Enhancing Neural Network Variational Monte Carlo through Basis Transformation

Zhixuan Liu[1,2], Dongheng Qian[1,2] and Jing Wang[1,2,3,4]

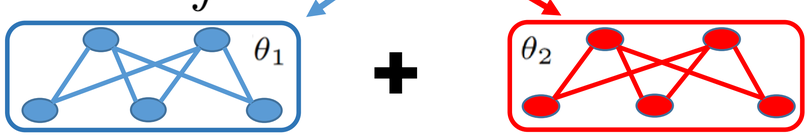
[1]State Key Laboratory of Surface Physics and Department of Physics, Fudan University, Shanghai 200433, China
[2]Shanghai Research Center for Quantum Sciences, Shanghai 201315, China
[3]Institute for Nanoelectronic Devices and Quantum Computing, Fudan University, Shanghai 200433, China
[4]Hefei National Laboratory, Hefei 230088, China

Introduction

Neural-network variational Monte Carlo (NN-VMC) has become a powerful tool for quantum many-body problems. Typically, there are two ways to improve performance: Enhancing the network architecture & Improving the update strategy.

Here, we take a different route: Introducing a many-body basis transformation with a variational description.

$$\tilde{\psi}_\theta(\mathbf{r}) = \int d\mathbf{x}\, \psi_{\theta_1}(\mathbf{x})\, G_{\theta_2}(\mathbf{x}, \mathbf{r})$$



This immediately raises two questions:

(1) What kind of basis transformation should be introduced?

(2) How to update parameters in the presence of basis transformation?

Basis Transformation

We consider a general basis transformation in a continuous model. The Hamiltonian matrix is modified accordingly: $H_\alpha(\mathbf{x}, \mathbf{x}') = \langle G_\alpha(\mathbf{x})|\hat{H}|G_\alpha(\mathbf{x}')\rangle$

When the basis set is non-orthogonal, the overlap matrix is no longer the identity matrix: $I_\alpha(\mathbf{x}, \mathbf{x}') = \langle G_\alpha(\mathbf{x})|G_\alpha(\mathbf{x}')\rangle$. The energy is rewritten as:

$$E_\theta = \frac{\int d\mathbf{x}d\mathbf{x}'\, \psi^*_{\theta_1}(\mathbf{x})\psi_{\theta_1}(\mathbf{x}')H_\alpha(\mathbf{x}, \mathbf{x}')}{\int d\mathbf{x}d\mathbf{x}'\, \psi^*_{\theta_1}(\mathbf{x})\psi_{\theta_1}(\mathbf{x}')I_\alpha(\mathbf{x}, \mathbf{x}')}. \quad (1)$$

Usually, the Hamiltonian becomes nonlocal after the basis transformation, necessitating the evaluation of an additional integral over x'. This is a high-dimensional integral in an Nd dimensional space, where N is the number of particles and d is the spatial dimension.

<u>How to handle this integral constitutes the main challenge of the basis transformation approach</u>

Gaussian Basis

Here, we choose Gaussian Basis with form : $G_\alpha(\mathbf{x}, \mathbf{r}) = \left(\frac{\alpha}{\pi}\right)^{3n/2} \exp\left(-\alpha \sum_{i=1}^{n} |\mathbf{r}_i - \mathbf{x}_i|^2\right)$, with variational parameter α.

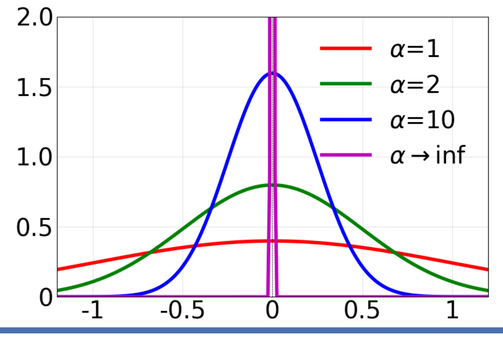
Only one parameter is added!

- **Property 1:** Its overlap matrix has form $I_\alpha(\mathbf{x}, \mathbf{x}') = \left(\frac{\alpha}{2\pi}\right)^{3n/2} \exp\left(-\frac{\alpha}{2}|\mathbf{x} - \mathbf{x}'|^2\right)$, which has the same form as normal distribution

  ➡ Integral over $\boldsymbol{x}'$ can be evaluated stochastically $\int dx' I(x, x') f(x') = \mathbb{E}_{x' \sim \mathcal{N}(x, 1/\sqrt{\alpha})} f(x')$

- **Property 2:** Parameter α controls the locality of $G_\alpha(\mathbf{x}, \mathbf{r})$, in the $\alpha \to \infty$ limit we have: $G_{\alpha\to\infty}(\mathbf{x}, \mathbf{r}) = \delta(r - x)$

$G_\alpha(r, x = 0)$ distribution in 1D projection



Energy Gradient

Traditional VMC calculate energy with Monte Carlo integral: $E_\theta = \mathbb{E}_{p_\theta(\mathbf{x})}[E_L(\mathbf{x})]$, with local energy E_L defined as $E_L(\mathbf{x}) = \int d\mathbf{x}' H(\mathbf{x}, \mathbf{x}')\psi_{\theta_1}(\mathbf{x}')/\psi_{\theta_1}(\mathbf{x})$ and distribution function $p_\theta(x)$ defined as $p_\theta(\mathbf{x}) \propto |\psi_{\theta_1}(\mathbf{x})|^2$

However, when non-orthogonal basis is considered, the denominator integrand in Eq. (1) is no longer a real positive function. We construct a positive distribution by taking absolute values: $p_\theta(\mathbf{x}) \propto \int d\mathbf{x}' \, |\psi^*_{\theta_1}(\mathbf{x})| \, |\psi_{\theta_1}(\mathbf{x}')| \, I_\alpha(\mathbf{x}, \mathbf{x}')$. Energy can be rewritten as:

$$E_\theta = \frac{\mathbb{E}_{p_\theta}[S_L(\mathbf{x})\, E_L(\mathbf{x})]}{\mathbb{E}_{p_\theta}[S_L(\mathbf{x})]},$$

$$S_L(\mathbf{x}) = \frac{\int d\mathbf{x}' \, I_\alpha(\mathbf{x}, \mathbf{x}') \, |\psi_{\theta_1}(\mathbf{x}')| \, \mathrm{sgn}\left[\psi^*_{\theta_1}(\mathbf{x})\psi_{\theta_1}(\mathbf{x}')\right]}{\int d\mathbf{x}' \, I_\alpha(\mathbf{x}, \mathbf{x}')|\psi_{\theta_1}(\mathbf{x}')|} \qquad E_L(\mathbf{x}) = \frac{\int d\mathbf{x}' \, H_\alpha(\mathbf{x}, \mathbf{x}')\psi_{\theta_1}(\mathbf{x}')}{\int d\mathbf{x}' \, I_\alpha(\mathbf{x}, \mathbf{x}')\psi_{\theta_1}(\mathbf{x}')}$$

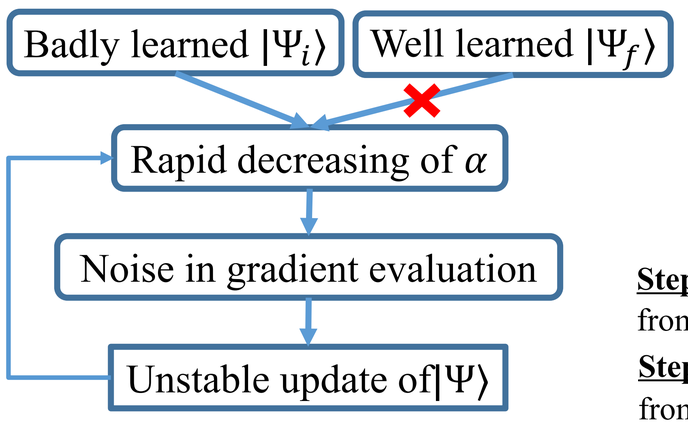
Differentiating Eq. (2) gradient yields the gradient with respect to $\theta = (\theta_1, \alpha)$

$$\nabla_\theta E_\theta = 2\,\mathrm{Re}\left\{\frac{\mathbb{E}_{p_\theta(\mathbf{x})}[O^*_\theta(\mathbf{x})\,\epsilon(\mathbf{x})\,S_L(\mathbf{x})]}{\mathbb{E}_{p_\theta(\mathbf{x})} S_L(\mathbf{x})}\right\}$$

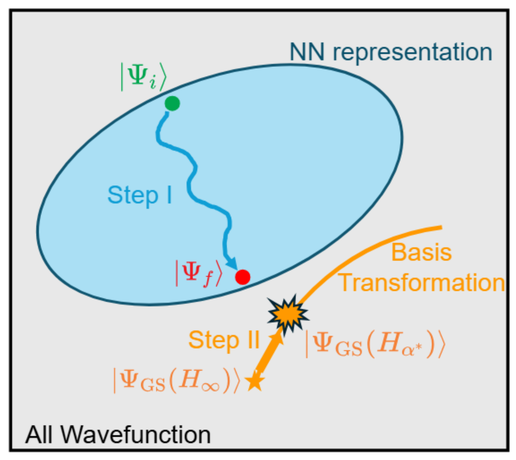
$$\epsilon(\mathbf{x}) = E_L(\mathbf{x}) - E_\theta \qquad O_{\theta_1}(\mathbf{x}) = \nabla_{\theta_1}\psi_{\theta_1}(\mathbf{x})/\psi_{\theta_1}(\mathbf{x}) \qquad O_\alpha(\mathbf{x}) = -\frac{1}{4\alpha^2}\nabla^2_{\mathbf{x}}\psi(\mathbf{x})/\psi(\mathbf{x})$$

Two Step Optimization

A naive strategy is to update θ_1 and α simultaneously. However, we find that this approach is prone to instability. A crucial observation is that the locality of the Gaussian basis, controlled by α, directly impacts the accuracy of gradient estimates. This may lead to a optimization failure:



Two step scheme



<u>Step I</u>: fix $\alpha \to \infty$ and update θ_1 from $|\Psi_i\rangle$ to $|\Psi_f\rangle$(traditional VMC)

<u>Step II</u>: fix $|\Psi_f\rangle$ and update α from $\alpha \sim \infty$ to $\alpha = \alpha^*$

Key point is to avoid a to small value of α

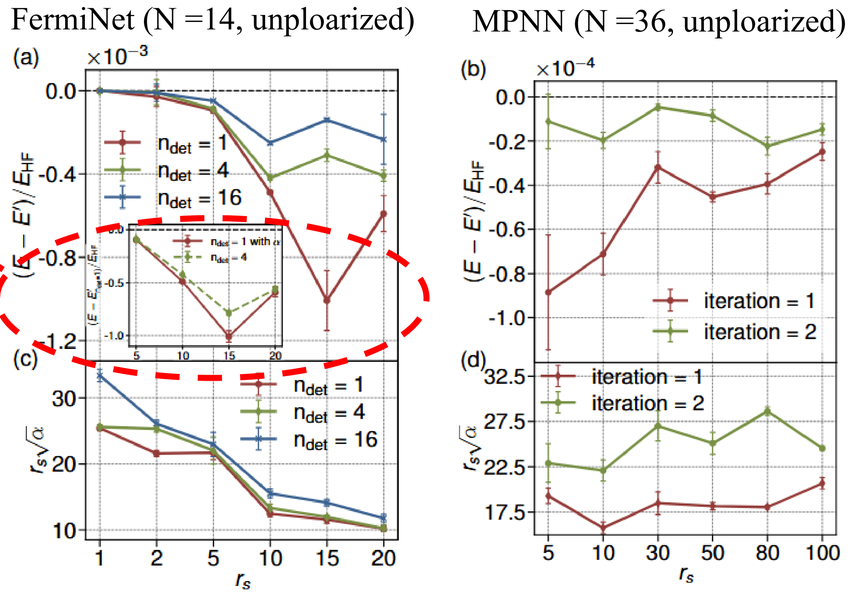
3D HEG

We benchmark the method on three-dimensional homogeneous electron gas. Its Hamiltonian, in Hartree atomic units, is given by $H = -\frac{1}{2}\sum_i \nabla_i^2 + \sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \text{b.g.}$

We found adding a single parameter α yields lower energies at $r_s = 5, 10, 20$ than increasing n_{det} in FermiNet from 1 to 4, which introduces over 10,000 additional parameters.

<u>Two set of network structures are tested</u>

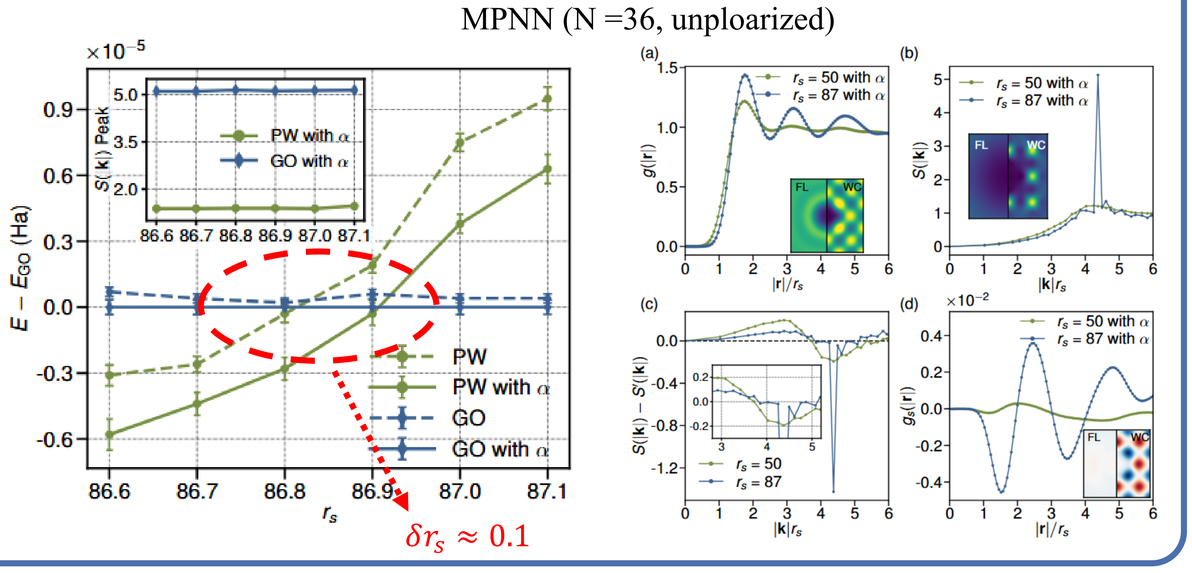
FermiNet (N =14, unploarized) MPNN (N =36, unploarized)



<u>More accurate estimation of the FL-WC phase transition point</u>

<u>Correlation functions</u>

MPNN (N =36, unploarized)



Conclusion

- Variation Monte Carlo under non-orthogonal basis transformation with variational basis paraments is derived, exploring new perspective for enhancing NN-VMC in continuous space.
- Using the three-dimensional homogeneous electron gas as an example, we verify the universality and efficiency of introducing the basis transformation. Furthermore, an important application of the method is presented, which enables a more accurate determination of the phase transition.