Do Neural Quantum States Exhibit Double Descent?

Rebecca Merber & Anthony Ashmore

Many-body QM

Many-body quantum mechanics is the study of large numbers of interacting quantum particles. For example, if we wanted to understand many electrons and how they interact, or the Ising Model, we need to understand many-body quantum mechanics. The simplest states of these systems have minimum energy, and their wavefunctions satisfy the equation:

$$H\,|\Psi\rangle = E_0\,|\Psi\rangle$$

For two-state particles, the Hamiltonian, H, acts as a $2^N \times 2^N$matrix, where N is the number of particles. It is hard to find the exact wavefunction Ψ when N is large, as the size of the Hilbert space scales exponeitally with the system, $\dim \mathscr{H} = 2^N$. One method to avoid this problem is to use a neural network to model the wavefunction.

Neural Quantum States

We find the ground state wavefunction by approximating Ψ using a neural network (Carleo & Troyer, `16). The network is trained to model the ground state by being given a sample of configurations chosen by Monte Carlo. Then, the network returns Ψ evaluated on those configurations, which it uses to estimate the energy, $E = \langle H \rangle$. Finally, the network uses gradient descent to change the network parameters, θ, to iteratively reduce the energy:

$$\Delta\theta = -\eta \nabla_\theta E$$

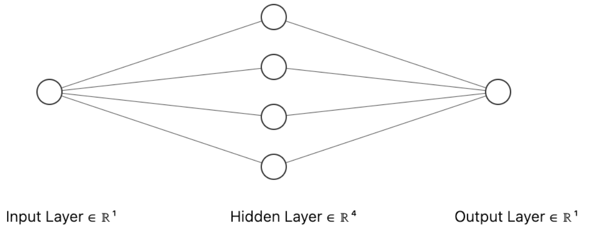


Figure 3: The basic outline of a dense network. The input takes in a configuration, and the output is interpreted as a probability amplitude, Ψ. The number and width of hidden layers can be varied to vary the number of parameters, to increase the expressivity of the network.

Double Descent

Deep neural networks often suffer from an effect called **double descent** - the network performance initially gets worse as the number of parameters increase, and then improves as the network becomes overparametrized (Nakkiran et al.,`19). Recent work claimed that neural quantum states exhibit this effect, roughly when the size of the Hilbert space is equal to the number of parameters (Moss et al., `25). This would be bad - it is impossible to reach the overparametrized regime for physically relevant systems. Our goal is to test different networks for this effect.

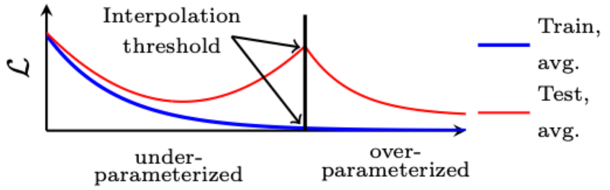


Figure 1: Double-Descent (Moss et al, 2025)

Lattice Gauge Theory

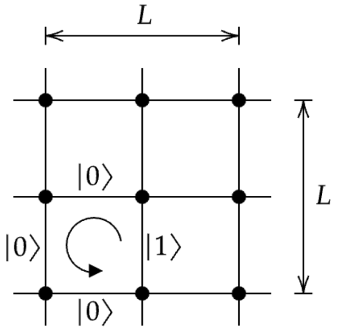


Figure 2: States on links between lattice sites for a 1x1 plaquette

Describes an interacting field force, like electromagnetism, where space is discretized (Horn et al., `79). The theory is described using unitary operators Q and P, which act on the states on each link:

$$Q\,|0\rangle = |0\rangle, \quad Q\,|1\rangle = -\,|1\rangle$$

$$P\,|0\rangle = |1\rangle, \quad P\,|1\rangle = |0\rangle$$

Physics given by a Hamiltonian:

$$H = g^2 \sum_l (1 - P_l) + \frac{1}{g^2} \sum_\square (1 - Q_{l_1} Q_{l_2} Q_{l_3} Q_{l_4})$$

$$\approx \frac{g^2}{2} E^2 + \frac{1}{2g^2} B^2$$

This theory is **gauge invariant** - highly nontrivial symmetry - and exhibits a **phase transition** as g is varied.

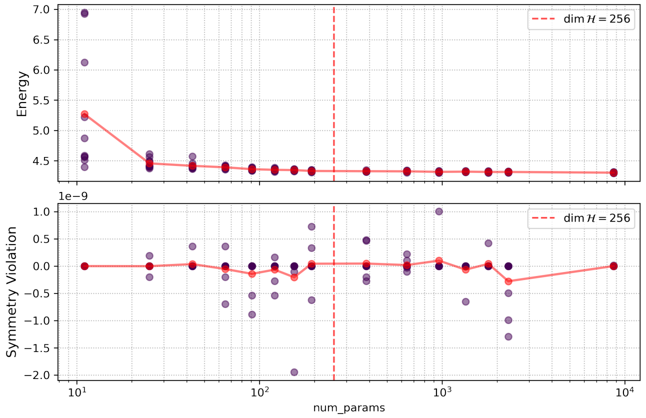
Phases around the loops are gauge invariant variables

e.g. $Q_1 Q_2 Q_3 Q_4 = (1)(-1)(1)(1) = -1$

Looking for Double Descent

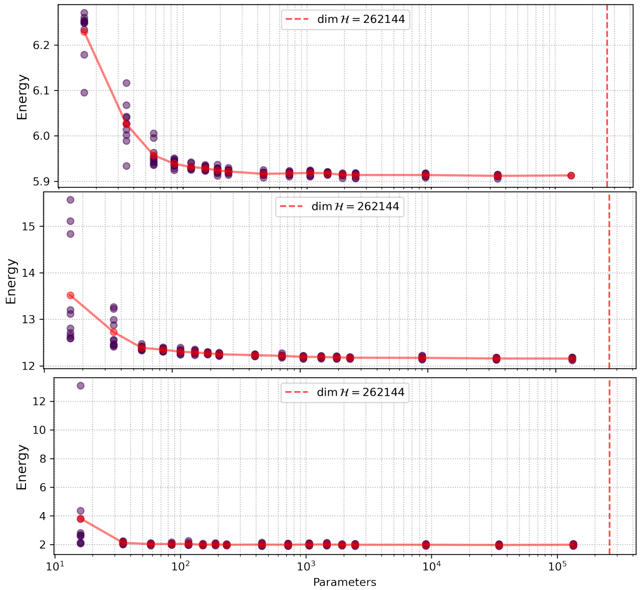
We constructed two types of neural quantum states to model the ground state of this lattice gauge theory to look for double descent: a **dense feed-forward network** which takes in the phases on the links, and a **gauge invariant network** which takes in the phases around the loops. The dense feed-forward network must learn the gauge invariance, while the gauge invariant network is "physics-informed." The networks were constructed and trained using NetKet and JAX (Vicentini et al., `22).

Results



We varied the number of parameters by changing the width of the hidden layers of the network.

- For each network we did ten training runs and examined how the final ground state energy varied with the network size.
- We observed large networks had smaller ground state energy with no obvious double descent peak.
- In all cases, the gauge symmetry is preserved to machine precision.
- We found similar results for the dense network, but its performance was worse overall.



To ensure that these results held for larger lattices, we repeated our experiments for a 3×3 lattice with coupling g set below, at, and above the phase transition.

Discussion & Conclusion

The dense networks were unable to accurately model the ground state wavefunction while the gauge invariant networks performed well. However, we do not observe double descent for either network

In all of our experiments, we saw no indication of double descent. Why?

Moss et al. used supervised learning to calculate the ground state and used the top 25% most probable configurations for training. We trained using variational Monte Carlo, which samples configurations according to their probabilities. This enables the network to generalize better without needing to be in the overparametrized regime.

References

Carleo, G., & Troyer, M. (2017). Solving the quantum many-body problem with artificial neural networks. *Science, 355*(6325), 602–606. https://doi.org/10.1126/science.aag2302

Horn, D., Weinstein, M., & Yankielowicz, S. (1979). Hamiltonian approach to Z(N) lattice gauge theories. *Physical Review D, 19*(12), 3715–3731.

Moss, M. S., Orfi, A., Roth, C., Sengupta, A. M., Georges, A., Sels, D., Dawid, A., & Valenti, A. (2026). Double descent: When do neural quantum states generalize? *Physical Review E, 113*(4).

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2019). *Deep double descent: Where bigger models and more data hurt*. arXiv.

Vicentini, F., Hofmann, D., Szabó, A., Wu, D., Roth, C., Giuliani, C., Pescia, G., Nys, J., Vargas-Calderón, V., Astrakhantsev, N., & Carleo, G. (2021). *NetKet 3: Machine learning toolbox for many-body quantum systems*. arXiv.