Complex Restricted Boltzmann Machines for Ground State Estimation

Adrien Kahn Reda Nabil Simon Dupouy

April 6, 2023

The quantum many-body problem

We are interested in studying a **Quantum Many-Body Problem** with **non-negligible correlations**.

A general framework would be **N interacting particles** with a certain **degree of freedom K**. Usually the **spin** (**dim 2** $\uparrow\downarrow$).

Our aim is to efficiently **calculate the ground-state** and **simulate the quantum dynamics** of a given Hamiltonian in this space of exponential dimension 2^n.

Reminder on notations

Bras and kets

- In quantum mechanics, a quantum state is an element of $\left(\mathbb{C}^2\right)^{\otimes n}$. It is denoted by a ket: $|\psi\rangle \in \left(\mathbb{C}^2\right)^{\otimes n}$. It has dimension $2^n \times 1$.
- We also use the notation bra to denote the adjoint of the state $\langle\psi| = (|\psi\rangle)^\dagger$. It has dimension 1×2^n.

Hamiltonian

Hamiltonian

- The Hamiltonian H is a $2^n \times 2^n$ complex matrix that defines the system through Schrödinger's equation:

$$i\hbar \frac{\partial}{\partial t} |\psi(t)\rangle = H |\psi(t)\rangle$$

- The energy of a state $|\psi\rangle$ is given by:

$$\frac{\langle\psi| H |\psi\rangle}{\langle\psi|\psi\rangle}$$

- It is hermitian, its real eigenvalues are the energies.
- In physics, we are interested in its ground state (eigenstate of lowest eigenvalue).

Local Hamiltonians

Local Hamiltonians

- A local Hamiltonian is a Hamiltonian that is a sum of a polynomial number of local terms:

$$\sum_{k=1}^{m} \alpha_k p_k$$

  with $m = \text{poly}(n)$.
- The terms p_k are products of n Pauli matrices, with only $O(1)$ of them not being the identity.

For example $Z_1 Z_2$ or X_{n-3} are local terms.

The curse of dimensionality

In order to fully describe a state $|\psi\rangle$ we need to know its decomposition in the Hilbert space of all possible spin configurations.

Let us notice that:

$$\begin{aligned}|\psi\rangle &= \sum_s |s\rangle\langle s|\psi\rangle \text{ as } \sum_s |s\rangle\langle s| = 1 \\ &= \sum_s \psi(s)\,|s\rangle\end{aligned}$$

$\psi(S)$ is thus the component of the projection of $|\psi\rangle$ on $|S\rangle$. It is a complex number with an amplitude and a phase than need to be calculated. Which means $2^{n+1} - 2$ real numbers to identify.

The Restricted Boltzmann Machine

Usual optimisation methods are not adapted to such dimensions, exponential in N. The limit is around $N \sim 10$.

Neural-Networks Quantum States (NQS) drastically lowers the dimension. It becomes polynomial in N.

It is based on the principle of the Restricted Boltzmann Machines (RBM). It allows us to extract the components $\psi(S)$ that interest us from the weights θ obtained by the RBM. θ is updated in order to make the complex amplitudes $\psi(S)$ converges towards the G-S.

A similar approach considering $\dot{\theta}(t)$ allows us to approximate the dynamical evolution of ψ.

Symmetries

The RBM uses a set of M hidden neurons. M generally scales with the system's size: $M = \alpha N$. Thus there are roughly $M \times N = \alpha N^2$ parameters to optimise.

The dimension can be further reduced by symmetry considerations.

For example, in a translation-invariant system, all the spins are invariant. This allows another drastic reduction of the number of parameters by a factor N. We are left with αN parameters which is a linear dependency in the system's size.

Problems successfully tackled by NQS

NQS has been successful in solving quantum many-body problems such as the quantum Ising model and the Hubbard model.

It accurately predicts the phase diagram of the quantum Ising model for various lattice geometries and magnetic field strengths.

It provides new insights into the ground-state properties of the Hubbard model for strongly correlated electrons, including the Mott transition.

Restricted Boltzmann Machines (RBM)

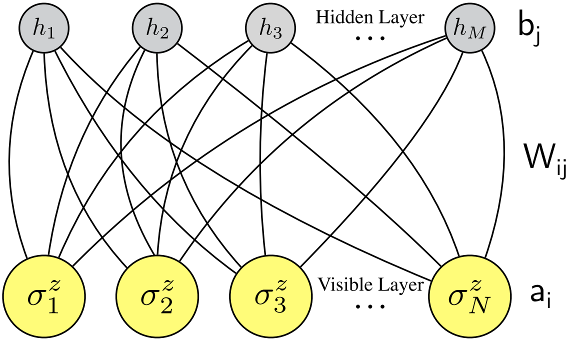


Figure: A Restricted Boltzmann Machine encoding a many-body quantum state of N spins. Spin configuration $S = (\sigma_1^z, \sigma_2^z, \ldots, \sigma_N^z)$, parameters $\theta = \{a_i, b_j, W_{ij}\}$.

Restricted Boltzmann Machines (RBM)

We define the energy of a configuration S, h for a set of parameters W:

$$E(S, h; \theta) = -\sum_j a_j \sigma_j^z - \sum_i b_i h_i - \sum_{i,j} W_{ij} h_i \sigma_j^z$$

The wave function on the full Hilbert space of visible and hidden spins is then:

$$\psi_M(S, h; \theta) = e^{-E(S,h;\theta)}$$

Restricted Boltzmann Machines (RBM)

$$E(S, h; \theta) = -\sum_j a_j \sigma_j^z - \sum_i b_i h_i - \sum_{i,j} W_{ij} h_i \sigma_j^z$$

Since there are no intra-layer connections, the hidden variables can be traced out, yielding the wave function ψ_M:

$$\begin{aligned}\psi_M(S; \theta) &= \sum_{\{h_i\}} e^{-E(S,h;\theta)} \\ &= e^{\sum_i a_i \sigma_i^z} \times \prod_{i=1}^{M} F_i(S; \theta)\end{aligned}$$

With $F_i(S; \theta) = 2\cosh[b_i + \sum_j W_{ij} \sigma_j^z]$.

Reinforcement learning approach

We want to find the ground state of a Hamiltonian H. It is the lowest-energy solution of the Schrödinger equation $H|\psi\rangle = E|\psi\rangle$.

As we do not have access to samples drawn according to the exact wave function, supervised learning is not possible. Hence, a reinforcement learning approach was used.

The loss function to be minimized is the energy:

$$E(\theta) = \frac{\langle\psi_\theta| H |\psi_\theta\rangle}{\langle\psi_\theta|\psi_\theta\rangle}$$

Measuring observables

In order to estimate the energy gradient, we must first be able to estimate the expectation value of observables at a given state $|\psi_\theta\rangle$:

$$\frac{\langle\psi_\theta|\,\hat{O}\,|\psi_\theta\rangle}{\langle\psi_\theta|\psi_\theta\rangle} = \frac{\sum_s \langle\psi_\theta|s\rangle\langle s|\,\hat{O}\,|\psi_\theta\rangle}{\sum_{s'} \langle\psi_\theta|s\rangle\langle s|\psi_\theta\rangle} = \frac{\sum_s \langle\psi_\theta|s\rangle\langle s|\psi_\theta\rangle \frac{\langle s|\hat{O}|\psi_\theta\rangle}{\langle s|\psi_\theta\rangle}}{\sum_{s'} |\langle\psi_\theta|s\rangle|^2}$$

$$= \sum_s \frac{|\langle\psi_\theta|s\rangle|^2}{\sum_{s'} |\langle\psi_\theta|s\rangle|^2} \frac{\langle s|\,\hat{O}\,|\psi_\theta\rangle}{\langle s|\psi_\theta\rangle} = \sum_s p(s) O_{\mathrm{loc}}(s)$$

Computing observables

The average of $\hat{O}$ over ψ_θ can be expressed as:

$$\frac{\langle \psi_\theta | \hat{O} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle} = \sum_s p(s) O_{\text{loc}}(s) = \langle O_{\text{loc}} \rangle$$

with the classical function $O_{\text{loc}}(s)$ and the distribution $p(s)$ given by:

$$O_{\text{loc}}(s) = \frac{\langle s | \hat{O} | \psi_\theta \rangle}{\langle s | \psi_\theta \rangle} \qquad p(s) = \frac{|\langle \psi_\theta | s \rangle|^2}{\sum_{s'} |\langle \psi_\theta | s \rangle|^2}$$

Hence we can estimate $\langle \hat{O} \rangle$ with a classical Monte-Carlo sampling of the classical spin space with the distribution $p(s)$.

Computing O_{loc} is reasonable for local observables

How to compute O_{loc} ?

$$O_{\text{loc}}(s) = \frac{\langle s| \hat{O} |\psi_\theta\rangle}{\langle s|\psi_\theta\rangle}$$
$$= \sum_{s'} \frac{\langle s| \hat{O} |s'\rangle \langle s'|\psi_\theta\rangle}{\langle s|\psi_\theta\rangle}$$

Because $\hat{O}$ is a sum of terms that act non-trivially only on a few spins, $\langle s| \hat{O} |s'\rangle$ is non-zero for a polynomial number of s'.

Imaginary time evolution

Real and imaginary time evolution

Real time evolution is defined as: $e^{-iHt} |\psi\rangle$
Imaginary time evolution is defined by substituting $t \to -i\tau$:

$$e^{-H\tau} |\psi\rangle$$

Effect of imaginary time evolution on a random state

$$e^{-H\tau} |\psi\rangle = \sum_k c_k e^{-E_k \tau} |\psi_k\rangle$$

The renormalized imaginary time evolved wavefunction converges exponentially fast towards the ground state!

Imaginary time evolution of a parametrized state

Evolution of $|\psi_\theta\rangle$ with small imaginary time δ

$$|\psi'\rangle = |\psi_\theta\rangle - \delta H |\psi_\theta\rangle + O(\delta^2)$$

How to find $\tilde{\theta}$ such that $|\psi_{\tilde{\theta}}\rangle \approx |\psi'\rangle$?

Expansion of $|\psi_\theta\rangle$ around θ

We can consider $\tilde{\theta} = \theta + \delta\dot{\theta}$. Then:

$$|\psi_{\tilde{\theta}}\rangle = |\psi_\theta\rangle + \delta \sum_p \dot{\theta}_p \partial_{\theta_p} |\psi_\theta\rangle + O(\delta^2)$$

Computing $\tilde{\theta}$

Many possible choices of metric to optimize. We take the infidelity.

Definition

The infidelity between $|\psi_{\tilde{\theta}}\rangle$ and $|\psi'\rangle$ is:

$$1 - \frac{\left|\langle\psi_{\tilde{\theta}}|\psi'\rangle\right|^2}{\langle\psi_{\tilde{\theta}}|\psi_{\tilde{\theta}}\rangle\langle\psi'|\psi'\rangle}$$

At the cost of great hardship, the fearless may undertake the minimization of this quantity over $\dot{\theta}$ at leading order in δ. (Hint: leading order is not 1...).

Update rule for θ

The optimal $\dot{\theta}$ is solution of:

$$\sum_{p'} \mathrm{Re}\left(S_{pp'}\right) \dot{\theta}_{p'} = -\mathrm{Re}\left(f_p\right)$$

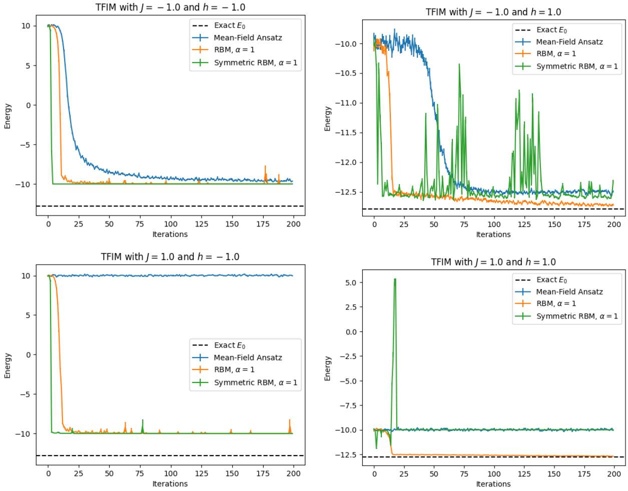
with

$$\begin{cases} S_{pp'} & = \langle O_p^* O_{p'} \rangle - \langle O_p^* \rangle \langle O_{p'} \rangle \\ f_p & = \langle O_p^* E_{\mathrm{loc}} \rangle - \langle O_p^* \rangle \langle E_{\mathrm{loc}} \rangle \end{cases}$$

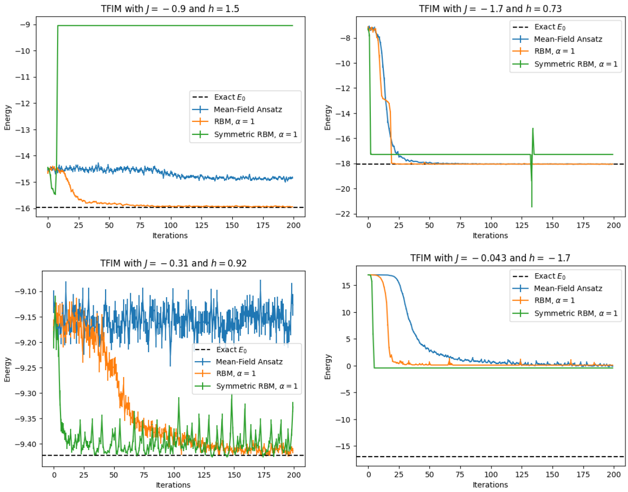
and the classical function:

$$O_p(s) = \partial_{\theta_p} \log \langle s | \psi_\theta \rangle$$

Results


TFIM with $J = -1.0$ and $h = -1.0$
TFIM with $J = -1.0$ and $h = 1.0$
TFIM with $J = 1.0$ and $h = -1.0$
TFIM with $J = 1.0$ and $h = 1.0$
Exact E_0
Mean-Field Ansatz
RBM, $\alpha = 1$
Symmetric RBM, $\alpha = 1$
Energy
Iterations

Results


TFIM with $J = -0.9$ and $h = 1.5$
TFIM with $J = -1.7$ and $h = 0.73$
TFIM with $J = -0.31$ and $h = 0.92$
TFIM with $J = -0.043$ and $h = -1.7$
Energy
Iterations
Exact E_0
Mean-Field Ansatz
RBM, $\alpha = 1$
Symmetric RBM, $\alpha = 1$

Conclusion

- Neural-Network Quantum States **capture efficiently the complexity** of some many-body systems in 1D and 2D, featuring non-local correlations.
- This method allows to estimate the **ground state** of a local Hamiltonian using imaginary time evolution. It can also be used to perform real time evolution.
- In both cases, the authors **claim state of the art accuracy**.
- The **accuracy can be improved** by increasing the size M of the hidden layer.
- Perspective to use **more complex architectures** such as Deep Neural Networks (DNN).