

Neural Quantum State tutorial notes

NGAI, Lee Yeung

September 2026

Please be aware that most of the materials here will NOT be covered in any test or exam.

1 Introduction

What is neural quantum states(NQS)? To summarize NQS in a simple sentence. NQS is a technique of using neural networks to represent ground state wavefunctions. The benefits of using NQS to find ground state are, it is unsupervised, “correct” outputs are not required to train this neural network and it is sign-problem free for most observables, which allowed researchers to study models that traditional Quantum Monte Carlo methods cannot study. Let say we are interested in studying the ground state of a Hamiltonian.

$$\hat{H} = -J \sum_{i=0}^L \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z - h \sum_{i=0}^{L-1} \hat{\sigma}_i^x$$

With periodic boundary condition $\hat{\sigma}_L^\alpha = \hat{\sigma}_0^\alpha$ Where $\hat{\sigma}^x, \hat{\sigma}^y, \hat{\sigma}^z$ are the Pauli matrices and J, h denote the interaction between two atoms and magnetic field strength respectively.

This is the famous Transverse Field Ising Model, which is trying to model the situation where a bunch of atoms that are only allowed to have 2 states, either spin-up or spin-down are placed under a transverse magnetic-field.

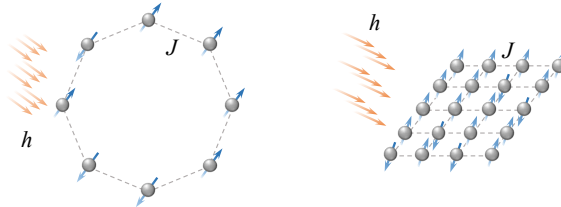


Figure 1: Left: 1-D TFIM with Periodic boundary condition Right: 2-D TFIM without Periodic boundary condition

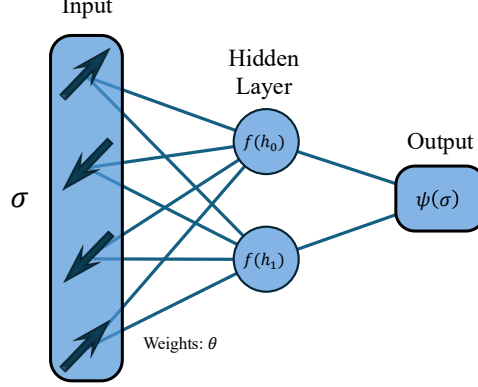


Figure 2: **Demonstration of a neural quantum state constructed by a single layer Feed Forward Neural Network** Where $f(x)$ denotes the activation function

Now, Let $|\psi\rangle = \sum_{\sigma \in S_z \text{ basis}} \psi(\sigma)|\sigma\rangle$ denotes the ground state of this Hamiltonian. Where $\psi(\sigma)$ is the wave function of the ground state *Then the goal of NQS is to use a neural network to represent $\psi(\sigma)$* . Let us denote this wavefunction represented by neural network be $\psi_\theta(\sigma)$, where θ denote the vector where its components are all parameters in the neural network.

2 Monte Carlo Sampling

One of the greatest problems in quantum many-body physics is the exponentially growing vector space. For example, for a 4×4 model there are 2^{16} number of spins combination while for an 8×8 model the size of it already grew into 2^{64} . It is impossible to either do exact-diagonalization or even represent the ground-state in vector form for most system size. However, common sense tells us that most basis are boring they are quite unlikely to contribute in the ground state. This is why we need Monte Carlo sampling, to let us focus on those basis which contribute the most.

One simple example is diagonal observables

$$\begin{aligned} \langle \hat{O} \rangle &= \frac{1}{\langle \psi | \psi \rangle} \langle \psi | \hat{O} | \psi \rangle \\ &= \frac{1}{\langle \psi | \psi \rangle} \sum_{\sigma, \sigma' \in S_z \text{ basis}} \psi^*(\sigma') \psi(\sigma) \langle \sigma' | \hat{O} | \sigma \rangle = \frac{1}{\langle \psi | \psi \rangle} \sum_{\sigma \in S_z \text{ basis}} |\psi(\sigma)|^2 O_{loc}(\sigma) \end{aligned}$$

This statistically equals

$$= \mathbb{E}_{\sigma \sim |\psi(\sigma)|^2 / \langle \psi | \psi \rangle} (O_{loc})$$

Where $O_{loc}(\sigma) = \langle \sigma | \hat{O} | \sigma \rangle$ is the local estimator of the diagonal observable.

This allow us to obtain $\langle \hat{O} \rangle$ by just letting σ follow the distribution of $\frac{|\psi(\sigma)|^2}{\langle \psi | \psi \rangle}$ and take the mean of $O_{loc}(\sigma)$. Such that we don't have to explore the entire set of basis.

It is always easier said than done. How do we sample σ according the distribution of $|\psi(\sigma)|^2 / \langle \psi | \psi \rangle$? This is where the **Metropolis-Hastings algorithm** comes in.

Roughly speaking, in practice, this algorithm goes as follows:

1. You are now having a state $\sigma_{old} = |\uparrow\uparrow\uparrow\rangle$
2. propose a new state by randomly flipping one spin $\sigma_{new} = |\uparrow\downarrow\uparrow\rangle$
3. Accept this newly proposed state with the probability of $P_{accept} = \min\left(1, \frac{|\psi(\sigma_{new})|^2}{|\psi(\sigma_{old})|^2}\right)$
4. If the newly proposed state is accepted, keep the new state, otherwise keep the old state and record it as a sample. Use the recorded state as the initial state for generate the next state.
5. Repeat steps 1 to 4 for a number of times, until the sample size is large enough for estimate the expectation value of the observable.

3 Loss function

The variational principle in quantum mechanics goes as follows:

For any states $|\psi\rangle$. We must have $\frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} \geq E_0$ Where E_0 denotes the ground

state energy such that the ground state $|\Psi\rangle$ satisfies $\frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = E_0$.

Therefore, it is natural to define energy as the loss function such that *as we keep minimizing the energy of the neural quantum state, the NQS will keep getting closer to the ground state.* Please note that energy is not diagonal. Therefore, the formulation of the expectation value of energy goes as follows.

$$\begin{aligned} \langle E \rangle &= \langle \hat{H} \rangle = \frac{1}{\langle \psi | \psi \rangle} \langle \psi | \hat{H} | \psi \rangle \\ &= \frac{1}{\langle \psi | \psi \rangle} \sum_{\sigma, \sigma' \in S_z \text{ basis}} \psi^*(\sigma) \psi(\sigma') \langle \sigma' | \hat{H} | \sigma \rangle \\ &= \frac{1}{\langle \psi | \psi \rangle} \sum_{\sigma \in S_z \text{ basis}} |\psi(\sigma)|^2 \sum_{\sigma' \in S_z \text{ basis}} \frac{\psi(\sigma')}{\psi(\sigma)} \langle \sigma | \hat{H} | \sigma' \rangle \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\langle \psi | \psi \rangle} \sum_{\sigma \in S_z \text{ basis}} |\psi(\sigma)|^2 E_{loc}(\sigma) \\
&= \mathbb{E}_{\sigma \sim |\psi(\sigma)|^2 / \langle \psi | \psi \rangle} (E_{loc})
\end{aligned}$$

Where $E_{loc} = \sum_{\sigma'} \frac{\psi(\sigma')}{\psi(\sigma)} \langle \sigma | \hat{H} | \sigma' \rangle$ is the local estimator of energy.

4 Minimize the loss function

How do we tell the parameters in θ that where should they go? We need them to go in a specific way such that $\langle \hat{H} \rangle$ is minimized. First of all, consider the partial derivative of the NQS with respect to a weight θ^i in the neural network

$$\frac{\partial \langle \hat{H} \rangle}{\partial \theta^i} = \frac{\partial}{\partial \theta^i} \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle}$$

Now let us denote the partial derivative as an operator such that $\frac{\partial}{\partial \theta^i} = \hat{O}_i$. Hence, $\langle \psi_\theta | \hat{O}_i^\dagger = \frac{\partial}{\partial \theta^i} \langle \psi_\theta |$ and $\hat{O}_i | \psi_\theta \rangle = \frac{\partial}{\partial \theta^i} | \psi_\theta \rangle$ Therefore,

$$\begin{aligned}
&\frac{\partial}{\partial \theta^i} \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} \\
&= \frac{\langle \psi_\theta | \psi_\theta \rangle \left(\langle \psi_\theta | \hat{H} \hat{O}_i | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i^\dagger \hat{H} | \psi_\theta \rangle \right) - \langle \psi_\theta | \hat{H} | \psi_\theta \rangle \left(\langle \psi_\theta | \hat{O}_i^\dagger | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle \right)}{\langle \psi_\theta | \psi_\theta \rangle^2} \\
&= \langle \hat{H} \hat{O}_i \rangle + \langle \hat{O}_i^\dagger \hat{H} \rangle - \langle \hat{H} \rangle \langle \hat{O}_i \rangle - \langle \hat{O}_i^\dagger \rangle \langle \hat{H} \rangle
\end{aligned}$$

It can be showed that

$$\langle \hat{H} \hat{O}_i \rangle = \overline{\langle \hat{O}_i^\dagger \hat{H} \rangle} \text{ and } \langle \hat{O}_i \rangle = \overline{\langle \hat{O}_i^\dagger \rangle}$$

Therefore,

$$\frac{\partial \langle \hat{H} \rangle}{\partial \theta^i} = 2\text{Re} \left(\langle \hat{H} \hat{O}_i \rangle - \langle \hat{H} \rangle \langle \hat{O}_i \rangle \right)$$

Where

$$\begin{aligned}
\langle \hat{O}_i \rangle &= \sum_{\sigma \in S_z \text{ basis}} |\psi(\sigma)|^2 \frac{\partial \ln(\psi_\theta(\sigma))}{\partial \theta^i} \\
&= \sum_{\sigma \in S_z \text{ basis}} |\psi(\sigma)|^2 O_{loc}(\sigma) = \mathbb{E}_{\sigma \sim |\psi(\sigma)|^2 / \langle \psi | \psi \rangle} (O_{loc})
\end{aligned}$$

and

$$\langle \hat{H} \hat{O}_i \rangle = \sum_{\sigma \in S_z \text{ basis}} |\psi_\theta(\sigma)|^2 \frac{\partial \ln(\psi_\theta(\sigma))}{\partial \theta^i} \sum_{\sigma' \in S_z \text{ basis}} \frac{\psi_\theta^*(\sigma')}{\psi_\theta^*(\sigma)} \langle \sigma' | \hat{H} | \sigma \rangle$$

$$= \sum_{\sigma \in S_z \text{ basis}} |\psi_\theta(\sigma)|^2 [HO_i]_{loc}(\sigma) = \mathbb{E}_{\sigma \sim |\psi(\sigma)|^2 / \langle \psi | \psi \rangle} ([HO_i]_{loc})$$

As to how to calculate $\frac{\partial \ln(\psi_\theta(\sigma))}{\partial \theta^i}$. Since $\psi_\theta(\sigma)$ is a neural network. Machine learning techniques can be applied. For example, by using **backpropagation algorithm**. We can find the partial derivative with respect to θ^i starting from the front most layer to the back most layer efficiently by chain-rule in calculus. Hence, we can perform **stochastic gradient descent** to update θ^i by

$$\theta_{n+1}^i = \theta_n^i - \eta \frac{\partial \langle \hat{H} \rangle}{\partial \theta^i} \quad (1)$$

Where n denotes the iteration and η denotes the learning rate, usually it is set to be 10^{-2} to 10^{-5} depending on your goal.

By repeating the update for enough times. The loss function, i.e. the NQS energy, will be close enough to the ground state energy and we can consider the NQS as “close enough” to the ground state.

5 A flaw in the vanilla gradient descent

Although for most people Eq.(1) is natural form of gradient descent. However, there is a hidden assumption behind that formulation. It assumed the **distance between two quantum states are Euclidean**, i.e the definition of distance is $||\vec{v}|| = \sqrt{v_1^2 + v_2^2 + \dots + v_N^2}$ for an N -dimensional vector $\vec{v}$ Which is not true apparently. Consider some basis transformation, the distance between two quantum states is mostly likely not conserved before and after the transformation.

Therefore, as a physicist, we need some other definition of distance, in this case we need the **Fubini-Study distance**. The definition of the Fubini-Study distance is as follows:

$$d(\psi, \phi) = \arccos \left(\sqrt{\frac{\langle \psi | \phi \rangle \langle \phi | \psi \rangle}{\langle \psi | \psi \rangle \langle \phi | \phi \rangle}} \right)$$

Now assume there is a very small change in all parameters, denote it by $\delta\theta$ The Fubini-Study distance between ψ_θ and $\psi_{\theta+\delta\theta}$ is given by

$$d(\psi_\theta, \psi_{\theta+\delta\theta}) = \arccos \left(\sqrt{\frac{\langle \psi_\theta | \psi_{\theta+\delta\theta} \rangle \langle \psi_{\theta+\delta\theta} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \psi_{\theta+\delta\theta} | \psi_{\theta+\delta\theta} \rangle}} \right)$$

Then let both states are normalized such that $\langle \psi_\theta | \psi_\theta \rangle = \langle \psi_{\theta+\delta\theta} | \psi_{\theta+\delta\theta} \rangle = 1$ Note that we can Taylor expand the NQS with infinitesimal change as follows

$$\psi_{\theta+\delta\theta} = \psi_\theta + \frac{\partial \psi_\theta}{\partial \theta^i} \delta\theta^i + \frac{1}{2} \frac{\partial^2 \psi_\theta}{\partial \theta^i \partial \theta^j} \delta\theta^i \delta\theta^j + O(\delta\theta^i \delta\theta^j \delta\theta^k)$$

Please be aware that we applied the Einstein summation notation, which means the summation signs are ignored.

Hence

$$\langle \psi_\theta | \psi_{\theta+\delta\theta} \rangle = 1 + \langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle \delta\theta^i + \frac{1}{2} \langle \psi_\theta | \hat{O}_i \hat{O}_j | \psi_\theta \rangle \delta\theta^i \delta\theta^j + O(\delta\theta^i \delta\theta^j \delta\theta^k)$$

Therefore,

$$\begin{aligned} & \langle \psi_\theta | \psi_{\theta+\delta\theta} \rangle \langle \psi_{\theta+\delta\theta} | \psi_\theta \rangle \\ &= 1 + \left(\langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_j^\dagger | \psi_\theta \rangle \right) \delta\theta^i \\ &+ \left(\frac{1}{2} \langle \psi_\theta | \hat{O}_i \hat{O}_j | \psi_\theta \rangle + \frac{1}{2} \langle \psi_\theta | \hat{O}_i^\dagger \hat{O}_j^\dagger | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle \langle \psi_\theta | \hat{O}_j^\dagger | \psi_\theta \rangle \right) \delta\theta^i \delta\theta^j + O(\delta\theta^i \delta\theta^j \delta\theta^k) \end{aligned} \quad (2)$$

Notice that

$$\frac{\partial}{\partial\theta^i} \langle \psi_\theta | \psi_\theta \rangle = \langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i^\dagger | \psi_\theta \rangle = 0$$

and

$$\frac{\partial^2}{\partial\theta^i \partial\theta^j} \langle \psi_\theta | \psi_\theta \rangle = \langle \psi_\theta | \hat{O}_i \hat{O}_j | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_j^\dagger \hat{O}_i | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i^\dagger \hat{O}_j | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i^\dagger \hat{O}_j^\dagger | \psi_\theta \rangle = 0$$

Hence, Eq.(2) can be rewritten as follows

$$= 1 + \left[\langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle \langle \psi_\theta | \hat{O}_j^\dagger | \psi_\theta \rangle - \frac{1}{2} \left(\langle \psi_\theta | \hat{O}_j^\dagger \hat{O}_i | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i^\dagger \hat{O}_j | \psi_\theta \rangle \right) \right] \delta\theta^i \delta\theta^j + \dots$$

Finally by arcsin approximation. If $x \approx 1$, then

$$\arccos^2(x) = 1 - x^2 + O((1 - x^2)^2)$$

Therefore

$$\begin{aligned} & \arccos^2(|\langle \psi_\theta | \psi_{\theta+\delta\theta} \rangle|) \\ &= - \left[\langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle \langle \psi_\theta | \hat{O}_j^\dagger | \psi_\theta \rangle - \frac{1}{2} \left(\langle \psi_\theta | \hat{O}_j^\dagger \hat{O}_i | \psi_\theta \rangle + \langle \psi_\theta | \hat{O}_i^\dagger \hat{O}_j | \psi_\theta \rangle \right) \right] \delta\theta^i \delta\theta^j + \dots \\ &= \text{Re} \left[\langle \psi_\theta | \hat{O}_i^\dagger \hat{O}_j | \psi_\theta \rangle - \langle \psi_\theta | \hat{O}_i | \psi_\theta \rangle \langle \psi_\theta | \hat{O}_j^\dagger | \psi_\theta \rangle \right] \delta\theta^i \delta\theta^j + \dots \\ &= \text{Re}(G_{ij}) \delta\theta^i \delta\theta^j = g_{ij} \delta\theta^i \delta\theta^j \end{aligned}$$

Where $g_{ij} = \text{Re}(G_{ij})$ is the **Fubini-Study metric tensor** and G_{ij} is the **Quantum Geometric Tensor**

Now consider the update of parameters by stochastic gradient descent as follows

$$\theta_{n+1}^i = \theta_n^i - \eta \delta\theta^i$$

The change in parameters with respect to Fubini-Study distance are now given by

$$(g_{ij}) \delta\theta^i = (\nabla E)_j$$

Hence

$$\delta\theta^i = (g^{-1})^{ij}(\nabla E)_j$$

Where $(\nabla E)_j = \frac{\partial \langle \hat{H} \rangle}{\partial \theta^j}$ is the tensor form of partial derivative of $\langle \hat{H} \rangle$ with respect to θ^j

Hence, we can be able to perform gradient descent with respect to Fubini-Study distance. This technique is also known as the **stochastic reconfiguration**. For more details please refer to [2].

6 Conclusion

Although the theory and math behind the Neural Quantum States seem tedious and complicated, all the stuff I brought out are just a tip of an iceberg. I would say some math here is not even valid. But for education purposes, I have to trade some mathematical rigorousness for simplicity.

However, if you don't care anything about theory. There is a python package that did everything for you.

It is the NetKet package! Unfortunately, It would be too much if I cover the coding part in this lecture. Therefore, I will leave the tutorial of NetKet in the reference page for your own practice. Please refer to [1] for more details.

References

- [1] NetKet. *Tutorials*. NetKet documentation. URL: <https://netket.readthedocs.io/en/latest/tutorials/index.html>.
- [2] James Stokes et al. “Quantum Natural Gradient”. In: *Quantum* 4 (May 2020), p. 269. DOI: 10.22331/q-2020-05-25-269. URL: <https://doi.org/10.22331/q-2020-05-25-269>.