Shanahan. Equivariant flow-based sampling for lattice gauge theory. *Phys. Rev. Lett.*, 125:121601, Sep 2020.

[43] Kyle Cranmer, Siavash Golkar, and Duccio Pappadopulo. Inferring the quantum density matrix with machine learning. *arXiv e-prints*, page arXiv:1904.05903, 2019.

[44] Jonas Khler, Leon Klein, and Frank No. Equivariant flows: sampling configurations for multi-body systems with symmetric energies. *arXiv e-prints*, page arXiv:1910.00753, 2019.

[45] Yang Li, Haidong Yi, Christopher Bender, Siyuan Shan, and Junier B Oliva. Exchangeable neural ode for set modeling. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 6936–6946. Curran Associates, Inc., 2020.

[46] Marin Biloš and Stephan Günnemann. Equivariant normalizing flows for point processes and sets. 2021.

[47] M Eger and EP Gross. Point transformations and the many body problem. *Annals of Physics*, 24:63–88, 1963.

[48] M. Moshinsky and C. Quesne. Linear canonical transformations and their unitary representations. *Journal of Mathematical Physics*, 12(8):1772–1780, 1971.

[49] Marcos Moshinsky. Canonical transformations and quantum mechanics. *SIAM Journal on Applied Mathematics*, 25(2):193–212, 1973.

[50] Gerardo Torres del Castillo, Hector Bello Martinez, Ricardo Mejia, and J.M. Paz. Representation of canonical transformations in quantum mechanics. *Revista Mexicana de Fsica*, 55, 04 2009.

[51] Maciej Baszak and Ziemowit Domański. Canonical transformations in quantum mechanics. *Annals of Physics*, 331:70–96, 2013.

[52] Bryce Seligman DeWitt. Point transformations in quantum mechanics. *Phys. Rev.*, 85:653–661, Feb 1952.

[53] James B. Anderson. A random-walk simulation of the Schrödinger equation: H+3. *The Journal of Chemical Physics*, 63(4):1499–1503, 1975.

[54] Esteban G Tabak and Eric Vanden-Eijnden. Density estimation by dual ascent of the log-likelihood. *Communications in Mathematical Sciences*, 8(1):217–233, 2010.

[55] Ariel Barr, Willem Gispen, and Austen Lamacraft. Quantum ground states from reinforcement learning. In Jianfeng Lu and Rachel Ward, editors, *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, volume 107 of *Proceedings of Machine Learning Research*, pages 635–653, Princeton University, Princeton, NJ, USA, 20–24 Jul 2020. PMLR.

[56] E. Weinan. A Proposal on Machine Learning via Dynamical Systems. *Communications in Mathematics and Statistics*, 5(1):1–11, 2017.

[57] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31, pages 6571–6583. Curran Associates, Inc., 2018.

[58] Michele Taddei, Michele Ruggeri, Saverio Moroni, and Markus Holzmann. Iterative backflow renormalization procedure for many-body ground-state wave functions of strongly interacting normal fermi liquids. *Phys. Rev. B*, 91:115106, Mar 2015.

[59] Michele Ruggeri, Saverio Moroni, and Markus Holzmann. Nonlinear network description for many-body quantum systems in continuous space. *Phys. Rev. Lett.*, 120:205302, May 2018.

[60] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[61] Linfeng Zhang, Jiequn Han, Han Wang, Wissam Saidi, Roberto Car, and Weinan E. End-to-end symmetry preserving inter-atomic potential energy model for finite and extended systems. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31, pages 4436–4446. Curran Associates, Inc., 2018.

[62] Peter Wirnsberger, Andrew J. Ballard, George Papamakarios, Stuart Abercrombie, Sbastien Racanire, Alexander Pritzel, Danilo Jimenez Rezende, and Charles Blundell. Targeted free energy estimation via learned mappings. *The Journal of Chemical Physics*, 153(14):144112, 2020.

[63] Shakir Mohamed, Mihaela Rosca, Michael Figurnov, and Andriy Mnih. Monte Carlo gradient estimation in machine learning. *Journal of Machine Learning Research*, 21(132):1–62, 2020.

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

[65] Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey. *Journal of Machine Learning Research*, 18(153):1–43, 2018.

[66] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[67] See `https://github.com/buwantaiji/FermiFlow` for a code implementation in PyTorch.

[68] Tobias Dornheim, Simon Groth, Alexey Filinov, and Michael Bonitz. Permutation blocking path integral Monte Carlo: A highly efficient approach to the simulation of strongly degenerate non-ideal fermions. *New Journal of Physics*, 17(7):073017, jul 2015.

[69] Tobias Dornheim, Michele Invernizzi, Jan Vorberger, and Barak Hirshberg. Attenuating the fermion sign problem in path integral Monte Carlo simulations using the bogoliubov inequality and thermodynamic integration. *The Journal of Chemical Physics*, 153(23):234104, 2020.

[70] A. Harju, S. Siljamäki, and R. M. Nieminen. Wigner molecules in quantum dots: A quantum Monte Carlo study. *Phys. Rev. B*, 65:075309, Jan 2002.

[71] S. M. Reimann, M. Koskinen, and M. Manninen. Formation of wigner molecules in small quantum dots. *Phys. Rev. B*, 62:8108–8113, Sep 2000.

[72] Francesco Pederiva, C. J. Umrigar, and E. Lipparini. Diffusion Monte Carlo study of circular quantum dots. *Phys. Rev. B*, 62:8120–8125, Sep 2000.

[73] C. H. Mak, R. Egger, and H. Weber-Gottschick. Multilevel blocking approach to the fermion sign problem in path-integral Monte Carlo simulations. *Phys. Rev. Lett.*, 81:4533–4536, Nov 1998.

[74] R. Egger, W. Häusler, C. H. Mak, and H. Grabert. Crossover from fermi liquid to wigner molecule behavior in quantum dots. *Phys. Rev. Lett.*, 82:3320–3323, Apr 1999.

[75] Jens Harting, Oliver Mülken, and Peter Borrmann. Interplay between shell effects and electron correlations in quantum dots. *Phys. Rev. B*, 62:10207–10211, Oct 2000.

[76] A. V. Filinov, M. Bonitz, and Yu. E. Lozovik. Wigner crystallization in mesoscopic 2d electron systems. *Phys. Rev. Lett.*, 86:3851–3854, Apr 2001.

[77] Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla. Recurrent neural network wave functions. *Phys. Rev. Research*, 2:023358, Jun 2020.

[78] Alex Kulesza and Ben Taskar. Determinantal point processes for machine learning. *Foundations and Trends® in Machine Learning*, 5(23):123–286, 2012.

[79] Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang. Unsupervised generative modeling using matrix product states. *Phys. Rev. X*, 8:031012, Jul 2018.

[80] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[81] T. Schoof, M. Bonitz, A. Filinov, D. Hochstuhl, and J.W. Dufty. Configuration path integral Monte Carlo. *Contributions to Plasma Physics*, 51(8):687–697, 2011.

[82] N. S. Blunt, T. W. Rogers, J. S. Spencer, and W. M. C. Foulkes. Density-matrix quantum Monte Carlo method. *Phys. Rev. B*, 89:245124, Jun 2014.

[83] Tim Schoof. *Configuration Path Integral Monte Carlo: ab initio simulations of fermions in the warm dense matter regime*. PhD thesis, 2017.

[84] Franz Wegner. Flow-equations for hamiltonians. *Annalen der physik*, 506(2):77–91, 1994.

[85] Stanislaw D. Glazek and Kenneth G. Wilson. Perturbative renormalization group for hamiltonians. *Phys. Rev. D*, 49:4214–4218, Apr 1994.

[86] Steven R. White. Numerical canonical transformation approach to quantum many-body problems. *The Journal of Chemical Physics*, 117(16):7472–7482, 2002.

[87] Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[88] Rob Cornish, Anthony L. Caterini, George Deligiannidis, and Arnaud Doucet. Relaxing bijectivity constraints with continuously indexed normalising flows. *arXiv e-prints*, page arXiv:1909.13833, 2019.

[89] Han Zhang, Xi Gao, Jacob Unterman, and Tom Arodz. Approximation capabilities of neural ODEs and invertible residual networks. In Hal Daum III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 11086–11095. PMLR, 13–18 Jul 2020.

[90] Zhifeng Kong and Kamalika Chaudhuri. The expressive power of a class of normalizing flow models. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 3599–3609. PMLR, 26–28 Aug 2020.

[91] Michal Bajdich, Lubos Mitas, Gabriel Drobný, and Lucas K. Wagner. Approximate and exact nodes of fermionic wavefunctions: Coordinate transformations and topologies. *Phys. Rev. B*, 72:075131, Aug 2005.

[92] Dario Bressanini. Implications of the two nodal domains conjecture for ground state fermionic wave functions. *Phys. Rev. B*, 86:115120, Sep 2012.

[93] D. M. Ceperley. Fermion nodes. *Journal of Statistical Physics*, 63(5):1237–1267, Jun 1991.

[94] Lubos Mitas. Structure of fermion nodes and nodal cells. *Phys. Rev. Lett.*, 96:240402, Jun 2006.

[95] Lubos Mitas. Fermion nodes and nodal cells of noninteracting and interacting fermions. *arXiv e-prints*, pages arXiv:cond–mat/0605550, 2006.

[96] William A. Glauser, Willard R. Brown, William A. Lester, D. Bressanini, Brian L. Hammond, and M. L. Koszykowski. Random-walk approach to mapping nodal regions of N-body wave functions: Ground-state Hartree-Fock wave functions for Li-C. *The Journal of Chemical Physics*, 97(12):9200–9215, 1992.

[97] Frank Krüger and Jan Zaanen. Fermionic quantum criticality and the fractal nodal surface. *Phys. Rev. B*, 78:035104, Jul 2008.

[98] M. Bajdich, L. Mitas, G. Drobný, L. K. Wagner, and K. E. Schmidt. Pfaffian pairing wave functions in electronic-structure quantum Monte Carlo simulations. *Phys. Rev. Lett.*, 96:130201, Apr 2006.

[99] M. Bajdich, L. Mitas, L. K. Wagner, and K. E. Schmidt. Pfaffian pairing and backflow wavefunctions for electronic structure quantum Monte Carlo methods. *Phys. Rev. B*, 77:115112, Mar 2008.

[100] Jens Behrmann, Will Grathwohl, Ricky T. Q. Chen, David Duvenaud, and Joern-Henrik Jacobsen. Invertible residual networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 573–582. PMLR, 09–15 Jun 2019.

[101] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(24):695–709, 2005.

[102] Yang Song, Sahaj Garg, Jiaxin Shi, and Stefano Ermon. Sliced score matching: A scalable approach to density and score estimation. In Ryan P. Adams and Vibhav Gogate, editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 574–584. PMLR, 22–25 Jul 2020.

[103] Tobias Dornheim, Simon Groth, Fionn D. Malone, Tim Schoof, Travis Sjostrom, W. M. C. Foulkes, and Michael Bonitz. Ab initio quantum Monte Carlo simulation of the warm dense electron gas. *Physics of Plasmas*, 24(5):056303, 2017.

[104] Herbert Goldstein, Charles P Poole, and John L Safko. *Classical Mechanics: Pearson New International Edition*. Pearson, third new international edition, 2013.