One More Time: Revisiting Neural Quantum States from a Reinforcement Learning Perspective

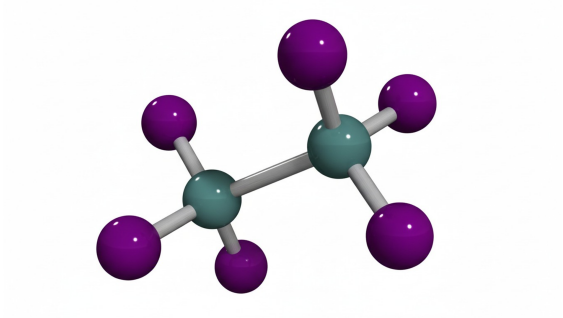
Juan Agustín Duque*, Sergio García Heredia*,
Vinicius Hernandes, Eliska Greplova,
Thomas Spriggs**, Aaron Courville**, Anna Dawid**

NeurIPS 2026

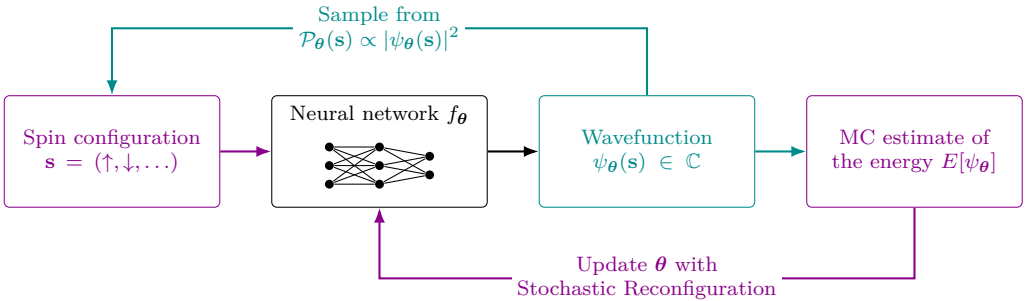
Quantum Many-Body Physics

It studies the collective behavior of systems of many quantum particles in:

- Condensed matter.
- Quantum chemistry.
- Nuclear physics.
- Etc.



Neural Quantum States



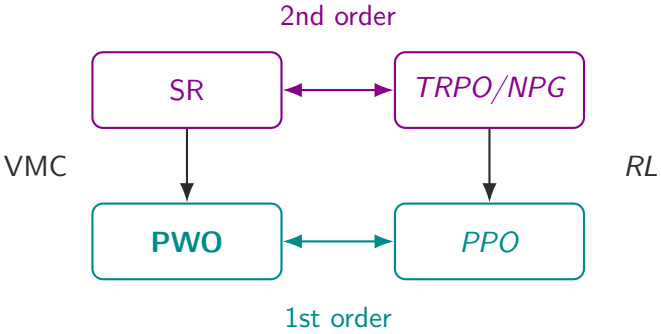
Limitations

Many **problems** in NQS have to do with **sampling**:

- Many **samples** are drawn for each training step and **discarded** afterwards.
- **minSR scales** with the **cube** of the number of **samples**.
- **MCMC** can be sample-**inefficient**.
- **AR** models are an alternative, but **minSR** is often **ill-conditioned**.

As a result, the **biggest models** used for NQS, which are only ~ 1 **M parameters**, can be expensive to train and rely on a **careful design**.

Connection with Reinforcement Learning



Proximal Wavefunction Optimization

If $A^{\mathrm{R}}_{\boldsymbol{\theta}}(\mathbf{s}) \equiv \mathrm{Re}\left\{\Delta E^{\mathrm{loc}}_{\boldsymbol{\theta}}(\mathbf{s})\right\}$ and $A^{\mathrm{I}}_{\boldsymbol{\theta}}(\mathbf{s}) \equiv \mathrm{Im}\left\{\Delta E^{\mathrm{loc}}_{\boldsymbol{\theta}}(\mathbf{s})\right\}$, then

$$\boldsymbol{\nabla}_{\boldsymbol{\theta}} E[\psi_{\boldsymbol{\theta}}] = \mathbb{E}_{\mathbf{s}\sim\mathcal{P}_{\boldsymbol{\theta}}}\left[2A^{\mathrm{R}}_{\boldsymbol{\theta}}(\mathbf{s})\boldsymbol{\nabla}_{\boldsymbol{\theta}} \log|\psi_{\boldsymbol{\theta}}|\right] + \mathbb{E}_{\mathbf{s}\sim\mathcal{P}_{\boldsymbol{\theta}}}\left[2A^{\mathrm{I}}_{\boldsymbol{\theta}}(\mathbf{s})\boldsymbol{\nabla}_{\boldsymbol{\theta}} \arg\psi_{\boldsymbol{\theta}}\right]$$

Motivated by this, PWO extends PPO from RL to NQS to overcome the problems of SR, minimizing the sum of two surrogates with a fixed reference $\boldsymbol{\theta}_{\mathrm{old}}$:

$$L_{\mathrm{PWO}}(\boldsymbol{\theta};\boldsymbol{\theta}_{\mathrm{old}}) = L^{\mathrm{clip}}_{\mathrm{mod}}(\boldsymbol{\theta};\boldsymbol{\theta}_{\mathrm{old}}) + L^{\mathrm{clip}}_{\mathrm{arg}}(\boldsymbol{\theta};\boldsymbol{\theta}_{\mathrm{old}}),$$

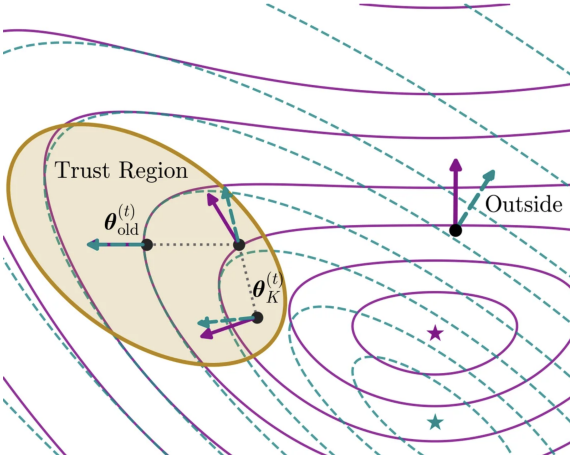
where

$$L^{\mathrm{clip}}_{\mathrm{mod}}(\boldsymbol{\theta};\boldsymbol{\theta}_{\mathrm{old}}) = \mathbb{E}_{\mathbf{s}\sim\mathcal{P}_{\boldsymbol{\theta}_{\mathrm{old}}}}\left[\max\left(r_{\boldsymbol{\theta}}A^{\mathrm{R}}_{\boldsymbol{\theta}_{\mathrm{old}}}, \mathrm{clip}(r_{\boldsymbol{\theta}}, 1-\epsilon_{\mathrm{mod}}, 1+\epsilon_{\mathrm{mod}})A^{\mathrm{R}}_{\boldsymbol{\theta}_{\mathrm{old}}}\right)\right],$$

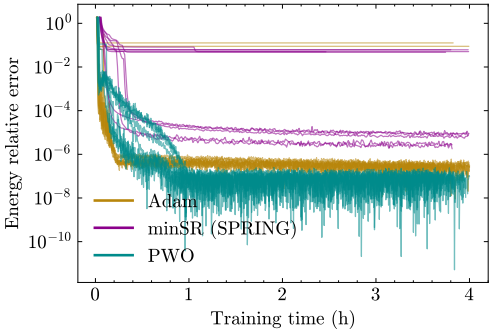
$$L^{\mathrm{clip}}_{\mathrm{arg}}(\boldsymbol{\theta};\boldsymbol{\theta}_{\mathrm{old}}) = \mathbb{E}_{\mathbf{s}\sim\mathcal{P}_{\boldsymbol{\theta}_{\mathrm{old}}}}\left[\mathrm{sg}(r_{\boldsymbol{\theta}})\max\left(\phi_{\boldsymbol{\theta}}A^{\mathrm{I}}_{\boldsymbol{\theta}_{\mathrm{old}}}, \mathrm{clip}(\phi_{\boldsymbol{\theta}}, -\epsilon_{\mathrm{arg}}, \epsilon_{\mathrm{arg}})A^{\mathrm{I}}_{\boldsymbol{\theta}_{\mathrm{old}}}\right)\right].$$

PWO guarantees

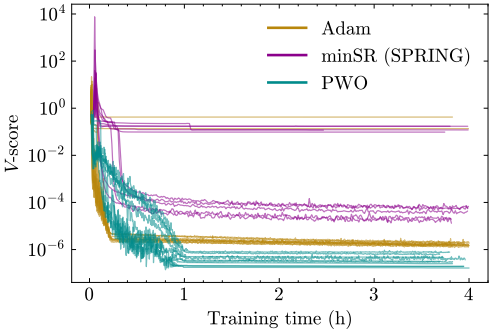
1. Gradient consistency.
2. Upper bound on the energy.
3. Infidelity control.



Spin systems



(a) Relative error.



(b) V-score.

Figure 1. J_1-J_2 Heisenberg model in a periodic chain of 12 spins with $J_2/J_1 = 0.5$.

Scalability

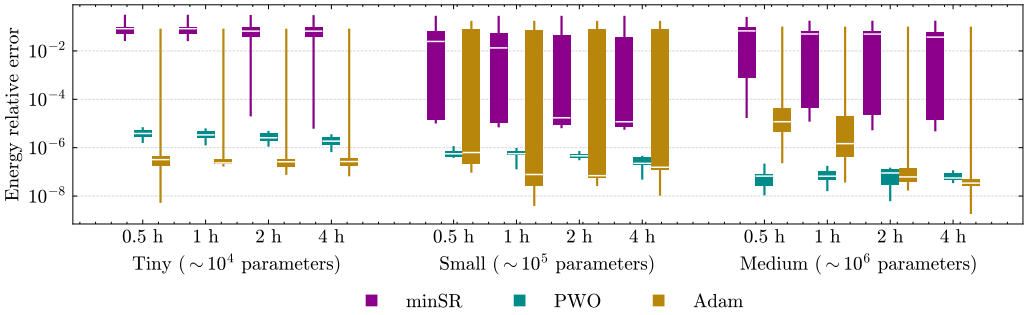
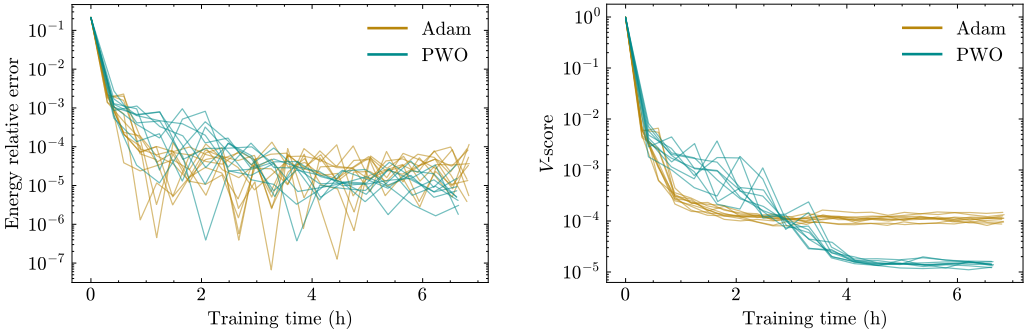


Figure 2. J_1-J_2 Heisenberg model in a periodic chain of 12 spins with $J_2/J_1 = 0.5$.

LLM as Wavefunction



(a) Relative error. (b) V-score.

Figure 3. 1.5 B RWKV-7 for TFIM in a periodic chain of 12 spins at the critical point.

Contributions

1. Explicit **connection** between **NQS** and **RL**.
2. **Proximal Wavefunction Optimization** (PWO).
3. **Theoretical** and **empirical evidence** supporting PWO.

Thanks for your attention



Paper



Code



Blog Post