



# Variational Quantum Many-Body Wave Functions Based on Restricted Boltzmann Machines Neural Networks

---

Master Thesis  
submitted to the Faculty of the  
Escola Tècnica d'Enginyeria de Telecomunicació de Barcelona  
Universitat Politècnica de Catalunya  
by  
Marc Fernández Martínez

In partial fulfillment  
of the requirements for the master in  
**Engineering Physics**

Advisor: Ferran Mazzanti, Enrique Romero  
Barcelona, Date: June 2024



# Contents

<b>List of Figures</b>	<b>4</b>
<b>List of Tables</b>	<b>4</b>
<b>1 Introduction</b>	<b>6</b>
1.1 Objectives . . . . .	6
1.2 Requirements and Specifications . . . . .	6
1.3 Methods and Procedures . . . . .	7
1.4 Gantt Diagram . . . . .	7
1.5 Deviations and Incidences . . . . .	8
<b>2 State of the art</b>	<b>10</b>
2.1 RBMs and DBMs . . . . .	10
2.2 Tensor Networks . . . . .	10
<b>3 Theoretical Background</b>	<b>11</b>
3.1 Transverse-field Ising model . . . . .	11
3.2 Pauli spin operator algebra . . . . .	11
3.3 Jordan-Wigner transformation . . . . .	13
3.4 Uniform Transverse-field Ising model . . . . .	15
3.4.1 Bogoliubov transformation . . . . .	17
3.4.2 Ground state and excited states . . . . .	18
3.4.3 Summary of theoretical results . . . . .	21
3.5 Magnetization of the ground state . . . . .	21
<b>4 Methodology</b>	<b>24</b>
4.1 Variational method . . . . .	24
4.2 Restricted Boltzmann Machines . . . . .	25
4.3 Optimization of RBM parameters . . . . .	26
4.3.1 Simplest case: $N = 2$ spins. . . . .	26
4.3.2 Next case: $N = 4$ spins. . . . .	30
4.4 Limits of the Simulated Annealing scheme . . . . .	33
4.5 Natural Gradient Learning . . . . .	34

---

4.5.1	Natural Gradient . . . . .	34
4.5.2	Stochastic Reconfiguration Learning . . . . .	35
<b>5</b>	<b>Results</b>	<b>42</b>
5.1	Varying number of spins and hidden units. . . . .	42
5.2	Ground state comparison . . . . .	43
5.2.1	Ground state energies and magnetizations . . . . .	43
5.2.2	Ground state distances and overlaps . . . . .	46
<b>6</b>	<b>Budget</b>	<b>51</b>
<b>7</b>	<b>Environment Impact</b>	<b>52</b>
<b>8</b>	<b>Conclusions and future development</b>	<b>53</b>
8.1	Conclusions . . . . .	53
8.2	Future Work . . . . .	54
	<b>References</b>	<b>55</b>

## List of Figures

1	Project's Gantt diagram . . . . .	8
2	Ground state energies for even number of lattice sites. . . . .	20
2	Ground state energies for even number of lattice sites. (cont.) . . . . .	20
3	Ground state magnetizations for (even) increasing lattice sites. . . . .	23
4	Ground state magnetic susceptibilities. . . . .	23
5	Diagram of an RBM. . . . .	25
6	Stochastic reconfiguration plots for a trial wavefunction with $N = 12$ spins and varying number of hidden units. The values in the plot are averages of 10 iterations.	43
7	Relative error between exact and RBM ground state energies. $J = 2, B = 1$ . . .	43
8	Relative error of the longitudinal magnetization $m_z$ . $J = 2, B = 1$ . . . . .	44
9	Relative error of the transversal magnetization $m_x$ . $J = 2, B = 1$ . . . . .	44
10	Transversal magnetic susceptibility error between exact and RBM ground states. $J = 2, B = 1$ . . . . .	44
11	RBM ground state magnetizations with respect to $B/J$ . $N_h = 4N$ in all cases. .	45
12	Energy difference between RBM and exact ground state with respect to $B/J$ . . .	45
13	Magnetic susceptibility. . . . .	46
14	Fubini-Study distance between exact and RBM ground states. . . . .	48
15	Relative error between RBM and first excited state. $J = 2, B = 1$ . . . . .	49
16	Relative error between RBM and second excited state. $J = 2, B = 1$ . . . . .	50

## List of Tables

1	Minimal energies found using divide and conquer strategy. $N = 2$ spins, $N_h = 1$ .	29
2	Minimal energies found using divide and conquer strategy. $N = 4$ spins, with 0 and 1 hidden units. . . . .	30
3	Minimal energies found using simulated annealing. $N = 4$ spins. . . . .	33
4	Minimal energies found using simulated annealing. $N = 5, 10, 12$ spins. . . . .	33
5	Minimal energies found using stochastic reconfiguration. $N = 5, 10, 12$ spins. . .	42

---

## Abstract

This thesis explores the application of variational wave functions based on Restricted Boltzmann Machines (RBMs) to solve spin quantum many-body problems, for instance, the quantum transverse field Ising model.

RBMs, a class of artificial neural networks, can be employed as an efficient and flexible ansatz for representing complex quantum states. By making use of the variational principle, the ground state energy and corresponding wave function of the system are approximated through the optimization of the RBM parameters. The effectiveness of this approach is demonstrated through a comparison with exact diagonalization results, highlighting the RBM's capability to capture critical phenomena and phase transitions inherent in the model.

This work highlights the potential of machine learning techniques in quantum physics, providing a new perspective on solving quantum many-body problems and contributing to the broader effort of integrating artificial intelligence with quantum computing.

# 1 Introduction

The quantum many-body problem poses a challenge in quantum physics involving the difficulty in predicting the properties and behaviour of systems with a large number of interacting particles. Traditional methods for solving these problems, such as exact diagonalization, often become infeasible as the system size increases, due to the exponential growth of the Hilbert space and the complexity of the many-body wavefunction. In recent years, artificial neural networks have been used as a powerful tool to approximate complex quantum states efficiently.

This thesis focuses on employing a variational quantum many-body wavefunction based on Restricted Boltzmann Machines to solve the one-dimensional (1D) quantum transverse field Ising model (TFI).

## 1.1 Objectives

The primary objectives of this research are:

- Solve the TFI model analytically, and obtain some exact results.
- To develop an RBM-based variational approach for the TFI.
- Optimizing the RBM parameters using variational principles, and machine learning methods.
- Validate the results by comparing them with the exact diagonalization results.
- To analyze the scalability and efficiency of the RBM approach for larger, or more complex systems.

## 1.2 Requirements and Specifications

To conduct this research project, a combination of theoretical knowledge and computational resources were needed:

1. A solid understanding of quantum many-body physics, specifically the transverse field Ising model.
2. Familiarity with machine learning, particularly restricted Boltzmann machines and variational Monte Carlo (VMC) methods.
3. Python was chosen for its extensive libraries suited for scientific computing and machine learning.
4. For hardware, a high-performance personal computer with a multi-core processor was necessary to handle the computational load, though having access to a High-Performance Computing cluster would have been beneficial for larger-scale simulations.

### 1.3 Methods and Procedures

The project began with an extensive literature review to understand the research in using RBMs to solve quantum many-body problems. This included studying the foundational work of Carleo and Troyer's "Solving the quantum many-body problem with artificial neural networks" (2017) [9]. The algorithms and software developed by [24] were adapted and extended to fit the specific requirements of solving the TFI model using RBMs.

Throughout the project, detailed documentation was maintained, including recording the methods, procedures and results that were compiled into this thesis document.

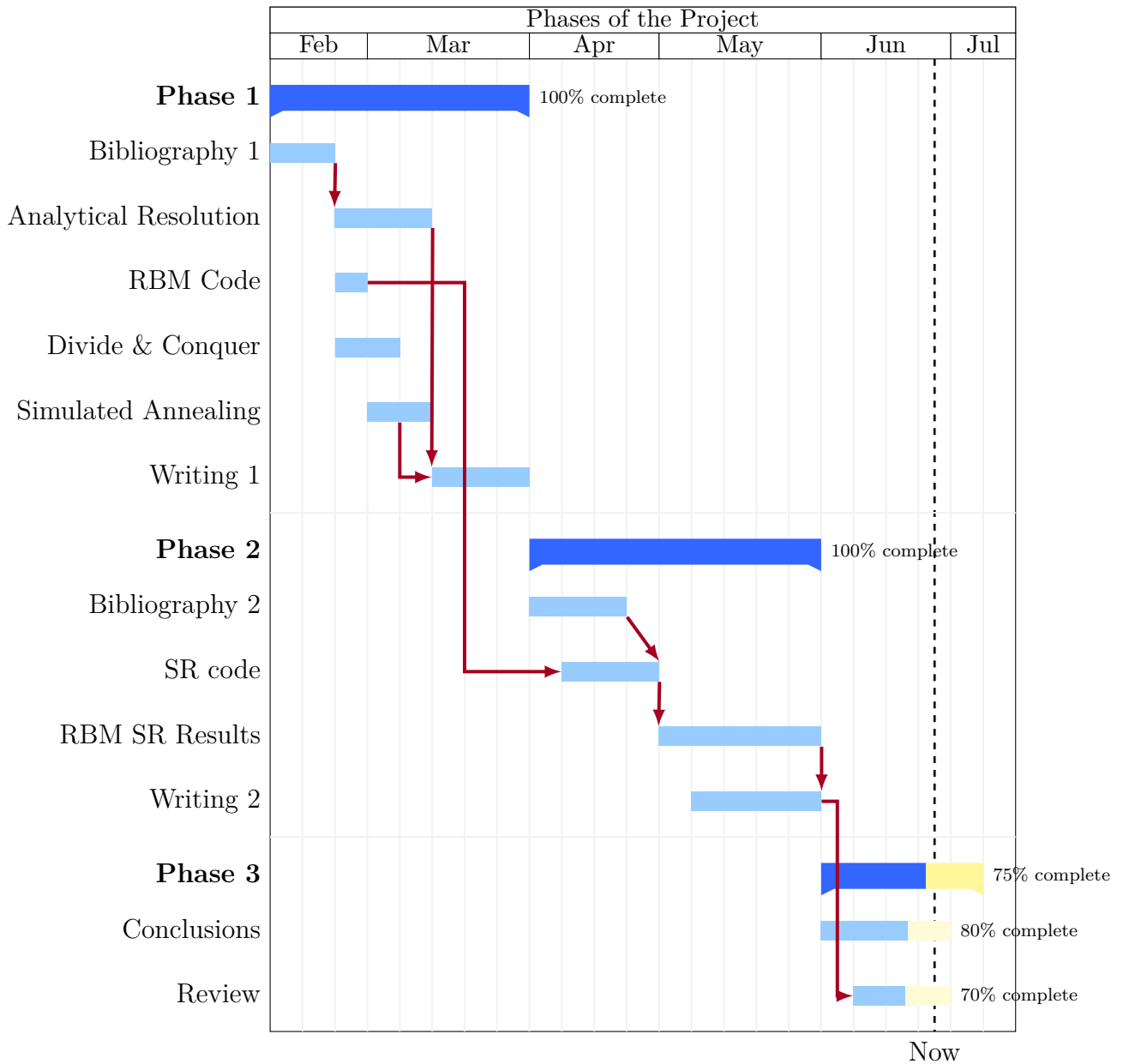
### 1.4 Gantt Diagram

The project was structured in 3 main phases, as shown in the Gantt Diagram (Fig.1). Each phase involved an initial bibliography research and lecture, and a final review and writing task of the main findings and results of the corresponding phase.

The first phase involved the two first objectives of the thesis. That is, solving the TFI model analytically and obtaining results for the RBM variational approach for small systems. Most of the programming work was done in parallel with the analytical resolution of the problem.

The next phase was the longest one, and it involved the core objective of the research. Optimizing the RBM parameters with stochastic reconfiguration and comparing the results with the exact results obtained in Phase 1 took almost two months. However, this step was crucial to ensure the reliability of the developed method.

The final phase consisted in reviewing all the results and documentation, repeating any experiments if needed, and writing this document.



**Figure 1:** Gantt diagram of the project

## 1.5 Deviations and Incidences

In this section, we discuss the deviations and incidences encountered during this research project, and how they were addressed.

### Deviations

1. **Complexity of analytical solution:** Initially, the plan was to derive and use analytical solutions for the magnetizations and susceptibility of the TFI ground state. However, the



complexity of obtaining explicit expressions from the diagonalized Hamiltonian led to the approach of calculating them through the diagonalization of the Hamiltonian matrix instead. This deviation meant that we could not obtain exact results for larger system sizes, but it was necessary to keep the project within scope and on schedule.

2. **Computational limitations:** The original plan underestimated the computational resources required for optimizing RBM parameters for larger systems. This implied that each experiment required additional time, and the result was to adjust the scope of the project by leaving the scalability and efficiency analysis for future work.
3. **Change in algorithms:** While the initial approach relied on simulated annealing for optimizing the RBM parameters, it became clear that a more suited technique for the problem like stochastic reconfiguration was needed for the experiments.

## Incidences

1. **Software issues:** During the development of the project, several issues were encountered with the programming languages that caused compatibility issues, bugs or numerical precision in the results. These incidences required troubleshooting and eventually switching between different programming languages: from Matlab to Julia, and finally Python.
2. **Resource management:** Handling large variables generated from large system simulations presented some challenges, as these variables ate away at the virtual memory of the computer, often leading to slow processing speeds and crashes. After implementing more efficient techniques and parallelization of the code, the issues were resolved.
3. **Overflow errors:** Validating the results obtained from the RBM wavefunction against the exact ones took longer than expected due to the high precision required and the overflow errors that arose in large system sizes. The large number of parameters used in these cases, combined with the expression of the wavefunction, made the calculations needed for the optimization process overflow and cause huge errors. Eventually, a numerically robust implementation of the functions solved the issue, causing some delays in the experimentation phase of the project.

To conclude this section, some lessons learned and mitigation strategies are provided to offer valuable insights for future projects.

First, being adaptable to changes and being ready to modify the approach in response to challenges was crucial for the project's success. Ensuring access to adequate computational resources from the start could have mitigated some delays, and allowed larger-scale simulations to be done. In addition, testing the RBM models incrementally helped identify and address issues early.

## 2 State of the art

The use of artificial neural networks (ANNs) to solve (spin) quantum many-body problems has significantly advanced since the publication of the Science article "Solving the quantum many-body problem with artificial neural networks" by Carleo and Troyer in 2017 [9]. Research in this field has seen substantial developments, particularly in the application of Variational Monte Carlo (VMC) methods combined with different architectures of neural networks.

### 2.1 RBMs and DBMs

Recent studies have explored the capabilities of both Restricted Boltzmann Machines (RBMs) and Deep Boltzmann Machines (DBMs) for representing quantum states [16]. The reason these models are advantageous is because they can serve as universal approximators for quantum states, meaning that they can represent any quantum state with arbitrary accuracy given enough hidden units [19].

### 2.2 Tensor Networks

Moreover, the research has expanded to compare neural network methods with other techniques such as tensor networks [20]. While neural networks efficiently represent quantum states with volume law entanglement entropy, there are examples in which tensor networks have efficient representations with fewer parameters [21] [22]. However, until recently, tensor networks could only tackle quantum systems in spatial dimensions higher than 1 by discretizing them [25], which can be an inadequate representation for some many-body problems. The variational approach, on the other hand, does not have this kind of problem.

Overall, the current state of the art involves making use of the representational power of neural networks and enhancing VMC methods with new architectures. In any case, there is a lot of synergy potential between neural networks and other quantum state approximation methods to explore.

### 3 Theoretical Background

As we mentioned in the introduction, the aim of this project is to explore the possibility of using variational wavefunctions of the RBM neural network form, and try to estimate the ground state of spin quantum many-body problems with them.

It would be interesting to show the effectiveness of this idea by comparing the results with the exact diagonalization. However, not many quantum mechanical problems can be solved analytically. So, in this thesis, we focus on solving the *transverse-field Ising model* in 1D, which can be solved analytically in 1D, as we will see in this section.

#### 3.1 Transverse-field Ising model

The **transverse-field Ising model** (TFI) [5] is a quantum version of the classical *Ising model*. It is characterized by having a lattice with nearest neighbours interactions determined by the alignment or anti-alignment of spin projections along the  $z$  axis. In addition, it also features an external magnetic field perpendicular to the  $z$  axis, which creates an energetic bias for this spin direction over the other.

The main quantum mechanical feature is that, since the spin operators along the  $z$  and the perpendicular axis do not commute, they cannot be observed simultaneously. As such, a quantum treatment is required.

The model consists of  $N$  spin- $1/2$ , distributed along  $L$  different lattice sites, and described by the Hamiltonian:

$$\hat{H} = - \sum_{j=1}^L J_j \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z - \sum_{j=1}^L (B_j^x \hat{\sigma}_j^x + B_j^y \hat{\sigma}_j^y) \quad (3.1)$$

We can choose, without loss of generality, that the perpendicular field goes along the  $x$  axis. Then, Equation 3.1 becomes

$$\hat{H} = - \sum_{j=1}^L J_j \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z - \sum_{j=1}^L B_j \hat{\sigma}_j^x \quad (3.2)$$

where  $\hat{\sigma}^\alpha$ ,  $\alpha = \{x, y, z\}$  are the Pauli matrices representing the elements of our spin algebra.

#### 3.2 Pauli spin operator algebra

Consider the Pauli spin matrices:

$$\hat{\sigma}^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

Notice that all three fulfill the involution relation

$$(\hat{\sigma}^x)^2 = (\hat{\sigma}^y)^2 = (\hat{\sigma}^z)^2 = \mathbb{1},$$

and the commutation relations

$$\begin{aligned}\left[\hat{\sigma}^\alpha, \hat{\sigma}^\beta\right] &= 2i\epsilon_{\alpha\beta\gamma}\hat{\sigma}^\gamma \\ \left\{\hat{\sigma}^\alpha, \hat{\sigma}^\beta\right\} &= 2\delta_{\alpha\beta}\mathbb{1}\end{aligned}\tag{3.3}$$

with all cyclic permutations. In particular, they verify  $[\hat{\sigma}^x, \hat{\sigma}^y] = 2i\hat{\sigma}^z$ .

We identify the spin states  $|\uparrow\rangle = (1, 0)$  and  $|\downarrow\rangle = (0, 1)$ , for which we will define the basis  $\mathcal{S} \equiv \{|\uparrow\rangle, |\downarrow\rangle\}$ . The action of the Pauli operators on these spins are

$$\begin{aligned}\hat{\sigma}^z |\uparrow\rangle &= |\uparrow\rangle, & \hat{\sigma}^z |\downarrow\rangle &= -|\downarrow\rangle \\ \hat{\sigma}^x |\uparrow\rangle &= |\downarrow\rangle, & \hat{\sigma}^x |\downarrow\rangle &= |\uparrow\rangle\end{aligned}$$

We have seen that the Hilbert space of a single spin is two-dimensional, hence the basis being  $\mathcal{S}$ . However, we can denote  $\hat{\sigma}_j^\alpha$  the Pauli matrix at different lattice sites  $j$ , therefore acting on *different* (distinguishable) Hilbert spaces. Then, they must fulfill

$$\left[\hat{\sigma}_i^\alpha, \hat{\sigma}_j^\beta\right] = 0 \quad \text{for } i \neq j.\tag{3.4}$$

But on the *same* site, they lead to the commutation rules in Equation 3.3.

---

**Definition 3.1** (Raising and lowering operators [3]). We define the raising and lowering operators acting on lattice site  $j$

$$\hat{\sigma}_j^+ = \frac{(\hat{\sigma}_j^x + i\hat{\sigma}_j^y)}{2}, \quad \hat{\sigma}_j^- = \frac{(\hat{\sigma}_j^x - i\hat{\sigma}_j^y)}{2}\tag{3.5}$$

which act on the basis states as  $\hat{\sigma}^+ |\downarrow\rangle = |\uparrow\rangle$  and  $\hat{\sigma}^- |\uparrow\rangle = |\downarrow\rangle$ .

They satisfy the *anti-commutator* relation

$$\left\{\hat{\sigma}^+, \hat{\sigma}^-\right\} = 1\tag{3.6}$$

which is typical of fermionic commutation rules.

---

Using bosons to describe spin systems would seem impossible, since we could always construct an infinite dimensional Hilbert space in second quantization with all states

$$|n\rangle = \frac{1}{\sqrt{n!}}(\hat{b}^\dagger)^n |0\rangle \quad \text{where } n = 0, 1, \dots, \infty.\tag{3.7}$$

where  $\hat{b}^\dagger$  represents a single boson associated with the vacuum state  $|0\rangle$  ( $\hat{b}|0\rangle = 0$ ).

Then, in order to solve the eigenvalue problem for the Hamiltonian in Equation 3.2, we need to add the constraint that at most one spin is allowed on each site. In the next section, we introduce the **Jordan-Wigner transformation**, which allows us to represent this constraint in terms of *spinless fermions*.

### 3.3 Jordan-Wigner transformation

As we just mentioned, to solve the TFI model problem, we want to truncate this Hilbert space to 2 states only  $\{|0\rangle, |1\rangle\}$ , assuming  $(\hat{b}^\dagger)^2 |0\rangle = 0$ . With this, we arrive at the Hilbert space of a single spin-1/2. Physically, this can be thought as adding an infinite on-site repulsion term to the boson Hamiltonian. This truncation, known as the *hard-core boson*, is done by transforming the Pauli spin-1/2 operators into hard-core bosons  $\hat{b}_j^\dagger$ , with the identification  $|0\rangle \leftrightarrow |\downarrow\rangle$  and  $|1\rangle = \hat{b}^\dagger |0\rangle \leftrightarrow |\uparrow\rangle$ :

$$\begin{cases} \hat{\sigma}_j^+ &= \hat{b}_j^\dagger \\ \hat{\sigma}_j^- &= \hat{b}_j \\ \hat{\sigma}_j^z &= 2\hat{b}_j^\dagger \hat{b}_j - 1 \end{cases} \implies \begin{cases} \hat{\sigma}_j^x &= \hat{b}_j^\dagger + \hat{b}_j \\ \hat{\sigma}_j^y &= i(\hat{b}_j - \hat{b}_j^\dagger) \\ \hat{\sigma}_j^z &= 2\hat{b}_j^\dagger \hat{b}_j - 1 \end{cases} \quad (3.8)$$

Notice that these operators commute at different sites (as  $\hat{\sigma}_j^\alpha$ ) but anti-commute on the same site  $\{\hat{b}_j^\dagger, \hat{b}_j\} = 1$ , this mix of (anti-)commutation relations indicate that these operators are not exactly bosonic. In fact, we can check that the commutation relation for bosonic creation/annihilation operators holds for  $\hat{b}_j$ :

$$[\hat{b}_j^\dagger, \hat{b}_j] = \hat{b}_j^\dagger \hat{b}_j - \hat{b}_j \hat{b}_j^\dagger = \hat{b}_j^\dagger \hat{b}_j - (1 - \hat{b}_j^\dagger \hat{b}_j) = 2\hat{b}_j^\dagger \hat{b}_j - 1 = \hat{\sigma}_j^z \quad (3.9)$$

where we used the on-site anticommutation relation. From Equation 3.9, we can see that  $\hat{b}$  are not ordinary bosons, since their commutator should be 1 on the same site. On the other hand, they verify the hard-core constraint (at most one spin per site) by definition.

This hard-core constraint can be represented in terms of *spinless fermions*  $\hat{c}_j^\dagger$ , where the double occupancy constraint is automatically enforced by the *Pauli exclusion principle*, and the operators anti-commute on the same site.

Unfortunately, writing  $\hat{b}_j^\dagger$  in terms of spinless fermions  $\hat{c}_j^\dagger$  is only useful in one-dimension, due to the fact that a natural ordering of the sites is only possible in (1D).

**Definition 3.2** (Jordan-Wigner transformation). The *Jordan-Wigner transformation* of hard-core bosons into fermions is defined as

$$\hat{b}_j = \hat{K}_j \hat{c}_j = \hat{c}_j \hat{K}_j \quad \text{with} \quad \hat{K}_j = e^{i\pi \sum_{j'=1}^{j-1} \hat{n}_{j'}} = \prod_{j'=1}^{j-1} (1 - 2\hat{n}_{j'}) \quad (3.10)$$

where the *non-local* string operator  $\hat{K}_j$  is simply a sign, counting the parity of the number of fermions between 1 and  $j - 1$ . Notice that  $\hat{K}_j = \hat{K}_j^\dagger = \hat{K}_j^{-1}$ , and  $\hat{K}_j^2 = 1$ .

We denote  $\hat{n}_j = \hat{c}_j^\dagger \hat{c}_j$  as the fermion number operator. One can show that the phase-factor  $\hat{K}_j$  cancels out in  $\hat{n}_j$ , i.e.,  $\hat{n}_j = \hat{c}_j^\dagger \hat{c}_j = \hat{b}_j^\dagger \hat{b}_j$ .

If  $\hat{c}_j$  are taken to be standard fermionic operators, with *anti-commutation* relations  $\{\hat{c}_j, \hat{c}_{j'}^\dagger\} = \delta_{j,j'}$  and  $\{\hat{c}_j, \hat{c}_{j'}\} = \{\hat{c}_j^\dagger, \hat{c}_{j'}^\dagger\} = 0$ , then we can use these relations to derive some of the  $\hat{b}_j$  properties:

**Proposition 3.3** (Hard-core boson properties).

$$\mathbf{P1:} \begin{cases} \{\hat{b}_j, \hat{b}_j^\dagger\} = 1 \\ \{\hat{b}_j, \hat{b}_j\} = 0 \\ \{\hat{b}_j^\dagger, \hat{b}_j^\dagger\} = 0 \end{cases} \quad \mathbf{P2:} \begin{cases} [\hat{b}_j, \hat{b}_{j'}^\dagger] = 0 \\ [\hat{b}_j, \hat{b}_{j'}] = 0 \quad \text{if } j \neq j' \\ [\hat{b}_j^\dagger, \hat{b}_{j'}^\dagger] = 0 \end{cases} \quad (3.11)$$

**Proof:**

Property **P1** is straightforward due to  $\hat{K}_j$  cancelling out in  $\hat{n}_j$ :

$$\hat{b}_j^\dagger \hat{b}_j = \hat{c}_j^\dagger \hat{K}_j^\dagger \hat{K}_j \hat{c}_j = \hat{c}_j^\dagger \hat{c}_j$$

and, similarly,  $\hat{b}_j \hat{b}_j^\dagger = \hat{c}_j \hat{c}_j^\dagger$ . Thus, on each site  $\hat{b}_j$  inherit the anti-commutation properties from the fermions  $\hat{c}_j$ .

To prove property **P2**, let us consider  $[\hat{b}_{j_1}, \hat{b}_{j_2}^\dagger]$ , assuming  $j_2 > j_1$ . Using equation 3.10 we can show that

$$\hat{b}_{j_1} \hat{b}_{j_2}^\dagger = \hat{c}_{j_1} e^{-i\pi \sum_{j=j_1}^{j_2-1} \hat{n}_j} \hat{c}_{j_2}^\dagger \quad (3.12)$$

Similarly, one can find that

$$\begin{aligned} \hat{b}_{j_2}^\dagger \hat{b}_{j_1} &= \hat{c}_{j_2}^\dagger e^{-i\pi \sum_{j=j_1}^{j_2-1} \hat{n}_j} \hat{c}_{j_1} \\ &= -e^{-i\pi \sum_{j=j_1}^{j_2-1} \hat{n}_j} \hat{c}_{j_1} \hat{c}_{j_2}^\dagger = \hat{c}_{j_1} e^{-i\pi \sum_{j=j_1}^{j_2-1} \hat{n}_j} \hat{c}_{j_2}^\dagger \end{aligned} \quad (3.13)$$

The first sign change is due to the fermionic anti-commutation. The second sign change, however, comes from the fact that the only non-zero contribution of  $\hat{n}_{j_1}$  to the left of  $\hat{c}_{j_1}$  is when the latter acts onto state  $|1\rangle$ , resulting in a term  $e^{-i\pi \hat{n}_{j_1}} \hat{c}_{j_1} |1\rangle = e^{-i\pi \hat{n}_{j_1}} |0\rangle \rightarrow e^{-i\pi \cdot 0} = 1$ . Then, in the final expression, the non-zero contribution to the right of  $\hat{c}_{j_1}$  is  $|1\rangle$ , so the previous term becomes  $\hat{c}_{j_1} e^{-i\pi \hat{n}_{j_1}} |1\rangle \rightarrow e^{-i\pi} = -1$ .

Thus, from eqs.3.12-3.13 it follows that

$$[\hat{b}_{j_1}, \hat{b}_{j_2}^\dagger] = \hat{b}_{j_1} \hat{b}_{j_2}^\dagger - \hat{b}_{j_2}^\dagger \hat{b}_{j_1} = \hat{c}_{j_1} e^{-i\pi \sum_{j=j_1}^{j_2-1} \hat{n}_j} \hat{c}_{j_2}^\dagger - \hat{c}_{j_1} e^{-i\pi \sum_{j=j_1}^{j_2-1} \hat{n}_j} \hat{c}_{j_2}^\dagger = 0$$

All the other commutation relations in **P2** are proven similarly.

Here is a summary of useful expressions where the string operator  $\hat{K}_j$  disappears completely:

$$\begin{aligned} \hat{b}_j^\dagger \hat{b}_j &= \hat{c}_j^\dagger \hat{c}_j \\ \hat{b}_j^\dagger \hat{b}_{j+1}^\dagger &= \hat{c}_j^\dagger (1 - 2\hat{n}_j) \hat{c}_{j+1}^\dagger = \hat{c}_j^\dagger \hat{c}_{j+1}^\dagger \\ \hat{b}_j^\dagger \hat{b}_{j+1} &= \hat{c}_j^\dagger (1 - 2\hat{n}_j) \hat{c}_{j+1} = \hat{c}_j^\dagger \hat{c}_{j+1} \\ \hat{b}_j \hat{b}_{j+1} &= \hat{c}_j (1 - 2\hat{n}_j) \hat{c}_{j+1} = \hat{c}_j (1 - 2(1 - \hat{c}_j \hat{c}_j^\dagger)) \hat{c}_{j+1} = -\hat{c}_j \hat{c}_{j+1} \\ \hat{b}_j \hat{b}_{j+1}^\dagger &= \hat{c}_j (1 - 2\hat{n}_j) \hat{c}_{j+1}^\dagger = \hat{c}_j (1 - 2(1 - \hat{c}_j \hat{c}_j^\dagger)) \hat{c}_{j+1}^\dagger = -\hat{c}_j \hat{c}_{j+1}^\dagger \end{aligned} \quad (3.14)$$

Where we have used

$$\prod_{j'=1}^{j-1} (1 - 2\hat{n}_{j'}) \prod_{j'=1}^j (1 - 2\hat{n}_{j'}) = 1 - 2\hat{n}_j. \quad (3.15)$$

since  $(1 - 2\hat{n}_j)^2 = 1$ , and different sites commute.

In our case, we will be assuming periodic boundary conditions (PBC) for the spin operators in Hamiltonian 3.1, which means that  $\hat{\sigma}_{L+1}^\alpha \equiv \hat{\sigma}_1^\alpha$ . We can impose the same PBC for the hard-core bosons, for instance:  $\hat{b}_L^\dagger \hat{b}_{L+1} \equiv \hat{b}_L^\dagger \hat{b}_1$ . However, if we rewrite this condition using spinless fermions:

$$\hat{b}_L^\dagger \hat{b}_1 = e^{i\pi \sum_{j=1}^{L-1} \hat{n}_j} \hat{c}_L^\dagger \hat{c}_1 = -e^{i\pi \sum_{j=1}^L \hat{n}_j} \hat{c}_L^\dagger \hat{c}_1 = -e^{i\pi \hat{N}} \hat{c}_L^\dagger \hat{c}_1 \quad (3.16)$$

where

$$\hat{N} = \sum_{j=1}^L \hat{c}_j^\dagger \hat{c}_j = \sum_{j=1}^L \hat{b}_j^\dagger \hat{b}_j, \quad (3.17)$$

is the total number of particles in the chain. The second equality comes from the fact that, to the left of  $\hat{c}_L^\dagger$ , we have  $\hat{n}_L = 1$  and  $e^{i\pi \hat{n}_L} \equiv -1$ . We can also verify that:

$$\hat{b}_L^\dagger \hat{b}_1 = e^{i\pi \sum_{j=1}^{L-1} \hat{n}_j} \hat{c}_L^\dagger \hat{c}_1 = -e^{i\pi \sum_{j=1}^L \hat{n}_j} \hat{c}_L^\dagger \hat{c}_1 = -e^{i\pi \hat{N}} \hat{c}_L^\dagger \hat{c}_1 \quad (3.18)$$

In other words, the boundary conditions are affected by the *fermion parity*  $e^{i\pi \hat{N}} = (-1)^{\hat{N}}$ . PBC become anti-periodic boundary conditions (ABC) when  $\hat{N}$  is *even*.

### 3.4 Uniform Transverse-field Ising model

Let us study the uniform case, where  $J_j = J$  and  $B_j = B$  constant. We can apply the Jordan-Wigner transformation to the Hamiltonian from Equation 3.2 assuming open boundary conditions (OBC) for now, so as to avoid the fermion parity issue seen in Equation 3.18. Later we will add the corresponding term to obtain the Hamiltonian with PBC.

The expression of the Hamiltonian with OBC after the Jordan-Wigner transformation is

$$\hat{H}_{\text{OBC}} = -J \sum_{j=1}^{L-1} (1 - 2\hat{n}_j)(1 - 2\hat{n}_{j+1}) - B \sum_{j=1}^L \underbrace{\prod_{j'=1}^{j-1} (1 - 2\hat{n}_{j'})}_{\text{Very non-local!}} (\hat{c}_j^\dagger + \hat{c}_j) \quad (3.19)$$

One way of getting rid of the non-local interactions is to rotate the coordinates around the  $y$ -axis such that  $\hat{\sigma}_i^x \rightarrow -\hat{\sigma}_i^z$ , and  $\hat{\sigma}_i^z \rightarrow \hat{\sigma}_i^x$  [6]. Then, the previous Hamiltonian is written as

$$\hat{H}_{\text{OBC}} = -J \sum_{j=1}^{L-1} \hat{\sigma}_j^x \hat{\sigma}_{j+1}^x - B \sum_{j=1}^L \hat{\sigma}_j^z \quad (3.20)$$

$$= -J \sum_{j=1}^{L-1} \prod_{j'=1}^{j-1} (1 - 2\hat{n}_{j'}) (\hat{c}_j^\dagger + \hat{c}_j) \prod_{j''=1}^j (1 - 2\hat{n}_{j''}) (\hat{c}_{j+1}^\dagger + \hat{c}_{j+1}) - B \sum_{j=1}^L (1 - 2\hat{n}_j) \quad (3.21)$$

Here we can use eqs.3.14 to simplify the expression to

$$= -J \sum_{j=1}^{L-1} (\hat{c}_j^\dagger + \hat{c}_j)(1 - 2\hat{n}_j)(\hat{c}_{j+1}^\dagger + \hat{c}_{j+1}) - B \sum_{j=1}^L (1 - 2\hat{n}_j) \quad (3.22)$$

$$\hat{H}_{\text{OBC}} = -J \sum_{j=1}^{L-1} (\hat{c}_j^\dagger \hat{c}_{j+1} + \hat{c}_{j+1}^\dagger \hat{c}_j + \hat{c}_j^\dagger \hat{c}_{j+1}^\dagger + \hat{c}_{j+1} \hat{c}_j) - B \sum_{j=1}^L (1 - 2\hat{n}_j) \quad (3.23)$$

and the Hamiltonian in the PBC case is

$$\hat{H}_{\text{PBC}} = \hat{H}_{\text{OBC}} + (-1)^{\hat{N}} J (\hat{c}_L^\dagger \hat{c}_1^\dagger + \hat{c}_L \hat{c}_1 + \text{H.c.}) \quad (3.24)$$

Notice that, if the number of fermions  $\hat{N}$  takes an odd value, then we effectively have  $\hat{c}_{L+1} \equiv \hat{c}_1$ ; if, instead, we have  $\hat{N}$  even value, then the  $L$ -th bond has an opposite sign ( $\hat{c}_{L+1} \equiv -\hat{c}_1$ ). In addition, there are two terms in the Hamiltonian which do not conserve the fermionic particle number, i.e.,  $\hat{c}_j^\dagger \hat{c}_{j+1}^\dagger$  and its hermitian conjugate.

In order to diagonalize this Hamiltonian, let us now introduce the discrete Fourier transformations of the fermionic operators  $\hat{c}_j$ ,

$$\hat{c}_k = \frac{e^{-i\phi}}{\sqrt{L}} \sum_{j=1}^L e^{-ikj} \hat{c}_j \quad (3.25)$$

and the inverse Fourier transform of the momentum-space operators  $\hat{c}_k$

$$\hat{c}_j = \frac{e^{i\phi}}{\sqrt{L}} \sum_k e^{+ikj} \hat{c}_k \quad (3.26)$$

---

**Observation 3.4** (Effect of boundary conditions). The values of  $k$  that should be used in the transformation shown in Equation 3.25 depend on the fermion parity.

For  $\hat{N}$  odd (or parity  $p = 1$ ) we have  $\hat{c}_{L+1} \equiv \hat{c}_1$ , which implies that  $e^{ikL} = 1$ . Thus, we should choose the  $k$ 's as:

$$p = 1 \Rightarrow \mathcal{K}_{p=1} = \left\{ k = \frac{2n\pi}{L}, \text{ with } n = -\frac{L}{2} + 1, \dots, 0, \dots, \frac{L}{2} \right\} \quad (3.27)$$

For  $\hat{N}$  even (or  $p = 0$ ) we have  $\hat{c}_{L+1} \equiv -\hat{c}_1$ , which implies  $e^{ikL} = -1$ . Hence, the choice for  $k$ 's becomes:

$$p = 0 \Rightarrow \mathcal{K}_{p=0} = \left\{ k = \pm \frac{(2n-1)\pi}{L}, \text{ with } n = 1, \dots, \frac{L}{2} \right\} \quad (3.28)$$

Notice how we can group together the  $(-k, k)$  terms, except for  $k = 0$  and  $k = \pi$  when  $p = 1$ . Let us define the *positive*  $k$  values as

$$\begin{aligned} \mathcal{K}_{p=1}^+ &= \left\{ k = \frac{2n\pi}{L}, \text{ with } n = 1, \dots, \frac{L}{2} - 1 \right\} \\ \mathcal{K}_{p=0}^+ &= \left\{ k = \frac{(2n-1)\pi}{L}, \text{ with } n = 1, \dots, \frac{L}{2} \right\} \end{aligned} \quad (3.29)$$


---



The Hamiltonian in terms of  $\hat{c}_k$  and  $\hat{c}_k^\dagger$ , with our choice of  $k$ -vectors, becomes

$$\hat{\mathbb{H}}_p = -J \sum_k^{\mathcal{K}_p} \left( 2 \cos k \hat{c}_k^\dagger \hat{c}_k + \left( e^{-2i\phi} e^{ik} \hat{c}_k^\dagger \hat{c}_{-k}^\dagger + \text{H.c.} \right) \right) - B \sum_k^{\mathcal{K}_p} (1 - 2\hat{n}_k) \quad (3.30)$$

However, using the positive  $k$  values defined in Equation 3.29 and grouping the terms in  $(-k, k)$  pairs simplifies the previous expression to

$$\hat{\mathbb{H}}_0 = \sum_{k \in \mathcal{K}_0^+} \hat{H}_k \quad \hat{\mathbb{H}}_1 = \sum_{k \in \mathcal{K}_1^+} \hat{H}_k + \hat{H}_{k=0,\pi} \quad (3.31)$$

where the Hamiltonians  $\hat{H}_k, \hat{H}_{k=0,\pi}$  have the following form

$$\hat{H}_{k=0,\pi} = -2J(\hat{n}_0 - \hat{n}_\pi) - 2B(1 - \hat{n}_0 - \hat{n}_\pi) \quad (3.32)$$

$$\hat{H}_k = 2(B - J \cos k) (\hat{c}_k^\dagger \hat{c}_k - \hat{c}_{-k} \hat{c}_{-k}^\dagger) - 2J \sin k (ie^{-2i\phi} \hat{c}_k^\dagger \hat{c}_{-k}^\dagger - ie^{2i\phi} \hat{c}_{-k} \hat{c}_k) \quad (3.33)$$

Once again, for us to be able to diagonalize the second Hamiltonian, we have to deal with the two terms that do not conserve the fermionic particle number. In order to do that, we essentially require a **Bogoliubov transformation**.

### 3.4.1 Bogoliubov transformation

Equation 3.33 can be expressed in matrix form using the space generated by the states

$$\{ \hat{c}_k^\dagger |0\rangle, \hat{c}_{-k} |0\rangle, \hat{c}_k^\dagger \hat{c}_{-k}^\dagger |0\rangle, |0\rangle \} \quad (3.34)$$

leading to the  $4 \times 4$  matrix:

$$\hat{H}_k = \left( \begin{array}{cc|cc} 2(B - J \cos k) & -2iJ e^{-2i\phi} \sin k & 0 & 0 \\ 2iJ e^{2i\phi} \sin k & -2(B - J \cos k) & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \quad (3.35)$$

We can solve the  $2 \times 2$  eigenvalue problem for the non-trivial block of the Hamiltonian as

$$\epsilon_{k\pm} = \pm \epsilon_k = \pm 2J \sqrt{\left( \cos k - \frac{B}{J} \right)^2 + \sin^2 k} \quad (3.36)$$

with their corresponding eigenvectors  $(u_{k\pm}, v_{k\pm})^T$ . Notice how the energy does not depend on the phase difference  $\phi$ , so we can set it to 0 for simplicity as well as introducing some shorthand names for  $z_k = 2(B - J \cos k)$  and  $y_k = 2J \sin k$ .

The eigenvalue problem for the positive-energy eigenvector is

$$\begin{cases} z_k u_{k+} - i y_k v_{k+} = \epsilon_k u_{k+} \\ i y_k u_{k+} - z_k v_{k+} = \epsilon_k v_{k+} \end{cases} \quad (3.37)$$

Thus, we find:

$$\begin{pmatrix} u_{k+} \\ v_{k+} \end{pmatrix} \equiv \begin{pmatrix} u_k \\ v_k \end{pmatrix} = \frac{1}{\sqrt{2\epsilon_k(\epsilon_k + z_k)}} \begin{pmatrix} \epsilon_k + z_k \\ iy_k \end{pmatrix} \quad (3.38)$$

Note that, since both  $\epsilon_k$  and  $z_k$  are even in  $k$ ,  $u_k = u_{-k}$ ; while  $v_k = -v_{-k}$  due to  $y_k$  being odd. The negative eigenvector can be found by a simple transformation of Equation 3.37. By changing the sign of the first equation, taking the complex-conjugate of both and swapping their order, we get:

$$\begin{cases} z_k(-v_k^*) - iy_k u_k^* = -\epsilon_k(-v_k^*) \\ iy_k(-v_k^*) - z_k u_k^* = -\epsilon_k u_k^* \end{cases} \quad (3.39)$$

which leads to the negative-energy eigenvector:

$$\begin{pmatrix} u_{k-} \\ v_{k-} \end{pmatrix} \equiv \begin{pmatrix} -v_k^* \\ u_k^* \end{pmatrix} = \frac{1}{\sqrt{2\epsilon_k(\epsilon_k + z_k)}} \begin{pmatrix} iy_k \\ \epsilon_k + z_k \end{pmatrix} \quad (3.40)$$

Finally, we can diagonalize the non-trivial  $2 \times 2$  Hamiltonian in Equation 3.35 with

$$\mathbf{U}_k^\dagger \mathbf{H}_k \mathbf{U}_k = \begin{pmatrix} \epsilon_k & 0 \\ 0 & -\epsilon_k \end{pmatrix} \quad (3.41)$$

$$\mathbf{U}_k = \begin{pmatrix} u_k & -v_k^* \\ v_k & u_k^* \end{pmatrix} \quad (3.42)$$

So, we can define the fermionic operator  $\hat{\Phi}_k$  as

$$\hat{\Phi}_k = \mathbf{U}_k^\dagger \begin{pmatrix} \hat{c}_k \\ \hat{c}_{-k}^\dagger \end{pmatrix} = \begin{pmatrix} u_k^* \hat{c}_k + v_k^* \hat{c}_{-k}^\dagger \\ -v_k \hat{c}_k + u_k \hat{c}_{-k}^\dagger \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_k \\ \hat{\gamma}_{-k}^\dagger \end{pmatrix} \quad (3.43)$$

where we used the fact that  $u_k = u_{-k}$  and  $v_k = -v_{-k}$ . We can check that  $\hat{\gamma}_k$  is indeed a fermion by taking its anticommutator

$$\begin{aligned} \{\hat{\gamma}_k, \hat{\gamma}_k^\dagger\} &= \{u_k^* \hat{c}_k + v_k^* \hat{c}_{-k}^\dagger, u_k \hat{c}_k^\dagger + v_k \hat{c}_{-k}\} \\ &= |u_k|^2 \{\hat{c}_k, \hat{c}_k^\dagger\} + |v_k|^2 \{\hat{c}_{-k}^\dagger, \hat{c}_{-k}\} = |u_k|^2 + |v_k|^2 = 1 \end{aligned} \quad (3.44)$$

the last equality comes from the fact that the eigenvectors are normalized.

In terms of  $\hat{\Phi}_k = (\hat{\gamma}_k, \hat{\gamma}_{-k}^\dagger)^\top$  and  $\hat{\Phi}_k^\dagger = (\hat{\gamma}_k^\dagger, \hat{\gamma}_{-k})$ , we end up with

$$\hat{H}_k = \hat{\Phi}_k^\dagger \begin{pmatrix} \epsilon_k & 0 \\ 0 & -\epsilon_k \end{pmatrix} \hat{\Phi}_k = \epsilon_k (\hat{\gamma}_k^\dagger \hat{\gamma}_k - \hat{\gamma}_{-k}^\dagger \hat{\gamma}_{-k}) = \epsilon_k (\hat{\gamma}_k^\dagger \hat{\gamma}_k + \hat{\gamma}_{-k}^\dagger \hat{\gamma}_{-k} - 1) \quad (3.45)$$

and the excitation spectrum given in Equation 3.36.

### 3.4.2 Ground state and excited states

Notice that we made the assumption that the number of sites  $L$  is even (see Observation 3.4). This allowed us to define a simple set of  $k$ -values in order to diagonalize the Hamiltonian. It can be shown, however, that with  $L$  taking an odd value, we can still diagonalize the Hamiltonian

and we arrive at the same expression for  $\hat{H}_k$  in Equation 3.45 (with the same excitation energy in Equation 3.36, see [12]&[6]).

In this section we show that the ground state of the Ising model is different depending on the parity of both the number of lattice sites  $L$  and the total number of fermions in the chain  $N$ .

We immediately see from Equation 3.45 that the ground state would be a "vacuum-like" state that annihilates all  $\hat{\gamma}_k$ . This is known as the *Bogoliubov vacuum*:

$$\hat{\gamma}_k |\emptyset_\gamma\rangle = 0 \quad \forall k \quad (3.46)$$

This state can be defined for both the  $p = 0$  (ABC) and the  $p = 1$  (PBC) sectors, and their energies depend on the parity of  $L$ .

**Even number of lattice sites:** Consider first the case of even  $L$ . The ground state for  $p = 0$  (ABC) can be written as the product

$$|\emptyset_\gamma\rangle^{\text{ABC}} \propto \prod_{k \in \mathcal{K}_0^+} \hat{\gamma}_{-k} \hat{\gamma}_k |0\rangle \quad (3.47)$$

with  $|0\rangle$  the original fermion vacuum,  $\hat{c}_k |0\rangle = 0$ . Developing the previous expression, we find

$$\begin{aligned} \prod_{k \in \mathcal{K}_0^+} \hat{\gamma}_{-k} \hat{\gamma}_k |0\rangle &= \prod_{k \in \mathcal{K}_0^+} (u_{-k}^* \hat{c}_{-k} + v_{-k}^* \hat{c}_k^\dagger) (u_k^* \hat{c}_k + v_k^* \hat{c}_{-k}^\dagger) |0\rangle \\ &= \prod_{k \in \mathcal{K}_0^+} v_k^* (u_{-k}^* + v_{-k}^* \hat{c}_k^\dagger \hat{c}_{-k}^\dagger) |0\rangle = \prod_{k \in \mathcal{K}_0^+} v_k^* (u_k^* - v_k^* \hat{c}_k^\dagger \hat{c}_{-k}^\dagger) |0\rangle \end{aligned} \quad (3.48)$$

This state can be expressed as a product of Cooper pairs [15] if we normalize it:

$$|\emptyset_\gamma\rangle^{\text{ABC}} = \prod_{k \in \mathcal{K}_0^+} (u_k^* - v_k^* \hat{c}_k^\dagger \hat{c}_{-k}^\dagger) |0\rangle = \prod_{k \in \mathcal{K}_0^+} (u_k + v_k \hat{c}_k^\dagger \hat{c}_{-k}^\dagger) |0\rangle \quad (3.49)$$

this last step can only be done because we are assuming  $\phi = 0$ , for which  $u_k$  is real and  $v_k$  purely imaginary. The energy of this state is simply the sum of all negative excitation energies

$$E_0^{\text{ABC}} = - \sum_{k \in \mathcal{K}_0^+} \epsilon_k \quad (3.50)$$

The ground state for the PBC-sector ( $p = 1$ ) must contain an *odd* number of particles. This imposes a restriction on the unpaired Hamiltonian terms,  $\hat{H}_{k=0,\pi}$ , since the BCS-paired states will always be fermion-even. Looking at Equation 3.32, it is easy to see that (for  $B > 0$ ), the ground state has  $\hat{n}_{k=0} \rightarrow 1$  and  $\hat{n}_{k=\pi} \rightarrow 0$ . This results in an extra contribution of

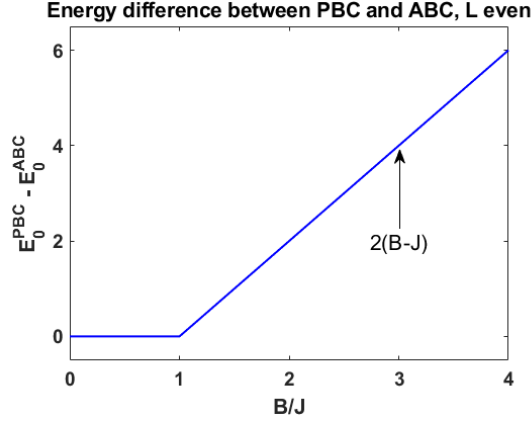
$$\delta E_{k=0,\pi} = \min(\hat{H}_{k=0,\pi}) = -2J$$

The PBC ground state results in:

$$|\emptyset_\gamma\rangle^{\text{PBC}} = \hat{c}_{k=0}^\dagger \prod_{k \in \mathcal{K}_1^+} (u_k^* - v_k^* \hat{c}_k^\dagger \hat{c}_{-k}^\dagger) |0\rangle \quad (3.51)$$

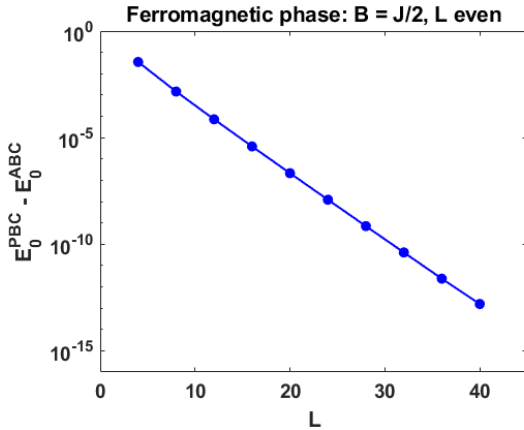
with its corresponding energy being

$$E_0^{\text{PBC}} = -2J - \sum_{k \in \mathcal{K}_1^+} \epsilon_k \quad (3.52)$$

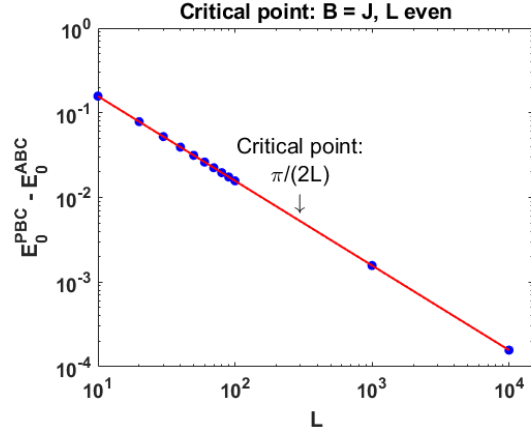


(a) PBC- and ABC-sector energy difference.

**Figure 2:** Ground state energies for even number of lattice sites.



(b) Energy difference in ferromagnetic phase.



(c) Energy difference at critical point.

**Figure 2:** Ground state energies for even number of lattice sites. (cont.)

Figure 2 depicts the difference between  $E_0^{\text{PBC}}$  and  $E_0^{\text{ABC}}$  for  $L \rightarrow \infty$ , leading to an energy gap  $\Delta E = 2(B - J)$  when  $B > J$  between the ground state  $|\emptyset_\gamma\rangle^{\text{ABC}}$  and the first excited state  $|\emptyset_\gamma\rangle^{\text{PBC}}$ . We can see an exponential decay in the energy gap in the ferromagnetic region (fig.2b) and a power-law behaviour at the critical point (fig.2c).

From this we can conclude that, for  $L$  even, the *true* ground state of the Ising model corresponds to the ABC-sector state  $|\emptyset_\gamma\rangle^{\text{ABC}}$ , but there is a double degeneracy when  $L \rightarrow \infty$  in the ferromagnetic phase  $B < J$ .

**Odd number of lattice sites:** For the  $L$  odd case, we can expect the behaviour to be asymptotically the same as in the even case. Indeed, as shown in [12],  $|\emptyset_\gamma\rangle^{\text{PBC}}$  becomes the

*true* ground state in this case, and the gap between the energies seen in Figure 2 is identical. Thus, for simplicity reasons, we will be assuming for the rest of this document that the number of lattice sites  $L$  is **always even** and equal to the total number of spins  $N$ .

### 3.4.3 Summary of theoretical results

To summarize all that we have seen in this first part,

- The system that we want to study is described by the Hamiltonian

$$\hat{H} = -J \sum_{i=1}^N \hat{\sigma}_i^z \hat{\sigma}_{i+1}^z - B \sum_{i=1}^N \hat{\sigma}_i^x \quad (3.53)$$

with periodic boundary conditions  $\hat{\sigma}_{N+1}^z \equiv \hat{\sigma}_1^z$ .

- The ground state  $|\emptyset_\gamma\rangle$  is given by Equation 3.49, since we are assuming even number of lattice sites  $L$ , and equal to the total number of spins  $N$ .
- The ground state energy can be calculated with Equation 3.50, or

$$E_0 = - \sum_{k \in \mathcal{K}_0^+} \epsilon_k = -2J \sum_{k \in \mathcal{K}_0^+} \sqrt{\left(\cos k - \frac{B}{J}\right)^2 + \sin^2 k} \quad (3.54)$$

where  $\mathcal{K}_0^+ = \{k = \frac{(2n-1)\pi}{N}, \text{ with } n = 1, \dots, \frac{L}{2}\}$ .

In the last section of this theoretical introduction, we calculate the magnetizations of the ground state  $|\emptyset_\gamma\rangle$ , in order to see their behaviour when the value of the external magnetic  $B$  field changes.

## 3.5 Magnetization of the ground state

In addition to its energy, we can calculate the magnetization of the ground state, assuming  $L$  is even and  $|\emptyset_\gamma\rangle^{\text{ABC}} \equiv |\emptyset_\gamma\rangle$ , with the following expressions:

$$\langle \hat{\sigma}^z \rangle = \langle \emptyset_\gamma | \hat{\sigma}^z | \emptyset_\gamma \rangle = \frac{1}{L} \sum_{i=1}^L \langle \emptyset_\gamma | \hat{\sigma}_i^z | \emptyset_\gamma \rangle \quad (3.55)$$

for the longitudinal magnetization, and

$$\langle \hat{\sigma}^x \rangle = \langle \emptyset_\gamma | \hat{\sigma}^x | \emptyset_\gamma \rangle = \frac{1}{L} \sum_{i=1}^L \langle \emptyset_\gamma | \hat{\sigma}_i^x | \emptyset_\gamma \rangle \quad (3.56)$$

for the transversal one. In both equations, we made use of the fact that  $|\emptyset_\gamma\rangle$  is normalized.

There is an issue with Equation 3.55 which is that, because our Hamiltonian (eq.3.2) is invariant under the exchange of  $\hat{\sigma}^z \leftrightarrow -\hat{\sigma}^z$ , the ground state and their contributions to the magnetization will be the same. However, the expectation value  $\langle \hat{\sigma}^z \rangle$  should be of opposite sign to  $\langle -\hat{\sigma}^z \rangle$ ,

which is only possible if  $\langle \hat{\sigma}^z \rangle$  is always 0. In this case, we trivially have  $\langle \hat{\sigma}^z \rangle = 0$  and we need a different estimator of the magnetization.

For that, we define the **order parameters** of the ground state as

$$m_z = \sqrt{\langle \emptyset_\gamma | (\hat{\sigma}^z)^2 | \emptyset_\gamma \rangle} = \sqrt{\langle \emptyset_\gamma | \left( \frac{1}{L} \sum_{i=1}^L \hat{\sigma}_i^z \right)^2 | \emptyset_\gamma \rangle} \quad (3.57)$$

$$m_x = \sqrt{\langle \emptyset_\gamma | (\hat{\sigma}^x)^2 | \emptyset_\gamma \rangle} = \sqrt{\langle \emptyset_\gamma | \left( \frac{1}{L} \sum_{i=1}^L \hat{\sigma}_i^x \right)^2 | \emptyset_\gamma \rangle} \quad (3.58)$$

As a simple example, if we look at the expression of  $(\hat{\sigma}^z)^2$  projected onto the spin basis  $\mathcal{S}$  for  $L = 2$ , we see that

$$\begin{aligned} \mathcal{S} &= \{ |\uparrow\uparrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\downarrow\downarrow\rangle \} \\ (\hat{\sigma}^z)^2 &= \left( \frac{\hat{\sigma}_1^z + \hat{\sigma}_2^z}{2} \right)^2 = \frac{1}{4} \left( (\hat{\sigma}_1^z)^2 + (\hat{\sigma}_2^z)^2 + \hat{\sigma}_1^z \hat{\sigma}_2^z + \hat{\sigma}_2^z \hat{\sigma}_1^z \right) \\ &= \frac{1}{4} \left( \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right) \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

If we take a look at the action of this operator on the spin basis, we can see it takes maximum value 1 for the ordered  $|\uparrow\uparrow\rangle$  and  $|\downarrow\downarrow\rangle$ , or **ferromagnetic** phase; while the disordered, or **antiferromagnetic** phase  $|\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle$  takes minimum value 0. We will use this operator to characterize the behaviour of the ground state when varying the magnitude of the transversal field.

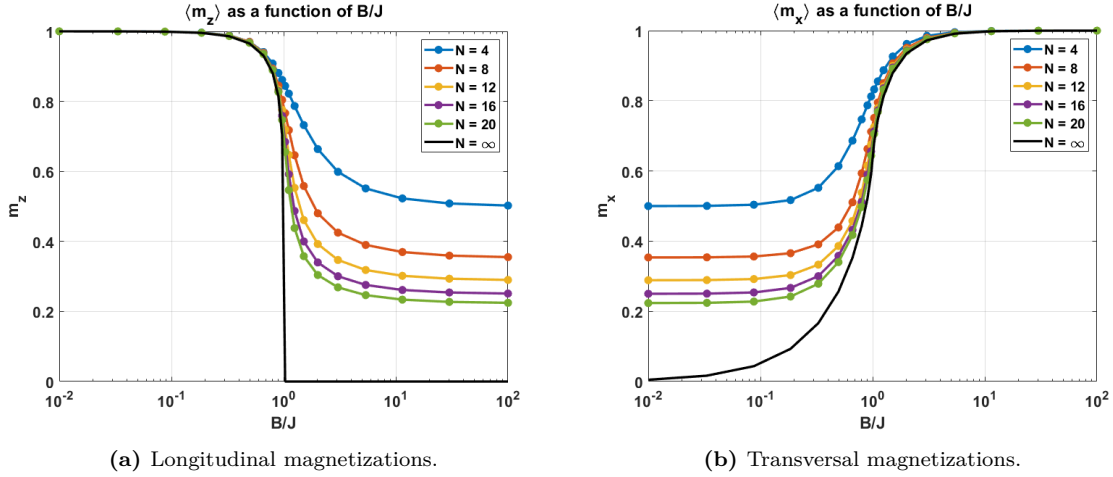
From Equations 3.57 & 3.58, we can use the result found in [5] & [4], which states that for  $L \rightarrow \infty$ , the expressions simplify to:

$$m_z = \left( 1 - \frac{B^2}{J^2} \right)^{\frac{1}{8}} \quad \text{if } B < J, \text{ 0 otherwise} \quad (3.59)$$

$$m_x = \frac{1}{\pi} \int_0^\pi \left( \frac{B + J \cos k}{\sqrt{J^2 + B^2 + 2JB \cos k}} \right) dk \quad (3.60)$$

Figure 3 shows the ground state magnetizations as well as the approximations made in Equations 3.59 & 3.60 as a function of  $L$ . We can see that the exact values tend to the expressions in Equations 3.59 & 3.60 as  $L \rightarrow \infty$ , as expected.

In Figure 3, we can also observe the phase transition with the order parameters. When  $B < J$ , we are in the *ordered phase*, where the ground state is degenerate as we saw in Figure 2b. However, beyond the critical point  $B/J = 1$ , the order is destroyed due to the strong external magnetic field, which tends to flip the spins.



**Figure 3:** Ground state magnetizations for (even) increasing lattice sites.

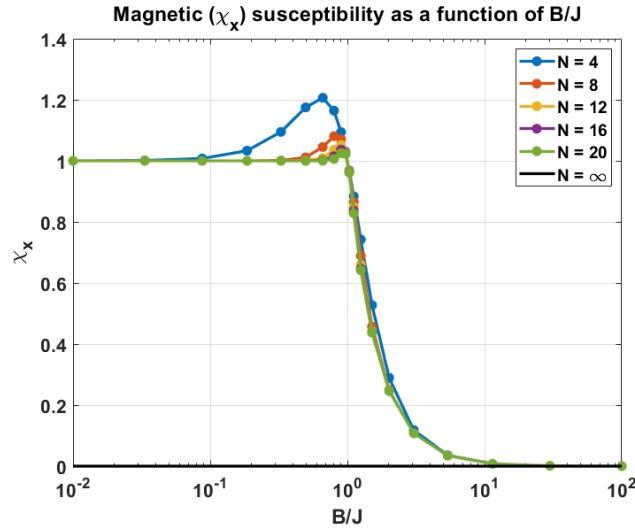
Lastly, we could look at the magnetic susceptibility of the ground state as defined in [2]

$$\chi_x = \frac{N}{T} \langle (\hat{\sigma}^x)^2 \rangle - \langle |\hat{\sigma}^x| \rangle^2 \quad (3.61)$$

where  $\hat{\sigma}^x$  is the global magnetization defined in Equation 3.56

$$\hat{\sigma}^x = \frac{1}{N} \sum_{i=1}^N \hat{\sigma}_i^x \quad (3.62)$$

As we are working in the  $T \rightarrow 0$  regime, we plot  $\chi_x$  without the temperature.



**Figure 4:** Ground state magnetic susceptibilities.

As it can be seen in Figure 4, the phase transition that occurs around the critical point  $B/J = 1$  is shown by the peak in the magnetic susceptibility. From this result, we can observe that as  $N$  increases, the phase transition moves towards  $B/J = 1$ , as expected from the thermodynamic limit shown in Figure 2.

## 4 Methodology

In this section, we introduce the variational method and the RBM wavefunction. We want to use them to find the ground state of the 1D TFI model in different cases, and compare the obtained results with those of the exact diagonalization that we have seen in the previous section.

### 4.1 Variational method

The *variational method* is very useful for estimating the ground state energy  $E_0$  when exact solutions of a problem are not known or available [3].

We define a *trial state*  $|\Psi\rangle$ , which tries to approximate the true ground ket  $|\psi_0\rangle$ , such that the mean energy of the system:

$$\langle \hat{H} \rangle \equiv \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (4.1)$$

is minimized.  $|\Psi\rangle$  does not need to be normalized.

---

**Theorem 4.1** (Upper bound energy). We can obtain an *upper bound* to the ground state energy  $E_0$  by considering different trial wavefunctions.

$$\langle H \rangle \geq E_0 \quad (4.2)$$

To prove this, we can expand  $|\Psi\rangle$  in terms of the eigenfunctions  $|\psi_n\rangle$  (which are unknown) as

$$|\Psi\rangle = \sum_n |\psi_n\rangle \langle \psi_n | \Psi \rangle, \text{ where } \hat{H} |\psi_n\rangle = E_n |\psi_n\rangle \quad (4.3)$$

and use this to calculate the energy  $\langle \hat{H} \rangle$

$$\langle H \rangle = \frac{\sum_n |\langle \psi_n | \Psi \rangle|^2 E_n}{\sum_n |\langle \psi_n | \Psi \rangle|^2} = \frac{\sum_n |\langle \psi_n | \Psi \rangle|^2 (E_n - E_0)}{\sum_n |\langle \psi_n | \Psi \rangle|^2} + E_0 > E_0 \quad (4.4)$$

where the last inequality comes from the fact that  $E_n - E_0$  must necessarily be positive.

**However!** In order to obtain an upper bound to the exact ground state, the trial wavefunction **must** have a *non-zero projection* on the ground state  $|\psi_0\rangle$ .

The reason is that the variational method provides an upper bound to the **lowest** energy eigenstate for which our trial wavefunction has a non-zero overlap.

---

Notice that the method does not tell us what kind of trial wavefunction we should use, nor does it say anything about the discrepancy between the estimated energy and the true ground state energy.

In theory, we could try *every possible function* in Hilbert space, keeping only the lowest energy

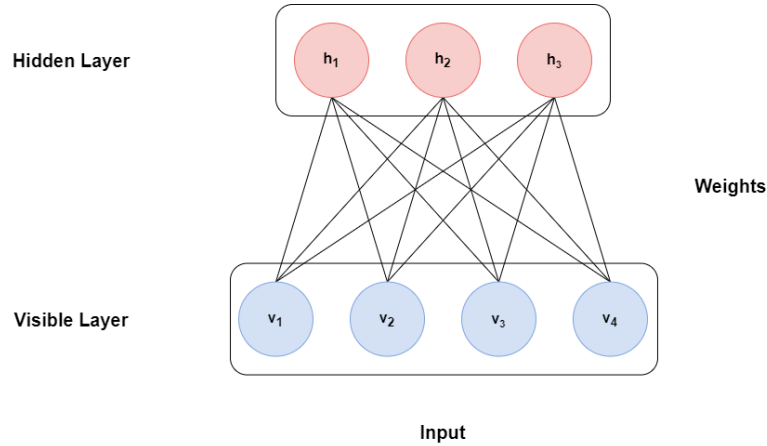


state, and we would end up finding the exact ground state. However, this is not a practical approach due to the size of the Hilbert space, so what it is usually done is to limit the search to a **small subset** of functions that are parameterized by some variables  $(\alpha, \beta, \gamma, \dots)$  that are later optimized for energy minimization.

In this thesis, we explore the idea of using a **Restricted Boltzmann Machine** as our **variational wavefunction** for solving the TFI model. But, first of all, we should explain what a Restricted Boltzmann Machine is, and how can we use it to solve our quantum many-body problem.

## 4.2 Restricted Boltzmann Machines

A Restricted Boltzmann Machine (RBM) is an energy based, undirected graphical model that consists of  $n$  visible units  $\vec{v} = (V_1, \dots, V_n)$  representing the observable data, and  $m$  inaccessible hidden units  $\vec{h} = (H_1, \dots, H_m)$  that capture the correlations between the observed variables [13]. It differs from the more general Boltzmann Machine model in that it adds a restriction such that there cannot be weights (or connections) between units in the same layer.



**Figure 5:** Diagram of an RBM.

In other words, its associated undirected graph is bipartite. Furthermore, the model has an energy function that is defined as

$$E(\vec{v}, \vec{h}) = -\vec{v}^T \mathbf{W} \vec{h} - \vec{v}^T \vec{b} - \vec{c}^T \vec{h}. \quad (4.5)$$

Our variational model is parameterized by the set of parameters  $\{\vec{b}, \vec{c}, \mathbf{W}\}$ , which we will generally refer to as  $\theta$ . These can be divided into two groups. The *offsets* or *bias*, composed by the vectors  $\vec{b}$  and  $\vec{c}$ , where  $b_i$  and  $c_j$  are terms associated with the  $i$ -th visible and  $j$ -th hidden variables, respectively. The second group corresponds to the weight matrix  $\mathbf{W}$ , where each element  $w_{ij}$  connects one visible unit  $v_i$  to one hidden unit  $h_j$ .

This flexible structure allows an RBM to be used as a variational wavefunction due to its ability to efficiently represent complex quantum states with a relatively *small* number of parameters. Its architecture, allows it to capture the intricate correlations and entanglement present in many-body quantum systems, having the potential to approximate their ground state wavefunctions within the variational principle framework.

**Definition 4.2** (RBM wavefunction). Let  $\Psi$  be a trial wavefunction of the form:

$$\Psi(\vec{s}; \boldsymbol{\theta}) = \sum_{\{h_i\}} e^{\sum_j b_j s_j + \sum_i c_i h_i + \sum_{ij} W_{ij} h_i s_j} \quad (4.6)$$

which corresponds to the exponential of an RBM artificial neural network [9] (eq.4.5), where:

- $s_j$  are the spins of the system, codified with the visible units with  $\{0, 1\}$  values.
- $h_i$  are auxiliary spins, codified by the hidden units with  $\{-1, 1\}$  values.
- $\boldsymbol{\theta} \equiv \{\vec{b}, \vec{c}, \mathbf{W}\}$  are the parameters of the RBM, which we will optimize in order to minimize the mean energy

$$E = \frac{\langle \psi_{\boldsymbol{\theta}} | \hat{H} | \psi_{\boldsymbol{\theta}} \rangle}{\langle \psi_{\boldsymbol{\theta}} | \psi_{\boldsymbol{\theta}} \rangle} \quad (4.7)$$

We can simplify Equation 4.6 further by considering binary hidden units  $h_i = \{-1, 1\}$ :

$$\Psi(\vec{s}; \boldsymbol{\theta}) = e^{\vec{b}^\dagger \cdot \vec{s}} \left( e^{\sum_i c_i + \mathbf{W}_i \cdot \vec{s}} + e^{-(\sum_i c_i + \mathbf{W}_i \cdot \vec{s})} \right) = e^{\vec{b}^\dagger \cdot \vec{s}} \prod_{i=1}^{N_h} 2 \cosh(c_i + \mathbf{W}_i \cdot \vec{s}) \quad (4.8)$$

$\mathbf{W}_i$  corresponds to the  $i$ -th row of the weight matrix.

Notice that, in order to allow for a complete description of both the amplitude and phase of the wavefunction, the network parameters are (in general) to be taken *complex-valued*.

### 4.3 Optimization of RBM parameters

In order to optimize the RBM parameters, we impose the condition that the optimal set of parameters  $\boldsymbol{\theta}_{\text{opt}}$  is the one that minimizes the mean energy in Equation 4.7. Namely, we fix  $\boldsymbol{\theta}$  with the condition:

$$\frac{\partial E}{\partial \theta_k} = 0, \quad \forall \theta_k \in \boldsymbol{\theta} \quad (4.9)$$

However, the difficult part of using an RBM as our wavefunction is deciding how many units should we use for each layer, and how to optimize the parameters in order to minimize the energy. For this reason, we first solve the simplest case of the TFI model: a chain of **two** spins.

#### 4.3.1 Simplest case: $N = 2$ spins.

For simplicity, we first consider the case where we have *no hidden units*. Then, the trial wavefunction defined in Equation 4.6 simplifies to

$$\Psi(\vec{s}) = e^{b_1 s_1 + b_2 s_2} \quad (4.10)$$

where  $s_1, s_2$  are the spins of particle 1 and 2, respectively. Notice that since the parameters can be complex, this implies that we may have up to 4 conditions to satisfy (real part + imaginary

part). However, we can consider a simpler case in which the parameters are  $b_1 = b_2 = b \in \mathbb{R}$ , this way we can easily calculate the expression of the derivative in Equation 4.9, and thus find the optimal parameters that minimize the energy.

For a system with 2 spins, we define the basis  $\mathcal{S} = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$  denoting 1 as an upwards spin  $|\uparrow\rangle$ , and 0 as the downwards spin  $|\downarrow\rangle$ . The Hamiltonian of the system is

$$\hat{H} = -J(\hat{\sigma}_1^z \hat{\sigma}_2^z + \hat{\sigma}_2^z \hat{\sigma}_1^z) - B(\hat{\sigma}_1^x + \hat{\sigma}_2^x) \quad (4.11)$$

or, rewriting the equation in the **basis of spins**:

$$= -J(\sigma_1^z \otimes \sigma_2^z + \sigma_2^z \otimes \sigma_1^z) - B(\sigma_1^x \otimes \mathbb{1}_2 + \mathbb{1}_2 \otimes \sigma_2^x) \quad (4.12)$$

$$\hat{H} = -2J \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} - B \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} -2J & -B & -B & 0 \\ -B & 2J & 0 & -B \\ -B & 0 & 2J & -B \\ 0 & -B & -B & -2J \end{pmatrix} \quad (4.13)$$

We can diagonalize this matrix to find the exact eigenenergies and eigenstates of the system, projected to our spin basis:

- **Eigenvalue:**  $+2J$

**Eigenstate:**

$$|\psi_{+2J}\rangle = \begin{pmatrix} 0 & -1 & 1 & 0 \end{pmatrix}^\top \equiv |\downarrow\uparrow\rangle - |\uparrow\downarrow\rangle \quad (4.14)$$

- **Eigenvalue:**  $-2J$

**Eigenstate:**

$$|\psi_{-2J}\rangle = \begin{pmatrix} -1 & 0 & 0 & 1 \end{pmatrix}^\top \equiv |\downarrow\downarrow\rangle - |\uparrow\uparrow\rangle \quad (4.15)$$

- **Eigenvalue:**  $2\sqrt{J^2 + B^2} \equiv 2k^2$

**Eigenstate:**

$$|\psi_{+2k^2}\rangle = \begin{pmatrix} B & -J - k^2 & -J - k^2 & B \end{pmatrix}^\top \equiv B(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle) - (J + k^2)(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \quad (4.16)$$

- **Eigenvalue:**  $-2\sqrt{J^2 + B^2} \equiv -2k^2$

**Eigenstate:**

$$|\psi_{-2k^2}\rangle = \begin{pmatrix} B & -J + k^2 & -J + k^2 & B \end{pmatrix}^\top \equiv B(|\uparrow\uparrow\rangle + |\downarrow\downarrow\rangle) - (J - k^2)(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) \quad (4.17)$$

Assuming  $J, B > 0$ , the lowest energy eigenvalue is  $-2k^2$ , with eigenstate  $|\psi_{-2k^2}\rangle \equiv |\psi_0\rangle$  being the actual ground state.

To calculate the mean energy of our RBM wavefunction, we project  $\Psi(\mathcal{S})$  to our spin basis and

use Equation 4.13:

$$\begin{aligned}\langle \Psi | \hat{H} | \Psi \rangle &= \begin{pmatrix} e^{2b} & e^b & e^b & 1 \end{pmatrix} \begin{pmatrix} -2J & -B & -B & 0 \\ -B & 2J & 0 & -B \\ -B & 0 & 2J & -B \\ 0 & -B & -B & -2J \end{pmatrix} \begin{pmatrix} e^{2b} \\ e^b \\ e^b \\ 1 \end{pmatrix} \\ &= -(2J + 4Be^b + 4Be^{3b} - 4Je^{2b} + 2Je^{4b}) \\ \langle \Psi | \Psi \rangle &= (e^{2b} + 1)^2\end{aligned}$$

Thus, the mean energy is equal to

$$E = -\frac{2J + 4Be^b + 4Be^{3b} - 4Je^{2b} + 2Je^{4b}}{(e^{2b} + 1)^2}$$

and we can find the minimum by calculating its derivative

$$\frac{\partial E}{\partial b} = \frac{4e^b(e^{2b} - 1)(B - 4Je^b + Be^{2b})}{(e^{2b} + 1)^3}$$

Now as a practical example, we consider the case  $J = 2, B = 1$ , so that the expression for the energy simplifies to

$$E = -\frac{4 + 4e^b + 4e^{3b} - 8e^{2b} + 4e^{4b}}{(e^{2b} + 1)^2}$$

and its gradient becomes

$$\frac{\partial E}{\partial b} = -\frac{4e^b(8e^{3b} - e^{4b} - 8e^b + 1)}{(e^{2b} + 1)^3}$$

This expression has 3 real solutions  $b = \{0, \log(4 + \sqrt{15}), \log(4 - \sqrt{15})\}$ , with their corresponding energies  $E = \{-2, -4.25, -4.25\}$ . If we compare our estimation of the ground state energy  $E_{\min} = -4.25$  with the real ground state energy in this case ( $E_0 = -2k^2 = -2\sqrt{J^2 + B^2} \simeq -4.472$ , for  $J = 2, B = 1$ ), we see that

$$E_0 \simeq -4.472 < E_{\min}$$

as expected from the variational principle, as a trial wavefunction always provides an upper bound to the exact ground state energy. Furthermore, to obtain a better approximation, we should consider more sophisticated wavefunctions. For instance, including one hidden unit in Equation 4.8

$$\Psi(\vec{s}) = 2e^{b_1 s_1 + b_2 s_2} \cosh(c + \omega_1 s_1 + \omega_2 s_2) \quad (4.18)$$

Compared to the previous wavefunction in eq.4.10, we notice the new term of the hyperbolic cosine appears when we consider an RBM with 1 hidden neuron. Proceeding as before, we now try to find the optimal values of these parameters such that we obtain the minimal energy.

We can see immediately that trying to find the optimal parameters by differentiating the

expression of the mean energy with respect to each one of them becomes too troublesome already. Instead, it is easier to simply brute force the exploration to find a set of parameters with the lowest energy, i.e., to explore a subspace of  $P$  parameters ( $D \subset \mathcal{C}^P$ ), which in this case is 5.

This exploration can be done using several different algorithms, as for instance:

---

**Definition 4.3** (Divide and Conquer). We define a *divide and conquer* strategy as recursively finding which is the point that minimizes the energy  $E(\boldsymbol{\theta})$  and updating the parameter vector with finer and finer grid points. Namely:

1. Calculate  $E(\theta_i)$  for all  $\theta_i \in \boldsymbol{\theta}_n$ .
2. Keep the value of  $\theta_{\min}$  that minimizes said energy.
3. Define the new vector  $\boldsymbol{\theta}_{n+1}$  as the points located at length  $L_{n+1} = L_n/2$  centered around  $\theta_{\min}$ :

$$\boldsymbol{\theta}_{n+1} = \theta_{\min} + L_{n+1}\boldsymbol{\theta}_0$$

We generalize this strategy for any number of parameters  $\alpha, \beta, \gamma, \dots$  (in our case  $\vec{\mathbf{b}}, \vec{\mathbf{c}}, \mathbf{W}$ ).

Notice that this strategy does **not guarantee** the convergence to the *global minimum*, and it can easily miss it, since it always follows the direction of decreasing energy.

---

The computational cost of this strategy is simply the number of times that we repeat the procedure (= the *depth* of our exploration  $d$ ), times the number of points that we explore at each step  $k^P$  (we visit  $k$  points for each parameter at each step). Thus, the computational cost is  $\mathcal{O}(d \cdot k^P)$ .

Trying out different values of the initial length  $L$  and the depth of the exploration  $d$ , we find the following values of the mean energy for  $J = 2$ ,  $B = 1$ :

	$L = 1$			$L = 2$			$L = 3$		
	$d = 5$	$d = 10$	$d = 20$	$d = 5$	$d = 10$	$d = 20$	$d = 5$	$d = 10$	$d = 20$
$E_{\min}$	-4.389	-4.403	-4.404	-4.342	-4.424	-4.426	-4.461	-4.472	<b>-4.472</b>
$\varepsilon_r(\%)$	1.859	1.539	1.530	2.911	1.085	1.038	0.260	$6.51e-4$	<b>5.18e-10</b>

**Table 1:** Minimal energies found using divide and conquer strategy.  $N = 2$  spins,  $N_h = 1$ .

The relative error in Table 1 is defined as

$$\varepsilon_r = \left| \frac{E - E_{\text{exact}}}{E_{\text{exact}}} \right| \quad (4.19)$$

The optimal values of the RBM parameters are, from the result corresponding to  $L = 3$ ,  $d = 20$ :

$$\begin{aligned}\vec{\mathbf{b}} &= \begin{pmatrix} -0.271 + 1.881i & -2.160 - 3.750i \end{pmatrix}^T \\ \vec{\mathbf{c}} &= \begin{pmatrix} -3.750 + 0i \end{pmatrix} \\ \mathbf{W} &= \begin{pmatrix} 0 + 0i & 0.750 + 0.750i \end{pmatrix}\end{aligned}$$

which lead to the state  $|\psi_\theta\rangle \equiv \langle \mathcal{S} | \Psi \rangle$ , that we can compare with the exact ground state  $|\psi_0\rangle$  seen in Equation 4.17 for  $J = 2$ ,  $B = 1$ :

$$|\psi_\theta\rangle = \begin{pmatrix} \langle \uparrow\uparrow | \Psi(\theta) \rangle \\ \langle \uparrow\downarrow | \Psi(\theta) \rangle \\ \langle \downarrow\uparrow | \Psi(\theta) \rangle \\ \langle \downarrow\downarrow | \Psi(\theta) \rangle \end{pmatrix} = \begin{pmatrix} 0.6882 \\ 0.1625 \\ 0.1625 \\ 0.6882 \end{pmatrix}, \quad |\psi_0\rangle = \begin{pmatrix} 0.6882 \\ 0.1625 \\ 0.1625 \\ 0.6882 \end{pmatrix}$$

We can see that we have obtained a really good solution just by adding one hidden unit to our RBM ansatz.

#### 4.3.2 Next case: $N = 4$ spins.

The next simplest case corresponds to a chain of 4 spins. We can diagonalize the Hamiltonian matrix, which in this case is  $2^4 \times 2^4 = 16 \times 16$  in size, to find that the ground state energy is  $E_0 = -8.543117$ , for  $J = 2$ ,  $B = 1$ .

Using the same divide and conquer scheme, we obtain table 2 corresponding to the simple case of 0 and 1 hidden units, initial square length  $L$  and search depth  $d$ .

	$L = 1$			$L = 2$			$L = 3$		
	$d = 5$	$d = 10$	$d = 20$	$d = 5$	$d = 10$	$d = 20$	$d = 5$	$d = 10$	$d = 20$
$N_h = 0 :$									
$E_{\min}$	-7.131	-7.229	-7.232	-8.498	-8.500	-8.500	-8.482	-8.500	-8.500
$\varepsilon_r(\%)$	16.53	15.38	15.34	0.528	0.505	0.505	0.710	0.505	0.505
$t_{\text{comp}}(s)$	0.020	0.035	0.064	0.028	0.038	0.075	0.027	0.038	0.056
$N_h = 1 :$									
$E_{\min}$	-8.493	-8.498	-8.498	-8.502	-8.505	-8.505	-8.481	-8.505	-8.505
$\varepsilon_r(\%)$	0.591	0.528	0.528	0.480	0.466	0.462	0.730	0.462	0.461
$t_{\text{comp}}(s)$	13.295	26.011	52.889	13.291	25.449	50.901	13.298	26.478	52.254

**Table 2:** Minimal energies found using divide and conquer strategy.  $N = 4$  spins, with 0 and 1 hidden units.

Notice in Table 2 that, despite the errors being quite small, the computational time  $t_{\text{comp}}$  of this strategy explodes very fast. In fact, not only does the cost grow as  $\mathcal{O}(d \cdot k^P)$ , but also

the memory required to generate all the possible permutations of the parameter values grows as  $\mathcal{O}(k^P)$  as well.

To summarize, the divide and conquer strategy gets too expensive already for  $N = 4$  spins, and it does not even guarantee **global convergence**, as we mentioned before. For this reason, we need to come up with an efficient way of exploring the parameter space while globally converging to the minimum of energy. Hence, we can introduce the **Simulated Annealing** Monte Carlo optimization method.

---

**Definition 4.4** (Simulated Annealing). Consider the multivariate function  $E(x_1, \dots, x_N)$  that should be minimized. We introduce a fictitious *temperature*  $T$  that we use to define the following probability distribution function:

$$P(x_1, \dots, x_N) = \exp\left(\frac{E(x_1, \dots, x_N)}{T}\right) \quad (4.20)$$

We then draw samples from it according to **Metropolis algorithm**:

1. Generate new solution  $\vec{x} \rightarrow \vec{x}'$  by randomly displacing the previous state.
2. Accept the new solution if  $E(\vec{x}') < E(\vec{x})$  or with probability

$$p_{\text{acc}} = \min\left\{1, \frac{P(\vec{x}')}{P(\vec{x})}\right\}$$

3. Additionally, in simulated annealing, we decrease the temperature after the system reaches equilibrium with some *cooling down* scheme. For instance, it is common to decrease  $T$  by a factor  $\gamma$  called the *cooling rate* usually very close to 1 (e.g. 0.999):

$$T_{i+1} = T_i \gamma$$

Simulated annealing is known to be an *asymptotic* algorithm. Essentially, if the initial temperature is large compared to the **characteristic energy** of the system, the algorithm accepts new solutions with high probability. Then, if we let the system thermalize and reach equilibrium each time we decrease the temperature infinitesimally, we can expect that the algorithm will converge to the **global minimum** if the final temperature is much smaller compared to the characteristic energy of the system. This is called *adiabatic descent*, and can only be done when  $N_{\text{iterations}} \rightarrow \infty$ .

---

Algorithms 1 & 2 show an implementation of this strategy. Regarding the computational cost, the most expensive calculation we are doing with this scheme is evaluating the wavefunction exactly for every possible spin state, which has a cost  $\mathcal{O}(2^N)$ , to calculate the mean energy at each iteration.

If we now consider the previous case using simulated annealing method, we obtain the results of table 3, corresponding to an initial temperature  $T = 20$ , and a cooling rate of  $\gamma = 0.999$ . This means that after  $N_{\text{iter}}$  iterations, where we wait for the system to reach equilibrium before

---

**Algorithm 1** Mean energy evaluation

```

1: function WAVEFUNCTION( $\vec{s}$ ,  $\vec{\theta}$ )                                ▷ Calculates projection of  $\langle \vec{s} | \Psi_{\vec{\theta}} \rangle$ 
2:    $\vec{b}, \vec{c}, \mathbf{W} \leftarrow \vec{\theta}$                                 ▷ Unravel  $\vec{\theta}$  for the different parameters
3:    $\text{exp\_term} \leftarrow \vec{b}^\dagger \cdot \vec{s}$ 
4:    $\text{cosh\_term} \leftarrow 0$ 
5:   for all  $i \in [1, \dots, N_h]$  do
6:      $\text{cosh\_term} \leftarrow \text{cosh\_term} + \log(\cosh(c_i + \vec{W}_i^\dagger \cdot \vec{s}))$ 
7:   return  $\exp(\text{exp\_term} + \text{cosh\_term})$ 

8: function MEANENERGY( $\hat{H}$ ,  $\vec{\theta}$ )                                ▷ Calculates the mean energy of  $|\Psi_{\vec{\theta}}\rangle$ 
9:    $\Psi_{\vec{\theta}} \leftarrow \{\}$ 
10:  for all  $\vec{s} \in \mathcal{S}$  do                                ▷ All  $2^N$  possible spin states
11:    append WAVEFUNCTION( $\vec{s}$ ,  $\vec{\theta}$ ) to  $\Psi_{\vec{\theta}}$ 
12:   $\text{norm} \leftarrow \Psi_{\vec{\theta}}^\dagger \cdot \Psi_{\vec{\theta}}$ 
13:   $\text{energy} \leftarrow \Psi_{\vec{\theta}}^\dagger \cdot \hat{H} \cdot \Psi_{\vec{\theta}}$ 
14:  return  $\text{energy} / \text{norm}$ 

```

---



---

**Algorithm 2** Simulated Annealing procedure

**Input:**  $N_{\text{iter}}, N_{\text{thermal}}, T_0, \gamma, \Delta t$  ▷ Adjust parameters so that  $P_{\text{acc}} \sim 0.5$

**Outputs:**  $\vec{\theta}_{\text{min}}, E_{\text{min}}$

```

1:  $\vec{\theta} \leftarrow$  Random initialization
2:  $E \leftarrow \text{MEANENERGY}(\hat{H}, \vec{\theta})$ 
3:  $T \leftarrow T_0$ 
4:  $n_{\text{acc}} \leftarrow 0$ 
5: for  $\text{iter} \leftarrow 1$  to  $N_{\text{iter}}$  do
6:   for  $\text{sample} \leftarrow 1$  to  $N_{\text{thermal}}$  do
7:      $\vec{\theta}' \leftarrow \vec{\theta} + \Delta t \cdot \vec{\xi}$                                 ▷ New solution by randomly displacing previous one
8:      $E' \leftarrow \text{MEANENERGY}(\hat{H}, \vec{\theta}')$ 
9:      $\Delta E \leftarrow E' - E$ 
10:     $u \leftarrow \text{UNIFORM}(0, 1)$                                 ▷ Uniform random variable
11:    if  $\Delta E < 0$  or  $\exp(-\Delta E / T) > u$  then
12:       $\vec{\theta} \leftarrow \vec{\theta}'$ 
13:       $E \leftarrow E + \Delta E$ 
14:       $n_{\text{acc}} \leftarrow n_{\text{acc}} + 1$ 
15:     $T \leftarrow T \cdot \gamma$ 
16:  $P_{\text{acc}} \sim n_{\text{acc}} / N_{\text{iter}}$ 

```

---



	$N_h = 3$			$N_h = 4$			$N_h = 5$		
	$N_{\text{iter}} = 10^4$	$N_{\text{iter}} = 10^5$	$N_{\text{iter}} = 10^6$	$N_{\text{iter}} = 10^4$	$N_{\text{iter}} = 10^5$	$N_{\text{iter}} = 10^6$	$N_{\text{iter}} = 10^4$	$N_{\text{iter}} = 10^5$	$N_{\text{iter}} = 10^6$
$E_{\min}$	-8.501610	-8.535389	-8.535646	-8.505239	-8.535555	-8.535661	-8.502652	-8.535519	-8.535858
$\varepsilon_r$	0.48585%	0.09046%	0.08745%	0.47668%	0.08851%	0.08728%	0.47366%	0.08894%	0.08497%

**Table 3:** Minimal energies found using simulated annealing.  $N = 4$  spins.

decreasing the temperature, the final temperature will be  $T_f = T_0 \cdot \gamma^{N_{\text{iter}}}$ . For example, with  $N_{\text{iter}} = 10^4$  iterations,  $T_f = 20 \cdot 0.999^{10^4} \simeq 9.03 \cdot 10^{-4}$ .

Notice that the relative errors are similar to our divide and conquer strategy, which means that we can obtain the same precision much more efficiently, at least in this case.

## 4.4 Limits of the Simulated Annealing scheme

As we just mentioned, our current approach for finding the ground state of the Hamiltonian (eq.3.2), involves generating new sets of parameters by randomly displacing the previous one, and then calculating the mean energy (see alg.1) over all possible spin states, and this has a computational cost of  $\mathcal{O}(2^N)$ .

In this section, we discuss the ground state energies found for the cases  $N = 5, 10, 12$  spins by diagonalizing the Hamiltonian as well as using simulated annealing, with varying number of hidden units. For that, we use  $N_h = \{N/2, N, 2N\}$  for each case. In addition, we measure the computational time and display it with the results in a table.

	$N = 5$			$N = 10$			$N = 12$		
	$N_h = 2$	$N_h = 5$	$N_h = 10$	$N_h = 5$	$N_h = 10$	$N_h = 20$	$N_h = 6$	$N_h = 12$	$N_h = 24$
$N_{\text{iter}} = 10^4$ :									
$E_{\min}$	-10.624692	-10.623765	-10.580834	-21.249803	-21.086341	-21.247562	-25.499926	-25.25018	-25.496572
$\varepsilon_r$	0.24406%	0.25277%	0.65585%	0.10063%	0.86910%	0.11117%	0.09786%	1.06999%	0.11189%
$t_{\text{comp}}$	1 s	1 s	2 s	72 s	63 s	73 s	814 s	857 s	949 s
$N_{\text{iter}} = 10^5$ :									
$E_{\min}$	-10.627724	-10.627971	-10.638927	-21.250639	-21.251775	-21.252391			
$\varepsilon_r$	0.21560%	0.21328%	0.11041%	0.09670%	0.09137%	0.08847%			
$t_{\text{comp}}$	9 s	12 s	17 s	610 s	629 s	720 s			
$N_{\text{iter}} = 10^6$ :									
$E_{\min}$	-10.628560	-10.628666	-10.639801	-21.251199	-21.252261	-21.252520			
$\varepsilon_r$	0.20774%	0.20675%	0.10220%	0.09407%	0.08908%	0.08787%			
$t_{\text{comp}}$	90 s	121 s	173 s	6238 s	5966 s	6983 s			

**Table 4:** Minimal energies found using simulated annealing.  $N = 5, 10, 12$  spins.

As one can see from Table 4, as the number of spins increases, the computational time  $t_{\text{comp}}$  grows exponentially, to the point where obtaining better relative errors is not worth the cost ( $N = 12$  case). Still, the results found are better than those obtained using the divide and conquer strategy.

In the next section, we introduce a different method for optimizing the RBM parameters  $\theta$ , known as the **Natural Gradient Descent**. While this method does not guarantee convergence

to the global minimum, we will see that the obtained energies are comparable with the ones found with Simulated Annealing, at least for the cases considered.

## 4.5 Natural Gradient Learning

A popular learning method in the field of neural networks (and machine learning in general) is the **stochastic gradient method** [1], used to minimize a proposed loss function, usually the energy of the system. It differs from the ordinary gradient descent algorithm, in the sense that, instead of computing the gradient of the entire dataset, it computes the gradient using only a subset of the data, chosen randomly. This makes it faster and more scalable for large datasets, though it introduces randomness in the optimization process.

However, in many cases, the parameter space is not **Euclidean** but has a **Riemannian** metric structure. In these cases, the ordinary gradient does not give the steepest direction of the loss function; rather, the direction of steepest descent is given by the *natural gradient* [8].

### 4.5.1 Natural Gradient

Let  $\mathcal{P} = \{\mathbf{w} \in \mathbf{R}^n\}$  be a parameter space in which we define a loss function  $L(\mathbf{w})$ . In Euclidean space, the coordinate system  $\mathbf{w}$  is orthonormal, and the squared length between  $\mathbf{w}$  and  $\mathbf{w} + d\mathbf{w}$  is given by

$$|\mathbf{w} + d\mathbf{w} - \mathbf{w}|^2 = |d\mathbf{w}|^2 = \sum_{i=1}^n (dw_i)^2 \quad (4.21)$$

with  $dw_i$  the components of  $d\mathbf{w}$ . However, when the coordinate system is Riemannian (nonorthonormal), the squared length is given by the quadratic form

$$|d\mathbf{w}|^2 = \sum_{i,j} g_{ij} dw_i dw_j, \quad \text{with } g_{ij} \text{ the metric of the space.} \quad (4.22)$$

In such a case we call  $\mathcal{P}$  a curved manifold.

---

**Proposition 4.5** (Parameter spaces of neural networks). The parameter spaces of neural networks have a *Riemannian* character [8]. Namely, their parameters define a curved manifold  $\mathcal{P}$  in which the squared length between infinitesimal vectors is given by Equation 4.22.

---

The steepest descent direction of our loss function  $L(\mathbf{w})$  is defined by the vector  $d\mathbf{w}$  that minimizes  $L(\mathbf{w} + d\mathbf{w})$  where we fix the length of  $|d\mathbf{w}|^2$  to  $\varepsilon^2$ , for a sufficiently small  $\varepsilon$ .

---

**Theorem 4.6** (Steepest direction in Riemannian space). The steepest descent direction of  $L(\mathbf{w})$  in a Riemannian space is given by

$$-\tilde{\nabla} L(\mathbf{w}) = -\mathbf{G}^{-1} \nabla L(\mathbf{w}) \quad (4.23)$$

where  $\mathbf{G}^{-1}$  is the inverse of the metric  $\mathbf{G} = (g_{ij})$  and  $\nabla L$  is the conventional gradient.

---

**Proof.** Let  $d\mathbf{w} = \varepsilon \mathbf{a}$ , with  $\mathbf{a}$  such that it minimizes

$$L(\mathbf{w} + d\mathbf{w}) = L(\mathbf{w}) + \varepsilon \nabla L(\mathbf{w})^T \mathbf{a}$$

Recall the constraint that  $|d\mathbf{w}|^2 = \varepsilon^2$ , which implies

$$|\mathbf{a}|^2 = \sum g_{ij} a_i a_j = 1$$

Using the Lagrange multipliers to incorporate this constraint, we have

$$\frac{\partial}{\partial a_i} \{ \nabla L(\mathbf{w})^T \mathbf{a} - \lambda \mathbf{a}^T \mathbf{G} \mathbf{a} \} = 0,$$

which gives,

$$\nabla L(\mathbf{w}) = 2\lambda \mathbf{G} \mathbf{a}$$

or,

$$\mathbf{a} = \frac{1}{2\lambda} \mathbf{G}^{-1} \nabla L(\mathbf{w})$$

where  $\lambda$  comes from the constraint.

We define  $\tilde{\nabla} L(\mathbf{w}) = \mathbf{G}^{-1} \nabla L(\mathbf{w})$ , the natural gradient of  $L$  in the Riemannian space.

Theorem 4.6 suggests a natural gradient descent algorithm of the form

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta_t \tilde{\nabla} L(\mathbf{w}_t) = \mathbf{w}_t - \eta_t \mathbf{G}^{-1} \nabla L(\mathbf{w}_t) \quad (4.24)$$

where  $\eta_t$  is the learning rate.

#### 4.5.2 Stochastic Reconfiguration Learning

Generally, the choice of  $\mathbf{G}$  in Equation 4.24 depends on the metric of the space in which the problem is defined. In our case, we will be following the procedure described in [10] & [11], which is specifically tailored for variational Monte Carlo in a quantum system.

In these references, it is seen that an adequate choice of  $\mathbf{G}$  is the Fubini-Study metric

$$d(\psi, \phi) = \arccos \left( \sqrt{\frac{\langle \psi | \phi \rangle \langle \phi | \psi \rangle}{\langle \psi | \psi \rangle \langle \phi | \phi \rangle}} \right) \quad (4.25)$$

also known as the **quantum angle**  $\gamma(\psi, \phi)$ . For normalized states  $|\phi\rangle$  and  $|\psi\rangle$ , it simplifies to:

$$\gamma(\psi, \phi) = \arccos(|\langle \psi | \phi \rangle|) \quad (4.26)$$

This expression can be expanded to second order in small parameter changes  $d\mathbf{w}$  as  $\gamma(\psi_{\mathbf{w}}, \psi_{\mathbf{w}+d\mathbf{w}})$ , which we will see next.

This specific application of the natural gradient method tailored for variational quantum problems is called **stochastic reconfiguration**.

## Minimization of expectation value of the energy

Going back to the Transversal Field Ising model, our objective is to optimize the values of the RBM parameters  $\theta$  that minimize the energy of our trial wavefunction in Equation 4.8. In other words, given  $|\psi_\theta\rangle = \langle \mathcal{S} | \Psi(\vec{s}; \theta) \rangle$ , find  $\theta_{\text{opt}}$  such that

$$\theta_{\text{opt}} = \arg \min_{\theta} \frac{\langle \psi_\theta | \hat{H} | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle}$$

**Definition 4.7** (Expectation value of an observable). We define the expectation value of any quantum mechanical observable  $\hat{O} = \hat{O}^\dagger$  as

$$\frac{\langle \psi | \hat{O} | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{\langle \psi | \mathbf{1} \hat{O} \mathbf{1} | \psi \rangle}{\langle \psi | \psi \rangle} \quad (4.27)$$

assuming  $|\psi\rangle$  is normalized:

$$= \sum_{\mathbf{s}} \sum_{\mathbf{s}'} \langle \psi | \mathbf{s} \rangle \langle \mathbf{s} | \hat{O} | \mathbf{s}' \rangle \langle \mathbf{s}' | \psi \rangle \quad (4.28)$$

$$= \sum_{\mathbf{s}} \sum_{\mathbf{s}'} \langle \mathbf{s} | \psi \rangle^* \langle \mathbf{s} | \hat{O} | \mathbf{s}' \rangle \langle \mathbf{s}' | \psi \rangle \quad (4.29)$$

$$= \sum_{\mathbf{s}} \sum_{\mathbf{s}'} \Psi(\mathbf{s})^* \langle \mathbf{s} | \hat{O} | \mathbf{s}' \rangle \Psi(\mathbf{s}') \quad (4.30)$$

$$= \sum_{\mathbf{s}} \sum_{\mathbf{s}'} \Psi(\mathbf{s})^* \frac{\Psi(\mathbf{s}')}{\Psi(\mathbf{s})} \langle \mathbf{s} | \hat{O} | \mathbf{s}' \rangle \Psi(\mathbf{s}') \quad (4.31)$$

$$= \sum_{\mathbf{s}} \Psi(\mathbf{s})^* \Psi(\mathbf{s}) \sum_{\mathbf{s}'} \langle \mathbf{s} | \hat{O} | \mathbf{s}' \rangle \frac{\Psi(\mathbf{s}')}{\Psi(\mathbf{s})} \quad (4.32)$$

$$= \sum_{\mathbf{s}} \rho(\mathbf{s}) O_{\text{loc}} \doteq \langle O_{\text{loc}} \rangle \quad (4.33)$$

where  $\mathbf{s}, \mathbf{s}'$  are complete basis and the local value of the observable at configuration  $\mathbf{s}$  is defined as

$$O_{\text{loc}}(\mathbf{s}) = \sum_{\mathbf{s}'} \langle \mathbf{s} | \hat{O} | \mathbf{s}' \rangle \frac{\Psi(\mathbf{s}')}{\Psi(\mathbf{s})} \quad (4.34)$$

Notice that the expectation value in Equation 4.33, can be approximated with a set of configurations obtained from a Markov chain Monte Carlo (MCMC) sampling of the wavefunction:

$$\frac{\langle \psi | \hat{O} | \psi \rangle}{\langle \psi | \psi \rangle} \approx \frac{1}{N_{MC}} \sum_{i=1}^{N_{MC}} O_{\text{loc}}(\mathbf{s}_i) \quad (4.35)$$

where, in our case,  $\mathbf{s}_i$  corresponds to sampled spin states.

**Definition 4.8** (Log-derivative operator). We define the log-derivative of a variational wave-

function with respect to its  $k$ -th parameter as

$$\Delta_k(\mathbf{s}) \doteq \frac{1}{\Psi(\mathbf{s}; \boldsymbol{\theta})} \partial_{\theta_k} \Psi(\mathbf{s}; \boldsymbol{\theta}) \quad (4.36)$$

Note that, in the case of complex-valued parameters, one has to take into account the complex conjugate

$$\Delta_k^*(\mathbf{s}) \doteq \frac{1}{\Psi^*(\mathbf{s}; \boldsymbol{\theta})} \partial_{\theta_k^*} \Psi^*(\mathbf{s}; \boldsymbol{\theta}) \quad (4.37)$$

In our case, we consider complex parameters and wavefunctions, namely  $\theta_i \in \mathbb{C}$  and  $\psi(\boldsymbol{\theta}) \in \mathbb{C}$ . To minimize the energy, which is a real-valued function, we use both the derivative of the parameters  $\partial_{\theta_k}$  and the derivative of the complex-conjugated parameters  $\partial_{\theta_k^*}$  [11]. Namely:

$$\partial_{\theta_k^*} \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{(\partial_{\theta_k^*} \langle \psi |) \hat{H} | \psi \rangle + \langle \psi | \hat{H} (\partial_{\theta_k^*} | \psi \rangle)}{\langle \psi | \psi \rangle} - \frac{[(\partial_{\theta_k^*} \langle \psi |) | \psi \rangle + \langle \psi | (\partial_{\theta_k^*} | \psi \rangle)] \langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle^2} \quad (4.38)$$

Now we use the fact that  $\partial_{\theta_k^*} |\psi(\boldsymbol{\theta})\rangle = 0$

$$\partial_{\theta_k^*} \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle} = \frac{(\partial_{\theta_k^*} \langle \psi |) \hat{H} | \psi \rangle + 0}{\langle \psi | \psi \rangle} - \frac{[(\partial_{\theta_k^*} \langle \psi |) | \psi \rangle + 0] \langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle^2} \quad (4.39)$$

$$= \langle \Delta_k^* E_{\text{loc}} \rangle - \langle \Delta_k^* \rangle \langle E_{\text{loc}} \rangle \quad (4.40)$$

We denote the gradient of the energy as

$$F_k = \langle \Delta_k^* E_{\text{loc}} \rangle - \langle \Delta_k^* \rangle \langle E_{\text{loc}} \rangle \quad (4.41)$$

$$= \langle \Delta_k^* (E_{\text{loc}} - \langle E_{\text{loc}} \rangle) \rangle \quad (4.42)$$

As mentioned before, the stochastic reconfiguration method is a technique that makes use of the information encoded in the quantum geometric tensor (QGT) to *precondition* the gradient used in stochastic optimization. It improves the convergence rate to the ground state of a Hamiltonian  $\hat{H}$  [11].

This QGT can be derived from the Fubini-Study distance in Equation 4.26, considering its infinitesimal version [11], [14], [23]. That is, for small perturbations of the parameters  $\delta\boldsymbol{\theta}$ , we denote  $|\psi(\boldsymbol{\theta})\rangle$  as  $|\psi\rangle$  and  $|\psi(\boldsymbol{\theta} + \delta\boldsymbol{\theta})\rangle$  as  $|\psi + \delta\psi\rangle$ . Then, the Fubini-Study distance between them is

$$\gamma(\psi, \psi + \delta\psi) = \arccos \left( \sqrt{\frac{|\langle \psi | \psi + \delta\psi \rangle|^2}{\|\psi\|^2 \|\psi + \delta\psi\|^2}} \right) \quad (4.43)$$

Assuming  $\psi$  is normalized, then the inner product  $|\langle\psi|\psi + \delta\psi\rangle|^2$  can be written

$$|\langle\psi|\psi + \delta\psi\rangle|^2 = |\langle\psi|\psi\rangle + \langle\psi|\delta\psi\rangle|^2 = |1 + \langle\psi|\delta\psi\rangle|^2 \quad (4.44)$$

$$= 1 + 2 \operatorname{Re} \langle\psi|\delta\psi\rangle + |\langle\psi|\delta\psi\rangle|^2 \quad (4.45)$$

On the other hand, the denominator can be expanded up to second order as

$$\frac{1}{\|\psi + \delta\psi\|^2} \simeq 1 - 2 \operatorname{Re} \langle\psi|\delta\psi\rangle - \|\delta\psi\|^2 + 4[\operatorname{Re} \langle\psi|\delta\psi\rangle]^2 \quad (4.46)$$

Then, inside the arccos, we have

$$\frac{|\langle\psi|\psi + \delta\psi\rangle|^2}{\|\psi + \delta\psi\|^2} \simeq 1 + |\langle\psi|\delta\psi\rangle|^2 - \|\delta\psi\|^2 \quad (4.47)$$

Finally, making use of the approximation  $\arccos \sqrt{1-x} \simeq \sqrt{x}$  for  $x \rightarrow 0^+$ , we obtain

$$\gamma(\psi, \psi + \delta\psi) \simeq \arccos\left(\sqrt{1 - (\|\delta\psi\|^2 - |\langle\psi|\delta\psi\rangle|^2)}\right) \simeq \sqrt{\|\delta\psi\|^2 - |\langle\psi|\delta\psi\rangle|^2} \quad (4.48)$$

or, if we square this distance

$$ds^2 \equiv \gamma(\psi, \psi + \delta\psi)^2 \simeq \|\delta\psi\|^2 - |\langle\psi|\delta\psi\rangle|^2 \quad (4.49)$$

However, it is interesting to write this final expression in terms of our variational derivatives  $\Delta_k$  as

$$S_{kk'} \doteq \langle\Delta_k^* \Delta_{k'}\rangle - \langle\Delta_k^*\rangle \langle\Delta_{k'}\rangle \quad (4.50)$$

$S_{kk'}$  denoting the quantum geometric tensor of our metric.

We finally have the update rule for our variational parameters, as seen below

$$\boldsymbol{\theta}_{t+1} \leftarrow \boldsymbol{\theta}_t - \delta t \mathbf{S}^{-1} \mathbf{F} \quad (4.51)$$

For our RBM ansatz, the variational derivatives of the parameters have the following expressions:

$$\begin{aligned} \log(\Psi(s)) &= \sum_{j=1}^N b_j s_j + N_h \log(2) + \sum_{i=1}^{N_h} \log\left(\cosh\left(c_i + \sum_{j=1}^N \omega_{ij} s_j\right)\right) \\ \Delta_{b_j}(s) &= \frac{\partial}{\partial b_j} \log(\Psi(s)) = s_j \end{aligned} \quad (4.52)$$

$$\Delta_{c_i}(s) = \frac{\partial}{\partial c_i} \log(\Psi(s)) = \tanh\left(c_i + \sum_{j=1}^N \omega_{ij} s_j\right) \quad (4.53)$$

$$\Delta_{\omega_{ij}}(s) = \frac{\partial}{\partial \omega_{ij}} \log(\Psi(s)) = s_j \tanh\left(c_i + \sum_{j=1}^N \omega_{ij} s_j\right) \quad (4.54)$$

With them, we can implement this strategy with some functions, seen in Algorithms 3, 4 & 5.

This strategy does not involve the evaluation of the entire spin basis  $\mathcal{S}$ , nor the calculation of the Hamiltonian matrix, so we should expect some speedup in the computation at the cost of some precision.

---

**Algorithm 3** Natural Gradient descent (part 1)

---

```

1: function FLIP( $\vec{s}, i$ ) ▷ Flips  $i$ -th spin
2:    $\vec{s}' \leftarrow \vec{s}$ 
3:    $s'_i \leftarrow 1 - s_i$ 
4:   return  $\vec{s}'$ 

5: function LOCALENERGY( $\vec{s}, \theta$ ) ▷ Local energy of state  $\vec{s}$ 
6:   couplings  $\leftarrow 0$ 
7:   for all  $s_i \in \vec{s}$  do
8:     if  $s_i == s_{i+1}$  then ▷ Including periodic boundary condition
9:       couplings  $\leftarrow$  couplings + 1
10:    else
11:      couplings  $\leftarrow$  couplings - 1
12:   energy_interaction  $\leftarrow -J \cdot$  couplings
13:   prob_ratios  $\leftarrow 0$ 
14:    $\Psi_\theta(\vec{s}) \leftarrow$  WAVEFUNCTION( $\vec{s}, \theta$ )
15:   for  $i \leftarrow 1$  to length( $\vec{s}$ ) do
16:      $\vec{s}' \leftarrow$  FLIP( $\vec{s}, i$ )
17:      $\Psi_\theta(\vec{s}') \leftarrow$  WAVEFUNCTION( $\vec{s}', \theta$ )
18:     prob_ratios  $\leftarrow$  prob_ratios +  $|\Psi_\theta(\vec{s}')|^2 / |\Psi_\theta(\vec{s})|^2$ 
19:   energy_field  $\leftarrow -B \cdot$  prob_ratios
20:   return energy_interaction + energy_field

21: function VARIATIONALDERIVATIVES( $\vec{s}, \theta$ )
22:    $\vec{b}, \vec{c}, \mathbf{W} \leftarrow \theta$ 
23:    $\Delta_{\vec{b}} \leftarrow \vec{s}$ 
24:    $\Delta_{\vec{c}} \leftarrow \tanh.(\vec{c} + \mathbf{W} \cdot \vec{s})$  ▷  $\tanh.()$  meaning element-wise  $\tanh()$ 
25:    $\Delta_{\mathbf{W}} \leftarrow \Delta_{\vec{c}}^\top \cdot \Delta_{\vec{b}}$ 
26:   append  $\{\Delta_{\vec{b}}, \Delta_{\vec{c}}, \Delta_{\mathbf{W}}\}$  to  $\Delta_\theta$ 
27:   return  $\Delta_\theta$ 

```

---

Algorithm 3 contains the functions

- **flip**: Takes the  $i$ -th spin in state  $\vec{s}$  and flips it.
- **localEnergy**: Calculates the local energy of state  $\vec{s}$  as

$$E_{\text{loc}}(\vec{s}) = \frac{\hat{H} |\Psi(\vec{s}; \theta)\rangle}{|\Psi(\vec{s}; \theta)\rangle}$$

- **variationalDerivatives**: Calculates the variational derivatives of the RBM ansatz defined in Equations 4.52-4.54.

---

**Algorithm 4** Natural Gradient descent (part 2)

---

```

28: function COVARIANCE( $\vec{s}_1, \vec{s}_2$ )
29:   samples  $\leftarrow$  length( $\vec{s}_1$ )
30:   mean1  $\leftarrow$  SUM( $\vec{s}_1$ )/samples
31:   mean2  $\leftarrow$  SUM( $\vec{s}_2$ )/samples
32:   mean12  $\leftarrow$   $\vec{s}_1^\top \cdot \vec{s}_2$  / samples
33:   return mean12 – mean1 · mean2

34: function SAMPLEBLOCK( $\theta$ , block_size,  $\vec{s}_0$ )
35:    $\vec{b}, \vec{c}, \mathbf{W} \leftarrow \theta$ 
36:    $n_v \leftarrow$  length( $\vec{b}$ )
37:   if  $\vec{s}_0 == \{\}$  then
38:      $\vec{s} \leftarrow$  random state ▷ Initialize  $\vec{s}$  as a random  $(0, 1)$  array with  $n_v$  elements
39:   states  $\leftarrow \{\}$ 
40:    $\Psi_\theta(\vec{s}) \leftarrow$  WAVEFUNCTION( $\vec{s}, \theta$ )
41:   for  $i \leftarrow 1$  to block_size do
42:     idx  $\leftarrow$  choose from 1 to  $n_v$ 
43:      $\vec{s}' \leftarrow$  FLIP( $\vec{s}$ , idx)
44:      $\Psi_\theta(\vec{s}') \leftarrow$  WAVEFUNCTION( $\vec{s}', \theta$ )
45:      $u \leftarrow$  UNIFORM(0, 1) ▷ Uniform random variable
46:     if  $|\Psi_\theta(\vec{s}')|^2 / |\Psi_\theta(\vec{s})|^2 \geq u$  then
47:        $\vec{s} \leftarrow \vec{s}'$ 
48:        $\Psi_\theta(\vec{s}) \leftarrow \Psi_\theta(\vec{s}')$ 
49:     append  $\Psi_\theta(\vec{s})$  to states
50:   return states

```

---

The function `covariance` in Algorithm 4 is used to calculate both the gradient  $\mathbf{F}$  and the QGT  $\mathbf{S}$ . For instance, to calculate the latter, we call the function as `covariance( $\Delta_\theta^*, \Delta_\theta$ )` since

$$\mathbf{S} = \langle \Delta_\theta^* \Delta_\theta \rangle - \langle \Delta_\theta^* \rangle \langle \Delta_\theta \rangle$$

In addition, the function `sampleBlock` samples `block_size` states starting from  $\vec{s}_0$  by randomly flipping its spins, and using  $|\Psi(\vec{s}; \theta)|^2$  as a probability distribution. The acceptance criteria of a sampled state is the same Metropolis algorithm used in Definition 4.4.



---

**Algorithm 5** Natural Gradient descent (part 3)

---

**Input:**  $\theta$ ,  $N_{\text{iter}}$ ,  $\lambda$ ,  $n_{\text{blocks}}$ , thermalization, block\_size

**Outputs:**  $\theta_{\text{min}}$ ,  $E_{\text{min}}$

```

51:  $N_B \leftarrow n_{\text{blocks}} - \text{thermalization}$ 
52: for iteration  $\leftarrow 1$  to  $N_{\text{iter}}$  do
53:   local_energies  $\leftarrow \{\}$ 
54:   derivatives  $\leftarrow \{\}$ 
55:   states  $\leftarrow \text{SAMPLEBLOCK}(\theta, \text{thermalization} \cdot \text{block\_size}, \vec{s}_0 = \{\})$ 
56:   state  $\leftarrow$  last element of states
57:   for  $k \leftarrow 1$  to  $N_B$  do
58:     block  $\leftarrow \text{SAMPLEBLOCK}(\theta, \text{block\_size}, \vec{s}_0 = \text{state})$ 
59:     state  $\leftarrow$  last element of block
60:     append  $\text{LOCALENERGY}(\text{state}, \theta)$  to local_energies
61:     append  $\text{VARIATIONALDERIVATIVE}(\text{state}, \theta)$  to derivatives
62:    $\mathbf{F} \leftarrow \text{COVARIANCE}(\text{derivatives}^\dagger, \text{local\_energies})$ 
63:    $\mathbf{S} \leftarrow \text{COVARIANCE}(\text{derivatives}^\dagger, \text{derivatives})$ 
64:    $\theta \leftarrow \theta - \lambda(\mathbf{S}^+ \cdot \mathbf{F})$   $\triangleright \mathbf{S}^+$  denotes the pseudo-inverse
65:  $\theta_{\text{min}} \leftarrow \theta$ 
66:  $E_{\text{min}} \leftarrow \text{SUM}(\text{local\_energies})/N_B$ 

```

---

Lastly, Algorithm 5 is the main part of the code, where the stochastic reconfiguration learning scheme is done. To summarize:

1. At each learning iteration we let the system thermalize by sampling a number of states starting from a random initial state  $\vec{s}_0$ .
2. Then, we choose the last state of this thermalization to start the Markov chain of block samples. We sample in blocks of a certain size to avoid picking states that are correlated. Next, we use the last sample of each block to calculate its local energy  $E$  and its variational derivatives  $\Delta_\theta$ .
3. After  $N_B$  iterations of block sampling, we calculate the gradient  $\mathbf{F}$  and the QGT  $\mathbf{S}$  by first averaging the results of the sampling, and then using Equations 4.41 & 4.50.
4. We update the parameters with the update rule in Equation 4.51. Here,  $\lambda$  acts as the learning rate, and  $\mathbf{S}^+$  is defined as the pseudo-inverse of  $\mathbf{S}$  to avoid numerical instability.
5. At the end of the learning process, the final average of the local energies is taken as the ground state energy of the RBM, and the final value of  $\theta$  as the optimized parameters of the variational wavefunction.

## 5 Results

In this section we will show and discuss the results obtained after training the RBM wavefunction with stochastic reconfiguration (SR), and compare the results with those of the exact diagonalization of the problem.

### 5.1 Varying number of spins and hidden units.

First, let us repeat the problem of table 4, this time using SR instead of simulated annealing. For the same values of the number of spins  $N$  and hidden units  $N_h$ , we obtain the following results:

	$N = 5$			$N = 10$			$N = 12$		
	$N_h = 2$	$N_h = 5$	$N_h = 10$	$N_h = 5$	$N_h = 10$	$N_h = 20$	$N_h = 6$	$N_h = 12$	$N_h = 24$
$N_{\text{iter}} = 125$ :									
$E_{\min}$	-10.473004	-10.625631	-10.610826	-20.841273	-20.795139	-20.815894	-24.723145	-24.867071	-24.927460
$\varepsilon_r$	1.66820%	0.23518%	0.37418%	2.02122%	2.23810%	2.14052%	3.14195%	2.57810%	2.34151%
$t_{\text{comp}}$	8 s	8 s	9 s	16 s	17 s	21 s	25 s	20 s	24 s
$N_{\text{iter}} = 250$ :									
$E_{\min}$	-10.510069	-10.621143	-10.648947	-21.250289	-21.251689	-21.271210	-25.387857	-25.484380	-25.486505
$\varepsilon_r$	1.32020%	0.27731%	0.01626%	0.09835%	0.09177%	0.09741%	0.53781%	0.15966%	0.15133%
$t_{\text{comp}}$	16 s	17 s	18 s	33 s	36 s	41 s	35 s	38 s	49 s
$N_{\text{iter}} = 375$ :									
$E_{\min}$	-10.517643	-10.645326	-10.648711	-21.254030	-21.254890	-21.271210	-25.421970	-25.504995	-25.525133
$\varepsilon_r$	1.24908%	0.05026%	0.01848%	0.08076%	0.07672%	0.09420%	0.40416%	0.07889%	0.05900%
$t_{\text{comp}}$	25 s	27 s	27 s	49 s	53 s	60 s	53 s	57 s	72 s

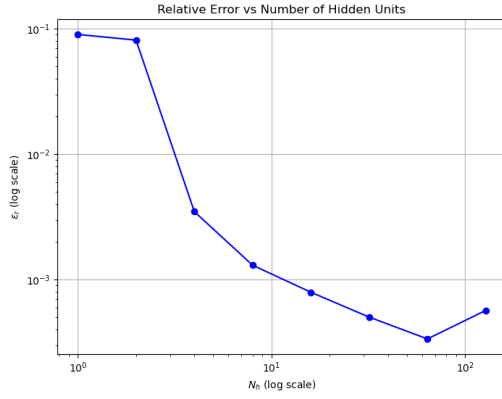
**Table 5:** Minimal energies found using stochastic reconfiguration.  $N = 5, 10, 12$  spins.

Notice how some of these results are even better than those obtained using simulated annealing in Table 4, specially those results obtained for higher number of learning iterations. Nevertheless, the important point is that we were able to obtain these results using this (computationally) much cheaper method, as well as being able to estimate the ground state energy for the  $N = 12$  case quite fast.

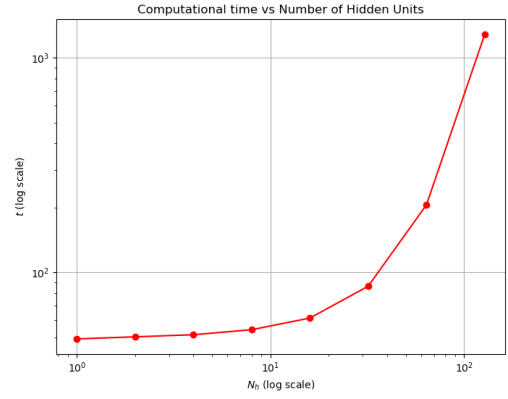
In fact, we expect to be able to generalize this to a much higher number of spins, with the only drawback being that we cannot diagonalize the corresponding Hamiltonian in order to calculate its exact ground state energy for comparison. In these cases, we can make use of Equation 3.54 to find the ground state energy.

Furthermore, we can plot how the relative error of the minimal energy decreases with the complexity of the trial wavefunction; i.e., by increasing the number of hidden units of the RBM wavefunction, we should expect to find lower energies after the learning process, if we have not already found the exact ground energy.

We can see in Figure 6a how the error indeed decreases with increasing  $N_h$ . However, there is a certain point at which we have diminishing returns, since increasing the number of hidden units also increases the computational cost of our algorithm (see fig.6b). Notice how at  $N_h = 128$ , not only does the error increase, but the computational time explodes from  $\sim 4$  to  $\sim 15$  minutes.



(a) Relative error of the ground state energy (Eq.4.19).



(b) Computational time of the algorithm.

**Figure 6:** Stochastic reconfiguration plots for a trial wavefunction with  $N = 12$  spins and varying number of hidden units. The values in the plot are averages of 10 iterations.

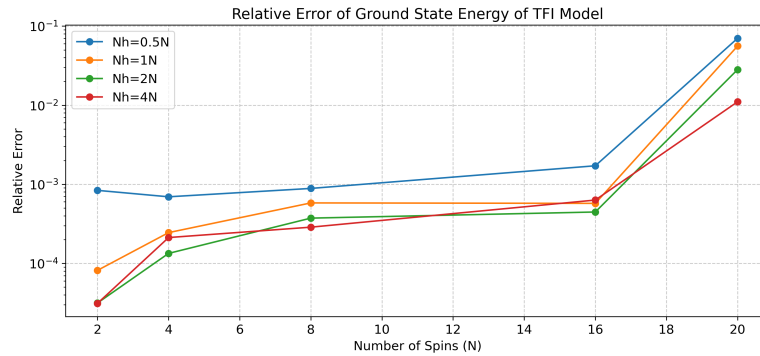
## 5.2 Ground state comparison

In this section, we compare the ground state found using stochastic reconfiguration with the exact ground state of the Ising transverse field model, by calculating the errors between the ground state energies and the magnetizations of the two states.

However, as I mentioned a few times throughout the document, we can only compare these results for  $N \lesssim 20$  as we cannot diagonalize the Hamiltonian  $\hat{H}$  for larger values (except the ground state energy, which can be calculated with Equation 3.54).

### 5.2.1 Ground state energies and magnetizations

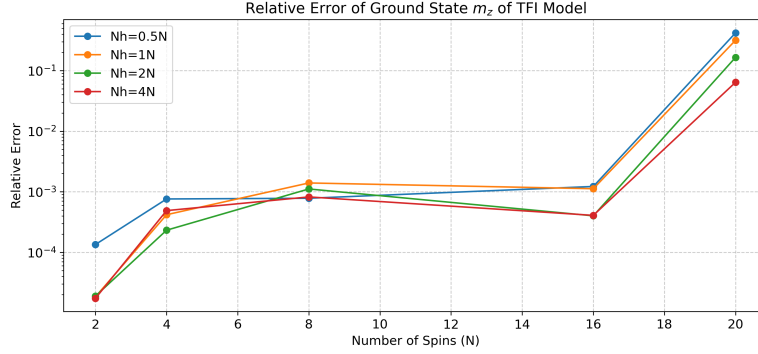
First, we plot the relative error of the ground energy (Eq.4.7) and magnetizations of the RBM wavefunction (Eqs.3.57&3.58). For the energy, we report both the exact value and the RBM ground state with SR learning, varying both  $N$  and  $N_h$ ; while keeping the rest of the learning parameters constant ( $N_{\text{iter}}, \lambda, n_{\text{blocks}}, \dots$ ) and setting  $J = 2, B = 1$ . We plot the relative error of their energies, longitudinal and transversal magnetizations.



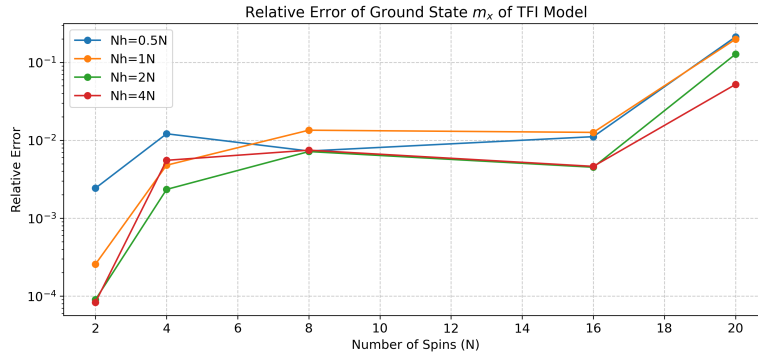
**Figure 7:** Relative error between exact and RBM ground state energies.  $J = 2, B = 1$ .

Figure 7 shows that the ground state energy found with the RBM wavefunction is quite similar to the exact ground state energy. In the  $N = 20$  case, when the number of hidden units is

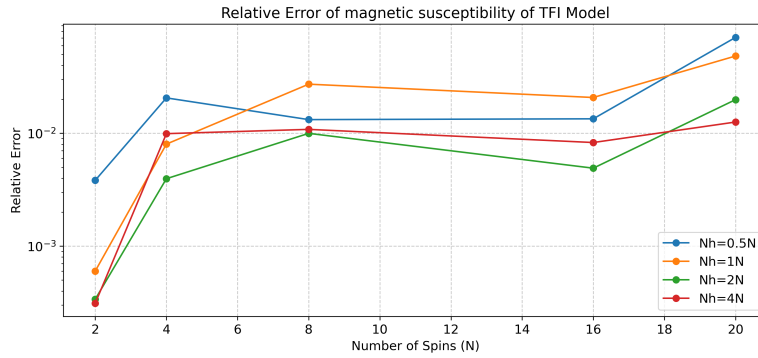
at least twice the number of spins, the error is less than 5%. The remaining error can be the residual of the stochastic exploration.



**Figure 8:** Relative error of the longitudinal magnetization  $m_z$ .  $J = 2$ ,  $B = 1$ .



**Figure 9:** Relative error of the transversal magnetization  $m_x$ .  $J = 2$ ,  $B = 1$ .

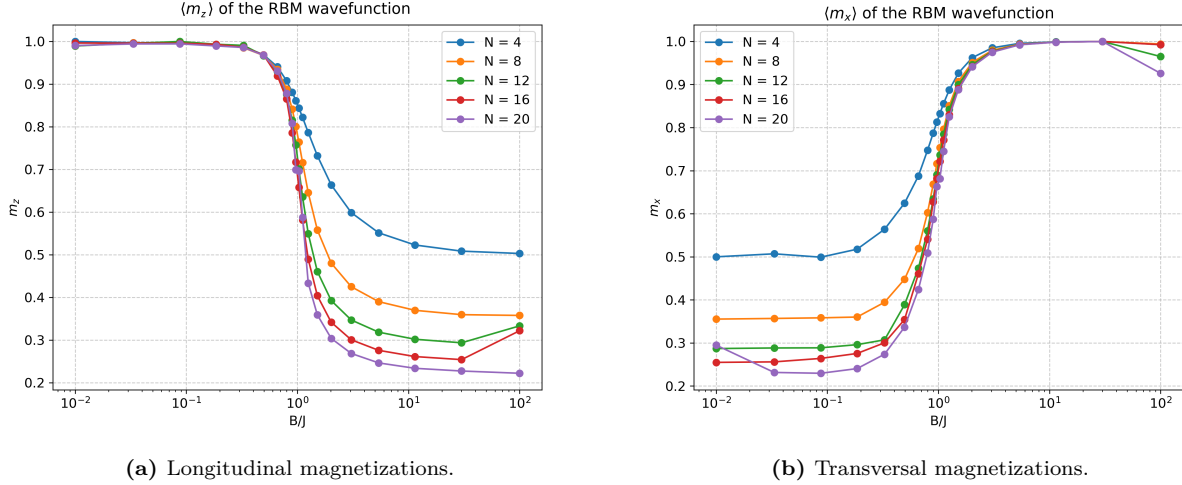


**Figure 10:** Transversal magnetic susceptibility error between exact and RBM ground states.  $J = 2$ ,  $B = 1$ .

Both Figures 8 & 9 show the magnetization of the RBM ground state, compared with the magnetizations that we calculated back in Figure 3. Increasing the complexity (number of hidden neurons) of the RBM seems to cause the overall errors to decrease. In this case we kept the learning parameters constant; i.e., number of iterations, block size, learning rate, etc, but one can see the effects of varying the number of iterations in Table 5.

Lastly, we can see in Figure 10 that the behaviour is similar to the one seen in Figure 9. The main difference being that the values of the relative errors are slightly larger for  $\chi_x$  at small values of  $N$ .

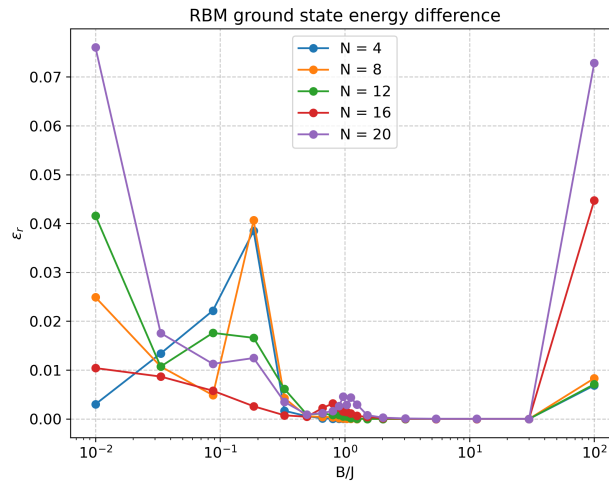
Next, following the same procedure as in Figure 3, we can plot the magnetizations of the RBM ground state versus the ratio  $B/J$  in order to observe the phase transition that occurs at the critical point  $B = J$ . In this case, we will plot different curves for different values of the number of spins  $N$ , and we will keep  $N_h = 4N$  constant in all cases.



**Figure 11:** RBM ground state magnetizations with respect to  $B/J$ .  $N_h = 4N$  in all cases.

Figure 11 shows the longitudinal and transversal magnetizations of the ground state of the RBM trial wavefunction with respect to the ratio  $B/J$ . In them, we can see that our estimated ground state undergoes the same phase transition as the actual ground state of the system at  $B/J = 1$ .

However, one could notice that for higher  $N$ , the values of the magnetizations have some errors on the edges of the plot. If we plot the relative energy difference of the RBM wavefunction with the exact ground state at these values (seen in Figure 12) we can confirm that the ground energy found with the RBM is not exactly  $E_0$ , thus explaining their difference in the magnetizations.

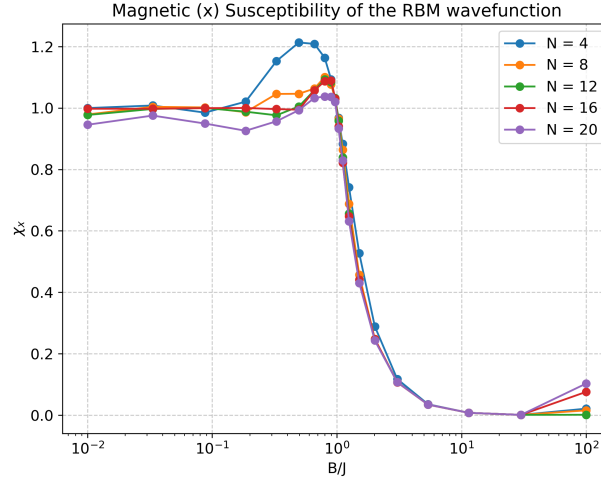


**Figure 12:** Energy difference between RBM and exact ground state with respect to  $B/J$ .

Note for instance, the point  $N = 20, B/J = 10^2$ , where the energy of the RBM state is  $\sim 7\%$

over the actual minimum energy. The same point can be identified in Figure 11b, showing an error of  $\sim 8\%$ .

Finally, we can plot the magnetic susceptibility of the RBM ground state and compare it with Figure 4.



**Figure 13:** Magnetic susceptibility.

Comparing Figure 13 with Figure 4, we observe how the phase transition for the RBM ground state behaves as in the exact ground state, showing the same tendency towards  $B/J = 1$  as  $N$  increases.

### 5.2.2 Ground state distances and overlaps

In the previous experiments, we have seen how our RBM wavefunction obtained good approximations for both the ground state energy and the magnetizations. In addition, the RBM state reproduced the phase transition when varying the ratio  $B/J$  in the Hamiltonian of the system. With these experiments, one would conclude that our variational method has accurately estimated the exact ground state of the transverse field Ising problem. However, from the variational principle [3], we know that the error of the mean energy is of **second order** in small variations of the trial wavefunction. Namely, if

$$|\psi_{\theta}\rangle = |\psi_0\rangle + |\delta\psi\rangle \quad (5.1)$$

then, we can decompose the error into:

$$|\delta\psi\rangle = |\delta\psi_{\parallel}\rangle + |\delta\psi_{\perp}\rangle = \alpha |\psi_0\rangle + |\delta\psi_{\perp}\rangle \quad (5.2)$$

which leads to an energy

$$E[\psi] = \frac{E_0|1 + \alpha|^2 + \langle\delta\psi_{\perp}|\hat{H}|\delta\psi_{\perp}\rangle}{|1 + \alpha|^2 + \langle\delta\psi_{\perp}|\delta\psi_{\perp}\rangle} = E_0 + \mathcal{O}(\delta\psi_{\perp})^2 \quad (5.3)$$

This is a very powerful theorem, since even a rather simple trial wavefunction can obtain good approximations to the ground state energy. Thus, it begs the question of whether our RBM wavefunction can properly describe the ground state of the problem, or if the errors are low due to the nature of the variational method.

What we can do is directly compare the exact ground state with the one found using SR on the RBM wavefunction for small values of  $N$ :

- For  $N = 2$ :

The exact ground state is found by diagonalizing the Hamiltonian in Equation 4.13, and it gives us the state seen in Equation 4.17. For  $J = 2$ ,  $B = 1$ , the normalized ground state is

$$|\psi_0\rangle \simeq \begin{pmatrix} 0.68819 \\ 0.16246 \\ 0.16246 \\ 0.68819 \end{pmatrix} \quad (5.4)$$

and the one we found using stochastic reconfiguration ( $N_h = 4N$ ,  $N_{\text{iter}} = 375$ , and the other adjusted learning parameters) on the RBM trial wavefunction is

$$|\psi_\theta\rangle = \begin{pmatrix} 0.68983 + 0.01332j \\ 0.16244 + 0.00234j \\ 0.16244 + 0.00234j \\ 0.68639 + 0.00654j \end{pmatrix} \quad (5.5)$$

We can immediately see that it is very similar to the exact ground state, and the RBM state almost verifies the symmetry under spin direction  $|\uparrow\rangle \leftrightarrow |\downarrow\rangle$ . If we calculate the distance between the two states (in Fubini-Study metric) we find

$$\gamma(\psi_0, \psi_\theta) = \arccos\left(\sqrt{\langle\psi_0|\psi_\theta\rangle \langle\psi_\theta|\psi_0\rangle}\right) = 0.005372 \text{ rad} = 0.307835^\circ \quad (5.6)$$

This means that our RBM state is almost parallel to the exact ground state of the TFI model for  $N = 2$  spins.

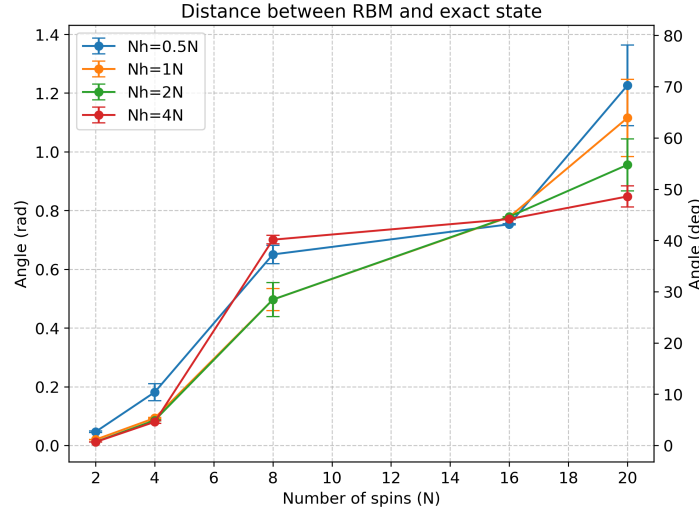
- For  $N = 4$ :

The normalized ground states for  $N = 4$  have  $2^4 = 16$  components, so instead of comparing them component by component, we can calculate the angle between the exact and RBM ground states for  $J = 2$ ,  $B = 1$  as

$$\gamma(\psi_0, \psi_\theta) = 0.083835 \text{ rad} = 4.803394^\circ \quad (5.7)$$

And, while they are close, they are not exactly parallel.

We also show in Figure 14 the angular distance between the RBM state and the exact ground state as a function of the number of spins  $N$ , for different values of the number of hidden units  $N_h$ .



**Figure 14:** Fubini-Study distance between exact and RBM ground states.

Looking at Figure 14, we can see that, if the RBM wavefunction does not have enough complexity, the ground state that we find at the end of the learning tends to be almost perpendicular to the exact ground state ( $70 \pm 10^\circ$ ) for  $N = 20$ . In quantum mechanical terms, this is close to an excited state of the system, since all Hamiltonian eigenstates form an orthonormal basis:  $\langle \psi_n | \psi_m \rangle = \delta_{nm}$ , and it would explain the errors in the energies and magnetizations.

In the  $N_h = 4N$  case, the angle between the RBM ground state and the exact one is between  $45^\circ$  to  $50^\circ$  for  $N = 20$ . It is important to highlight, however, that  $\gamma(\psi_0, \psi_\theta)$  is non-linear with respect to the projection  $|\langle \psi_0 | \psi_\theta \rangle|$  and that an angle of  $45^\circ$  corresponds to an overlap of  $\cos 45^\circ \sim 70.7\%$ .

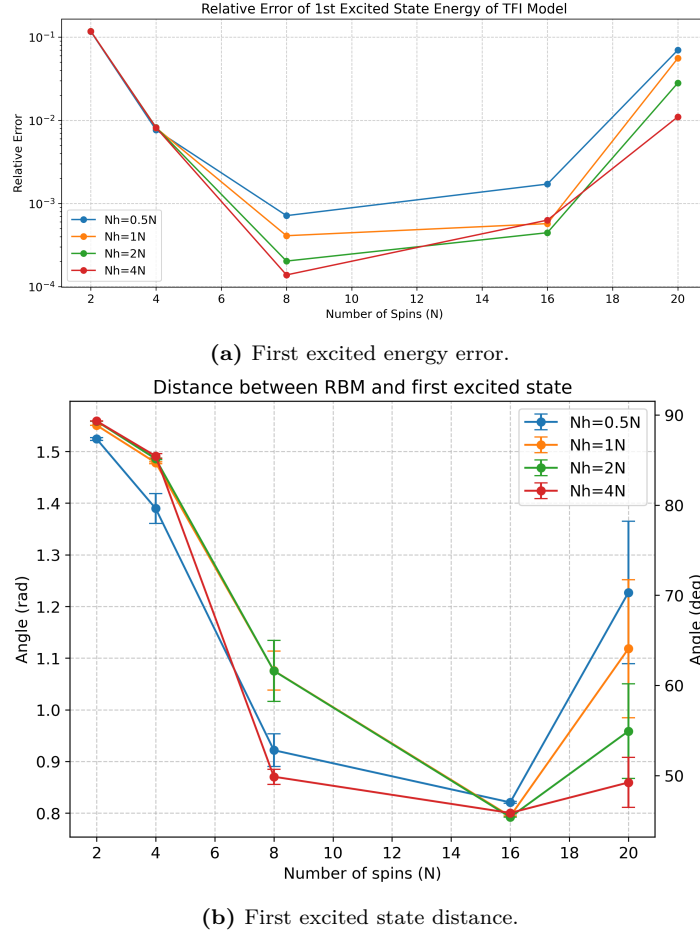
## Discussion of the Results

The results of the energy errors in Figure 7 and the quantum angles between the exact and RBM ground states in Figure 14 show interesting discrepancies. These differences may appear due to the fact that the variational principle used in optimizing the RBM parameters focuses on minimizing the energy expectation value. This means the learning process is aiming to achieve the lowest possible energy, but the algorithm might be getting trapped in local minima where the wavefunction is not an accurate representation of the true ground state.

To check this, let us take the same results from the experiments used in Figures 7 & 14, and also calculate the relative error of the RBM energy with the first and second excited states of the Hamiltonian in Equation 3.53.

Figure 15a provides some valuable insights on the previous results, in that the relative errors between the RBM energies and the exact energies of the ground and first excited states become increasingly similar when  $N$  increases. When  $N$  is small, the difference between the ground state energy and the first excited energy is relatively large. However, as  $N$  grows larger, this difference becomes insignificant (for  $N = 20$ , the relative difference between  $E_0$  and  $E_1$  is  $1.01 \cdot 10^{-8}$ ), to the point where finding a state that has an energy close to  $E_0$ , will also be close to  $E_1$ .





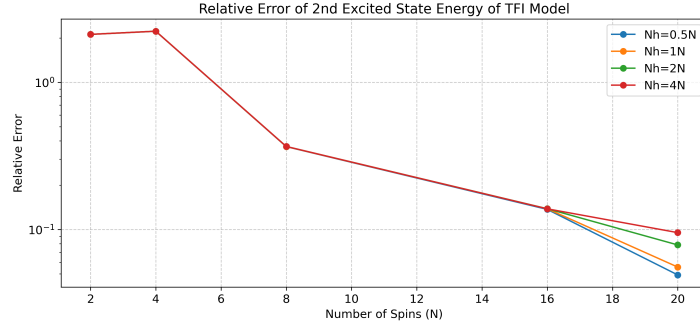
**Figure 15:** Relative error between RBM and first excited state.  $J = 2$ ,  $B = 1$ .

Notice now in Figure 15b how the RBM states found for  $N_h = 2N$ ,  $4N$  are also at an angle of  $52 \pm 5^\circ$  with the first excited state. As expected, our algorithm has converged to a local minimum where the found RBM state is a combination of the ground and first excited states. The reason, as we mentioned, may be that the algorithm only focuses in minimizing the energy and does not consider any symmetries or properties that only the ground state may have. So, when  $N$  is large, the RBM state tends to be close to both  $\psi_0$  and  $\psi_1$ .

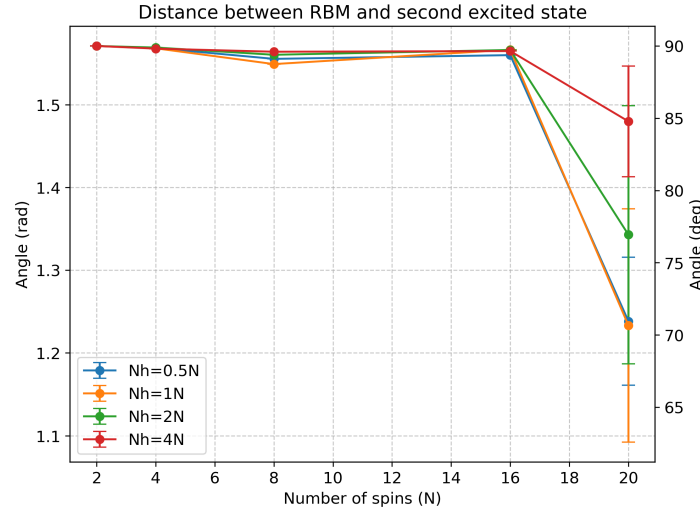
Figure 16 on the other hand, shows the comparison of the RBM state with the second excited state. In this case, the difference between  $E_0$  and  $E_2$  is larger even with  $N$  large (for  $N = 20$ , the relative difference between  $E_0$  and  $E_2$  is 0.0963). Notice that the cases  $N_h = 0.5N$ ,  $N_h = N$ , obtained states closer to the second excited state than the ground state. One could expect that, in these cases, the RBM converges to a combination of higher excited states.

Overall, this inability of the spins to simultaneously minimize their interaction energy with their neighbours is highly resemblant of the behaviour of **quantum spin-glass** systems, in which the resulting ground state is highly degenerate with many nearly equivalent configurations. One property of these spin-glass systems, for instance, is that they are characterized by frustrated interactions and randomness, which generally lead to complex energy landscapes with loads of local minima.

In conclusion, the trial wavefunction that we decided on might be neglecting the fidelity of the state representation. Namely, the translational symmetry of the problem is not reflected



(a) Second excited energy error.



(b) Second excited state distance.

**Figure 16:** Relative error between RBM and second excited state.  $J = 2$ ,  $B = 1$ .

on the variational wavefunction, so every parameter may take different values in the complex plane. Also, discussing the possibility of  $\psi_0$  having nodes and changes of sign may lead to other restrictions on the RBM parameters that could help speeding up the convergence, or decreasing the discrepancy between the energy and quantum angle errors as the size of the problem increases.

## 6 Budget

The following budget outlines the estimated costs sustained during the course of this research project, which primarily involved extensive online research, programming and periodic consultations with my thesis' advisors.

**1. Transportation costs:** Fuel costs for commuting to university.

- Number of trips: 10 (once every other week).
- Distance per trip:  $38.7 \times 2 \text{ km} = 77.4 \text{ km}$ .
- Fuel cost per km:  $6.9 \text{ l}/100 \text{ km} \times 1.622 \text{ €/l} = 11.19 \text{ €/100 km}$
- Total transportation cost: 86.62 €.

**2. Internet access and utilities:** Internet subscription and electricity for computer use.

- Internet cost: 45.90 €/month.
- Electricity cost: 0.1781 €/kWh (average cost in a household).
- Computer usage: 450 W. At 10 hours a week, for 5 months, the total is  $\sim 90 \text{ kWh}$ .
- Total cost:  $45.90 \times 5 + 0.1781 \times 90 = 245.53 \text{ €}$ .

**3. Software:** Computer usage and software used.

- Depreciation cost of personal computer: laptop cost/average lifetime  $\times$  duration of the project  $\sim 66.58 \text{ €}$ .
- Python and related libraries: Free (Open Source).
- Matlab license: Free (UPC License).

**4. Dedicated Hours:** Research and development hours.

- Total hours: 150 hours.
- Hourly rate: 20 €/hour.
- Total cost: 3000 €.

**5. Books and Online Resources:**

- Estimated cost: Free (used UPC student e-mail to access the documents).

**6. Stationery and Printing:**

- Estimated cost:  $0,10 \text{ € per page} \times \sim 60 \text{ pages} = 6 \text{ €}$ .

**Total Estimated Cost:** 3404.5 €.

The primary expenses for this project were indirect costs associated with maintaining internet connectivity, computer usage, and transportation for consultations. Given the reliance on open-source software and personal computing resources, the overall financial cost remained minimal.

## 7 Environment Impact

This section outlines the environmental impact of this research project. As I mentioned, this thesis primarily involved internet-based research, programming and commuting for meetings with my advisors.

### 1. Digital Activities: Energy consumption and accessing online resources.

- Energy consumption: Use of personal computer. The impact is relatively small but accumulates over extended periods.
- Data centers: Accessing online resources, involves connecting to data centers, which consume significant amounts of energy.

### 2. Transportation: Commuting to university.

- Fuel consumption: Commuting to university contributed to greenhouse gas emissions. The impact was relatively small due to the frequency of these commutes.

### 3. General Waste: Use of paper, and electronic devices.

- Paper and printing: Minimal use of paper and printing was involved, as most research materials were accessed and stored digitally.
- Electronic waste: I changed my laptop during the course of the project.

**Mitigation Efforts:** Efforts were made to minimize the environmental impact of this project by:

- Using energy-efficient computing practices, such as optimizing and parallelizing code to reduce computational load and energy usage.
- Accessing digital copies of research materials instead of printing.
- Where possible, opting for public transportation to reduce fuel consumption.
- Re-usage of my previous laptop to avoid contributing to electronic waste by giving it to another person.

In conclusion, while the environmental impact of this project is quite low, awareness and proactive measures were taken to minimize energy consumption and reducing waste.

## 8 Conclusions and future development

The primary goal of this thesis was to explore the idea and develop an RBM artificial neural network-based approach to solving spin quantum many-body problems, specifically the quantum transverse field Ising model in one dimension. This section summarizes the key findings and accomplished objectives of the research that were set at the beginning of the project.

### 8.1 Conclusions

#### Analytical Solution of the TFI model

The initial phase of the project involved learning about the TFI model, and solving it analytically to obtain some exact results. By applying Jordan-Wigner and Bogoliubov transformations, the Hamiltonian of the problem was diagonalized. This led to the expression of the exact ground state energy, and the values of the magnetizations for the TFI model ground state, serving as a reference for evaluating the performance of the RBM-based solutions.

#### Development of the RBM-based Variational Approach

A core objective was to make use of the variational method using RBMs to approximate the ground state of the model. This part involved defining the RBM wavefunction, and exploring a few simple cases where we could find the optimal parameters in order to minimize its ground state energy by differentiating or exploring the parameter space using Divide and Conquer. The errors found in these simple cases were quite good, leveraging the capability of the RBM wavefunction to capture complex quantum correlations with a relatively compact expression.

#### Optimization of the RBM Parameters

The next objective was to optimize the trial wavefunction parameters making use of the variational principle combined with machine learning methods. Namely, the optimization process aimed to minimize the energy expectation value and iteratively adjust the RBM parameters to converge towards the ground state.

The first approach consisted in using a Simulated Annealing scheme to randomly explore the parameter space, and slowly obtain lower energies by accepting the parameter changes according to a Metropolis algorithm. Although the method produced accurate results and ensured global convergence, the time required to achieve these results made it impractical for use in our experiments.

The second, and ultimately final approach, was a stochastic reconfiguration learning scheme. This technique was chosen for the experiments due to its efficiency and the accuracy of the results.

#### Validation of the Results

To validate the results obtained with the RBM-based wavefunction, the results were compared with those obtained from exact diagonalization methods. The comparison showed that the RBM approach could reproduce the ground state energies and magnetizations of the TFI model ground state for smaller ( $N \lesssim 20$ ) system sizes.

However, while the obtained ground state energy errors showed the potential of RBMs for solving spin quantum many-body problems, the discrepancies in the quantum angle between the RBM ground state and the true ground state suggest that further improvements are needed in the wavefunction representation. Future work could focus on integrating additional constraints on the RBM wavefunction to improve the fidelity of the quantum state representation, such as implementing translational symmetry on the parameters, or restricting them to the real numbers.

## 8.2 Future Work

This research has underscored the potential in applying RBMs to solve the TFI model. Nevertheless, there are several areas that remain open for future exploration, including the final objective of this thesis:

- Analyze the scalability of the RBM approach for larger system sizes, and extending the methodology to higher-dimensional systems or other quantum models with more complex interactions, such as the Heisenberg model on a square lattice.
- Explore some optimization techniques such as considering the translational symmetry of the problem in order to reduce the number of parameters needed, and thus decrease the computational cost of the method.
- Trying other machine learning frameworks, such as deep neural networks, to see if they could further improve the accuracy and efficiency of quantum state approximations.

In conclusion, this thesis has shown the potential of using restricted boltzmann machine neural networks as variational wavefunctions in order to solve spin quantum many-body problems, providing a foundation for future research in the intersection between machine learning and quantum physics.

## References

- [1] Suvrit Sra, Sebastian Nowozin, and Stephen J. Wright, editors. *The Tradeoffs of Large Scale Learning*, page 351–368. Neural information processing. MIT Press, 2012.
- [2] H. Ishizuka, Y. Motome, N. Furukawa, and S. Suzuki. Quantum monte carlo study of the transverse-field ising model on a frustrated checkerboard lattice. *Journal of Physics: Conference Series*, 320:012054, Sep 2011.
- [3] Cohen Claude Tannoudji, Bernard Diu, and Franck Laloë. *Quantum Mechanics*. John Wiley & Sons, 2020.
- [4] Pierre Pfeuty. The one dimensional ising model with a transverse field. *Annals of Physics*, 57:79–90, Jul 1969.
- [5] Elliot Lieb, Theodore Schultz, and Daniel Mattis. Two soluble models of an antiferromagnetic chain. *Annals of Physics*, 15(3):407–466, Dec 1961.
- [6] Glen Bigan Mbeng, Angelo Russomanno, and Giuseppe E. Santoro. The quantum ising chain for beginners, Sep 2020.
- [7] D M Wood and A Zunger. A new method for diagonalising large matrices. *Journal of Physics A: Mathematical and General*, 18(9):1343–1359, Jun 1985.
- [8] Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, Feb 1998.
- [9] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, 355(6325):602–606, Feb 2017.
- [10] Sandro Sorella. Generalized lanczos algorithm for variational quantum monte carlo. *Physical Review B*, 64(2), Jun 2001.
- [11] Sandro Sorella. Wave function optimization in the variational monte carlo method. *Physical Review B*, 71(24), Jun 2005.
- [12] Jian-Jun Dong, Peng Li, and Qi-Hui Chen. The a-cycle problem for the transverse ising ring. *Journal of Statistical Mechanics: Theory and Experiment*, 2016, May 2016.
- [13] Asja Fischer and Christian Igel. Training restricted boltzmann machines: An introduction. *Pattern Recognition*, 47(1):25–39, 2014.
- [14] Ran Cheng. Quantum geometric tensor (Fubini-Study metric) in simple quantum system: A pedagogical introduction. December 2010.
- [15] Leon N. Cooper. Bound electron pairs in a degenerate fermi gas. *Phys. Rev.*, 104:1189–1190, Nov 1956.
- [16] Yusuke Nomura. Boltzmann machines and quantum many-body problems. June 2023.
- [17] Nobuyuki Yoshioka, Wataru Mizukami, and Franco Nori. Solving quasiparticle band spectra of real solids using neural-network quantum states. *Commun. Phys.*, 4(1), May 2021.

- 
- [18] Zhih-Ahn Jia, Biao Yi, Rui Zhai, Yu-Chun Wu, Guang-Can Guo, and Guo-Ping Guo. Quantum neural network states: A brief review of methods and applications. *Advanced Quantum Technologies*, 2(7–8), March 2019.
  - [19] Stephen R Clark. Unifying neural-network quantum states and correlator product states via tensor networks. *J. Phys. A Math. Theor.*, 51(13):135301, April 2018.
  - [20] Jing Chen, Song Cheng, Haidong Xie, Lei Wang, and Tao Xiang. Equivalence of restricted boltzmann machines and tensor network states. *Phys. Rev. B.*, 97(8), February 2018.
  - [21] I Affleck, I, T Kennedy, E H Lieb, and H Tasaki. Rigorous results on valence-bond ground states in antiferromagnets. *Phys. Rev. Lett.*, 59(7):799–802, August 1987.
  - [22] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Ann. Phys. (N. Y.)*, 349:117–158, October 2014.
  - [23] Ran Cheng. Quantum geometric tensor (fubini-study metric) in simple quantum system: A pedagogical introduction, 2013.
  - [24] Borja Requena. Introduction to variational monte carlo with neural network quantum states, Mar 2021.
  - [25] Anastasiia A. Pervishko and Jacob Biamonte. Pushing tensor networks to the limit, May 2019.