

Neural Network Thermodynamics

by
Dennis Wagner



A Thesis submitted to

the School of Mathematics and Natural Sciences of
The University of Wuppertal
in partial fulfillment of the requirements of the degree of
Dr. rer. nat.

and

the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements of the degree of
Doctor of Philosophy
Department of Physics and Astronomy of the University of Manitoba, Winnipeg

Copyright © 2024 by Dennis Wagner
Wuppertal, August 26, 2024

Abstract

In this work, the variational Monte Carlo (VMC) method using the restricted Boltzmann machine (RBM), a feed-forward neural network, is investigated to allow for the computation of finite-temperature results of quantum spin models in one and two dimensions. To accomplish this task, there are, at least, three different approaches possible that are discussed and compared for the Heisenberg model in one dimension before they are deployed to the Heisenberg and the J_1 - J_2 model in two dimensions. In the latter case, the correct detection of the presumed finite-temperature phase transition is studied. The three different types of methods studied in a variational setting are comprised by the purification method, the sampling method, and the use of variational thermal pure quantum states and thus the notion of quantum typicality. In the proceeding of this study, the reason for choosing the RBM as the predestined variational ansatz is motivated by general (artificial) neural network theory and investigations and findings related to the ground state optimization method. In the purification and the sampling method, finite temperature correlation functions for the one-dimensional Heisenberg model are computed successfully and even the free energy is attainable. Unfortunately, the idea of using quantum typicality in a combined variational ansatz, using the RBM and a pair-product ansatz, is shown to lead to inaccurate results at low temperatures due to the high entanglement entropy of those variational states. The issue seems to be persistent for any variational ansatz wave function and renders the applicability of quantum typicality in a variational setting highly problematic. To complete this investigation, the real-time evolution at infinite temperature is studied and the unfavourable scaling of the number of hidden units documented. Finally, the purification and the sampling method are applied to the J_1 - J_2 model on the square lattice. Unfortunately, the computation of accurate results for very low temperatures is difficult due to an increase in the rejection probability, an issue related to the Monte Carlo sampling and known as critical slowing down.

Zusammenfassung

In dieser Arbeit wird die variationelle Monte-Carlo-Methode unter Verwendung der *restricted Boltzmann Machine* (RBM), einem neuronalen *Feedforward*-Netz, untersucht, um die Berechnung von thermodynamischen Variablen bei endlicher Temperatur in Quantenspin-Modellen in einer und zwei Dimensionen zu ermöglichen. Dazu sind mindestens drei verschiedene Ansätze anwendbar, die zunächst für das eindimensionale Heisenberg-Modell diskutiert werden, bevor sie auf das zweidimensionale Heisenberg- und das J_1 - J_2 -Modell angewendet werden. Im letzten Fall wird dann der vermutete Phasenübergang bei endlicher Temperatur untersucht. Die drei verwendeten Methoden umfassen die *Purification*-Methode, die *Sampling*-Methode und die Verwendung von variationellen, typischen Quantenzuständen. Der Grund für die Wahl des RBM als Variationsansatz wird in dieser Studie durch die allgemeine Theorie der künstlichen neuronalen Netze und die Untersuchungen und Erkenntnisse im Zusammenhang mit der Grundzustandsoptimierungsmethode motiviert. Mit der *Purification*- und der *Sampling*-Methode können endliche Temperaturkorrelationsfunktionen im eindimensionalen Heisenberg-Modell erfolgreich berechnet werden. Sogar die Bestimmung der freien Energie ist möglich. Leider zeigt sich, dass die Idee, die typischen Zustände in einem kombinierten variationellen Ansatz, bestehend aus dem RBM und einem *pair-product* Ansatz, zu verwenden, aufgrund der hohen Verschränkungsentropie dieser Zustände zu ungenauen Ergebnissen bei niedrigen Temperaturen führt. Das Problem scheint allgemein für jeden Variationsansatz zu bestehen und stellt die Anwendbarkeit typischer Zustände in einem variationellen Zugang grundsätzlich in Frage. Um diese Untersuchung abzurunden, wird die Echtzeitentwicklung bei unendlicher Temperatur durchgeführt und die Skalierung der Anzahl der *hidden units* dokumentiert. Zuletzt werden die *Purification*- und die *Sampling*-Methode mit Erfolg auch auf das J_1 - J_2 Modell auf dem Quadratgitter angewendet. Leider verhindert ein Anstieg der Ablehnungswahrscheinlichkeit im Monte-Carlo-Algorithmus, dass genaue Ergebnisse für niedrige Temperaturen berechnet werden können. Dieses Problem hängt mit der Art des Monte-Carlo-Samplings zusammen und ist bekannt als kritische Verlangsamung.

Acknowledgements

I am extremely grateful to my supervisors Andreas Klümper and Jesko Sirker, whose guidance, feedback and insights were key for this project carried out. I am deeply honoured to have had the opportunity to learn from Andreas Klümper and to benefit from his profound understanding of quantum many-body physics, his passion for science and his critical thinking.

I am also deeply indebted to Jesko Sirker, firstly, for proposing the topic of this work and this way allowing me to discover the application of neural networks in the field of quantum many-body physics. Furthermore, his expertise in both numerical and analytical methods in this field of research has been indispensable for the success of this project.

I would of course also like to acknowledge the funding and opportunities afforded to me by my supervisors and the DFG Research Unit 2316. I would like to thank everyone in the process of establishing the Collaborative Ph.D. program and thus allowing me to benefit from the academic opportunities at two great universities. I gratefully acknowledge the PhD fellowship awarded to me by the University of Wuppertal. Special thanks to the Digital Research Alliance of Canada and the University of Manitoba for providing the high performance computing resources needed for this project.

Furthermore, I want to thank Yahya Öz, Eyzo Stouten, Mouhcine Azhari and Henrik Jürgens, Andrew Urichuk, Kyle Monkman and Maximilian Emmanouilidis for the fruitful discussions and insights into their research topics.

I would like to express my deepest gratitude to my wife Judith Wagner and my entire family for emotional support and strongly required distraction from physics in general and the up and downs of this project in specific.

List of Symbols

Symbol	Meaning
\mathcal{H}	Hilbert space
H	Hamiltonian
\mathcal{S}	Set
$\mathcal{P}(\mathcal{H})$	Projective Hilbert space
\mathbb{P}	Set of real parameters
\mathbb{W}	Set of complex parameters
$\mathcal{V}(\mathcal{H}, \mathbb{R}), \mathcal{V}(\mathcal{H}, \mathbb{C})$	Variational submanifold of the Hilbert space over the field of real, complex numbers
$\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{R}), \mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$	Variational submanifold of the projective Hilbert space over the field of real, complex numbers
A	Arbitrary operator
A_Ψ	Expectation value of an operator A with respect to the state $ \Psi\rangle$
$\langle A \rangle_{\Psi_{\mathbb{W}}}$	Functional of an operator A with respect to the variational state $ \Psi_{\mathbb{W}}\rangle$ parametrized by a set of parameters \mathbb{W}
G	Quantum geometric tensor
F	Forces
N	Total number of spins and/or visible nodes
\mathbf{v}	Visible nodes
M	Number of hidden nodes
\mathbf{h}	Hidden nodes
$\alpha = \frac{M}{N}$	Number of hidden nodes in terms of visible nodes
N_H	Number of hidden layers
T	Temperature
β	Inverse temperature $1/T$
e	Energy per site
f	Free energy per site
χ	Magnetic susceptibility
M	Magnetization
\mathcal{H}	Magnetic field strength
B	Magnetic flux density
c	Specific heat capacity
\mathcal{L}	Quantum-number projector
N_{Met}	Number of Monte Carlo samples

Contents

1	Introduction	1
2	Physical Framework	4
2.1	Magnetism - The Exchange Interaction	4
2.2	Heisenberg Model	6
2.3	J_1 - J_2 Model	8
2.4	Hubbard Model	8
2.5	Marshall-Peierls Sign Rule and <i>Stoquastic</i> Hamiltonians	9
2.6	Phase Transitions and Critical Exponents	10
3	Quantum Geometry	13
3.1	Projective Hilbert Space and Tangent Space	14
3.2	Variational Manifold	15
3.2.1	Variational Manifold of the Projective Hilbert Space	15
3.2.2	Variational Manifold of the Hilbert Space	19
3.3	Quantum Natural Gradient (QNG)	19
3.4	Quantum Velocity	20
4	Variational Monte Carlo Method	22
4.1	Ground State Optimization	22
4.1.1	Stochastic Gradient Descent (SGD)	24
4.1.2	Quantum Natural Gradient (QNG)	26
4.2	Time Evolution	27

4.2.1	Imaginary-Time Evolution	28
4.2.2	Real-Time Evolution	29
4.3	Adaptive Heun Method	29
4.4	Imaginary-Time Evolution on the Manifold and β Modification	30
4.5	Regularization	32
4.5.1	Ground State Optimization	33
4.5.2	Time Evolution	34
5	Neural Network (NN)	36
5.1	Hilbert's 13-th Problem	36
5.2	Kolmogorov-Arnold Representation Theorem	37
5.3	Single-Layer Feed-Forward Neural Network (SLFFNN)	38
5.4	Multi-Layer Feed-Forward Neural Network (MLFFNN)	40
5.5	Feed-Forward Neural Network (FFNN) over the Complex Field	45
5.6	Neural Network Quantum State (NNQS)	45
5.6.1	Sigmoidal Ansatz (SA)	46
5.6.2	Boltzmann Machine (BM)	46
5.7	Effective Update Scheme	49
5.8	Details of the Learning Algorithm	50
5.9	Curse of Dimensionality	51
5.10	Entanglement Entropy and Expressive Power of RBMs	52
6	Variational Wave Functions for Strongly Correlated Systems	54
6.1	H ₂ -Molecule	54
6.1.1	Heitler-London Ansatz	55
6.1.2	Hartree-Fock Ansatz	55
6.2	Many-Body Ansatz Wave Function	55
6.2.1	Hartree-Fock Ansatz	56
6.2.2	Gutzwiller and Jastrow Factor	56
6.3	Pair-Product (PP) Wave Function	57

6.3.1	Bardeen-Cooper-Schrieffer (BCS) Wave Function	57
6.3.2	Resonating Valence Bond (RVB) Wave Function	58
6.3.3	Pair-Product (PP) Wave Function	58
6.3.4	Efficient Computation of the Pfaffian and its Derivative	60
6.4	Transformation from Fermions to Spins	61
6.5	PP-NN Ansatz	61
7	Ground State Optimization with Feed-Forward Neural Networks (FFNNs)	63
7.1	Barren Plateaus	64
7.2	Improvement by the PP-NN Ansatz	64
7.3	Exact Representations with Deep Boltzmann Machines (DBMs)	65
7.4	Wave Function for Finite-Temperature Computations	66
8	Finite-Temperature Results for the One-Dimensional Heisenberg Model	67
8.1	Purification Method	68
8.1.1	Purification	68
8.1.2	Macroscopic Thermodynamic Variables	70
8.1.3	Numerical Results	71
8.2	Sampling Method	77
8.2.1	Minimally Entangled Thermal States (METS)	77
8.2.2	Numerical Results	78
8.3	Thermal Pure Quantum State (TPQS)	81
8.3.1	Measure Concentration on Normed Spaces	81
8.3.2	Infinite-Temperature Thermal Pure Quantum State (TPQS)	82
8.3.3	Extension to Finite Temperatures	84
8.3.4	Canonical Thermal Pure Quantum State (CTPQS)	84
8.3.5	Variational Canonical Thermal Pure Quantum State (VCTPQS)	85
8.3.6	PP-NN Ansatz	86
8.3.7	The Infinite-Temperature VCTPQS	86
8.3.8	Numerical Results	90

9	Dynamical Correlation Functions for the One-Dimensional Heisenberg Model at Infinite Temperature	92
9.1	Spin-Spin Autocorrelation Function and Entanglement Entropy	93
9.2	Numerical Results	93
10	Finite-Temperature Results for the Two-Dimensional J_1-J_2 Model on the Square Lattice	96
10.1	Ground State Phase Diagram	96
10.2	Modifications of the Purification and the Sampling Algorithm	97
10.3	Finite-Temperature Results	98
10.3.1	Finite-Temperature Results for the Two-Dimensional Heisenberg Model	98
10.3.2	Finite-Temperature Results for the Two-Dimensional J_1 - J_2 Model	100
10.4	Phase Transition at Finite Temperatures	101
10.5	Critical Slowing Down	103
11	Conclusion	108
12	Outlook	110
	Bibliography	127
	Appendix A Schrödinger Equation on the Variational Manifold of the Projective Hilbert Space	128
	Appendix B Schrödinger Equation on the Variational Manifold of the Hilbert Space	129

List of Figures

2.1	Crystal structure of copper benzoate [1].	7
2.2	Unit cell of the Ising model on a triangular lattice.	8
2.3	Phase diagram of the ground state of the J_1 - J_2 Heisenberg model [5].	12
3.1	The Bloch sphere [2].	13
4.1	Flowchart of the VMC algorithm.	25
5.1	Depiction of the SLFFNN described by Eq. (5.8).	40
5.2	Depiction of the MLFFNN described by Eq. (5.10).	42
5.3	Depiction of the MLFFNN described by Eq. (5.13).	44
5.4	Depiction of the RBM described by Eq. (5.26).	49
6.1	Illustration of four singlet pair patterns of the RVB wave function ansatz in two dimensions [3].	58
8.1	Inner and free energy for an open Heisenberg chain with $N = 7$ sites obtained using the purification algorithm without the β modification.	72
8.2	Imaginary time τ in the Heun algorithm versus number of iterative steps for an open Heisenberg chain with $N = 7$ sites.	72
8.3	Inner and free energy for an open Heisenberg chain with $N = 7$ sites obtained using the purification algorithm and Monte Carlo sampling of the Hilbert space.	73
8.4	Imaginary time τ in the Heun algorithm versus number of iterative steps for an open Heisenberg chain with $N = 7$ sites.	73
8.5	Inner and free energy compared to the QMC solution for an open Heisenberg chain with $N = 30$ sites obtained using the purification algorithm.	75

8.6	Imaginary time τ in the Heun algorithm versus number of iterative steps for an open Heisenberg chain with $N = 30$ sites.	75
8.7	Correlation functions $C_{\parallel,\perp}^2$ for an open Heisenberg chain with $N = 30$ sites obtained using the purification algorithm are compared to the QMC solution.	76
8.8	Inner and free energy compared to the QMC solution for an open Heisenberg chain of length $N = 30$ obtained using the sampling algorithm.	79
8.9	Imaginary time in the Heun algorithm for an open Heisenberg chain of length $N = 30$	79
8.10	Correlation functions $C_{\parallel,\perp}^2$ compared to the QMC solution for an open Heisenberg chain with $N = 30$ sites obtained using the sampling algorithm.	80
8.11	Illustration of the width within which 90% of randomly chosen states are located [4].	82
8.12	Square root of the mean of the absolute of the pairwise Pearson correlation coefficients of the wave function components for n different sets of parameters and a spin chain of length $N = 12$	88
8.13	Standard deviation of the operator $\langle S_0^z \rangle$ computed using 10^4 Metropolis samples and $n = 10^3$ different sets of parameters.	88
8.14	Entanglement entropy of the PP, the PP-local and the METS-NN initial states compared to the Gaussian case for different chain lengths of an open Heisenberg chain.	89
8.15	Inner energy of the PP-NN and the PP-NN-local ansatz compared to the QMC solution for an open Heisenberg chain of length $N = 30$ obtained using the typicality algorithm.	90
8.16	Imaginary time τ in the Heun algorithm versus number of iterative steps for the PP-NN and the PP-NN-local ansatz for an open Heisenberg chain of length $N = 30$ using the typicality algorithm.	90
9.1	Autocorrelation function $\langle S_n^z(0)S_n^z(t) \rangle$ for $n = N/2$ and corresponding entanglement entropy at infinite temperatures for an open Heisenberg chain with $N = 4$ sites.	94
9.2	Autocorrelation function $\langle S_n^z(0)S_n^z(t) \rangle$ for $n = N/2$ and corresponding entanglement entropy at infinite temperatures for an open Heisenberg chain with $N = 6$ sites.	94
9.3	Autocorrelation function $\langle S_n^z(0)S_n^z(t) \rangle$ for $n = N/2$ at infinite temperatures for an open Heisenberg chain with $N = 8$ sites.	94
10.1	Inner and free energy for the periodic Heisenberg model on the square lattice with $N = 4 \times 4$ sites obtained using the purification and the sampling algorithm.	99
10.2	Imaginary time τ in the Heun algorithm versus number of iterative steps for the periodic Heisenberg model on the square lattice with $N = 4 \times 4$ sites obtained using the purification and the sampling algorithm.	99

10.3	Inner and free energy for the periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites obtained using the purification and the sampling algorithm.	100
10.4	Specific heat capacity for the periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites and $J_2 = 0.85$ obtained using the purification and the sampling algorithm.	100
10.5	Absolute value of the correlation function o_{vh} for a periodic J_1 - J_2 model on the square lattice for various system sizes using the sampling algorithm.	102
10.6	Specific heat capacity for a periodic J_1 - J_2 model on the square lattice with $J_2 = 0.85$ for various system sizes using the sampling algorithm.	102
10.7	Results for the energy, the specific heat capacity and the correlation function o_{vh} computed by the corner transfer matrix renormalization group (CTMRG) algorithm [6].	103
10.8	Energy for a periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm.	105
10.9	Rejection probability for a periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm.	105
10.10	Integrated autocorrelation time for for a periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm.	106

Chapter 1

Introduction

Simulating physical quantum many-body problems for finite system sizes larger than just a few particles is in general a very difficult task. Since the Hilbert space grows exponentially in system size, the computation of the real- and imaginary-time evolution is a complex task. Therefore, a number of numerical methods for different applications have been established. While the quantum Monte Carlo (QMC) method is well-established and provides accurate results for a number of models, it suffers from the sign problem, which is especially prevalent in frustrated spin models and fermionic models [7]. Thus, the density-matrix renormalization group (DMRG), which optimizes matrix-product states (MPS) efficiently, is the most accurate method in this field of research [8]. Unfortunately, it is not straightforward to generalize this approach to two dimensional models. However, there are various propositions for this generalization [9].

The Variational Monte Carlo (VMC) method was first proposed in the field of ground state optimization making use of the variational principle. A variational ansatz for the wave function is proposed and depends on a set of variational parameters. The idea is generalized by using a variational ansatz to approximate any state and any evolution of interest. While the reasoning for these ansatz wave functions has often been physically motivated, for example by peculiarities of the model at hand, the huge practical success of (artificial) neural networks ((A)NNs) in other fields has paved the way for a universal ansatz of the wave function. It is solely motivated by the universal approximation theorem. Hence, the ansatz is proven to be an approximator independent of the concrete nature of the wave function to be approximated. Furthermore, this ansatz is readily generalized to two-dimensional models.

In this work, the variational Monte Carlo (VMC) method using feed-forward neural networks (FFNNs) is investigated to allow for the computation of finite-temperature results of quantum spin models in one and two dimensions. To do so, there are, at least, three different approaches possible that are discussed for the Heisenberg model in one dimension before they are deployed to the J_1 - J_2 model in two dimensions. Additionally, the scaling of the number of variational parameters in the real-time evolution is studied.

Before diving into this topic, the interest in the covered spin models, namely the Heisenberg model and the J_1 - J_2 model, is motivated in Ch. 2. Furthermore, some of their peculiarities for the

computations at hand are explained. In Ch. 3, a geometric view on quantum physics, known as quantum geometry is presented. Therein, the embedding of variational wave functions is discussed. They can be viewed as elements of the variational submanifold of the (projective) Hilbert space. The peculiarities of this quantum space, especially its metric tensor known as the quantum geometric tensor, is used to define the (quantum) natural gradient upon this manifold. Therewith, an arbitrary state is optimized with respect to the expectation value of a physical operator, for example the energy, or evolved in real or imaginary time. The respective algorithm making use of the Monte Carlo sampling for larger system sizes is explained in detail in Ch. 4.

The first ideas for such an ansatz in the ground state optimization were inspired by physical considerations and expectations for the ground states of the models under consideration. These are discussed in Ch. 6. A well-known example is the so-called Jastrow-Gutzwiller factor and the pair-product (PP) wave function [10]. However, these are always restricted to the expected form of the wave function and therefore biased by what is the assumed outcome. The crucial difference in this approach comes with the use of a specific form of neural networks (NNs), namely feed-forward neural networks (FFNNs), for the ansatz wave function. Those are not tailored to a specific model and do not prescribe certain features but are universal approximators, a decisive feature that is explained in detail in Ch. 5. Furthermore, the effect of using multiple layers in deep neural networks is discussed and compared to the single layer case. The quantum states based on neural networks are named neural network quantum states (NNQS). The optimization of those states with respect to a specific loss function, for example the energy, is in neural network theory considered a recurrent learning algorithm and was first proposed in Ref. [11]. Due to the universal feature of FFNNs, there is a number of network architectures varying in depth and width available. They have been deployed in the ground state optimization method of spin models and are compared and elaborated on in chapter 7. This way, especially the potential benefit of using multi-layer FFNNs compared to single-layer FFNNs is explored and the restricted Boltzmann machine (RBM) identified as a suitable choice for the variational wave function in the finite-temperature computations carried out in this work. Thus, it is from there on referred to simply as the NN ansatz.

In Ch. 8, the NN ansatz is used to compute finite-temperature results for the one-dimensional Heisenberg model. In the purification algorithm, the NN ansatz is modified to allow for an accurate approximation of the infinite-temperature state. This modified ansatz is referred to as the modified NN (mNN) ansatz. The fine-tuning of this state is checked for small system sizes and the effect of the Monte Carlo sampling is studied. Subsequently, the method is applied to larger system sizes yielding results for the energy, the free energy and local correlation functions. Secondly, the sampling method is investigated for which the minimally entangled thermal states (METs) are expressed easily in the NN ansatz by selecting a particular choice of parameters. The results for the inner energy, the free energy and spin-spin correlation functions are comparable to the ones obtained in the purification method and also limited by the sampling error. In the final section of this chapter, the criteria for the representation of typical pure quantum states (TPQSs) by a variational ansatz are investigated and understood. Unfortunately, the unavoidable increase in the entanglement entropy is spotted as the decisive feature inhibiting the numerical computation of any correlation function for low temperatures in an accuracy that is competitive with the other two methods discussed, although, the quantum typicality approach of course provides more accurate results than the sampling method at high temperatures.

To conclude this discussion, the real-time evolution is considered in Ch. 9 using the purification

method at infinite temperature. Unfortunately, the number of hidden units, that is needed to provide accurate results for larger system sizes, grows much faster than linear, at least like $\propto N^4$. However, it is more likely to grow even exponentially.

In the last Ch. 10, the most promising two methods, the purification method and the sampling method, are applied to the two-dimensional Heisenberg and the two-dimensional J_1 - J_2 model to obtain finite-temperature results. Especially for the J_1 - J_2 model, the purification algorithm needs to be adjusted by a Hadamard transform to allow for an efficient sampling at high temperatures. The reason for a more substantial deviation of the finite-temperature results are discussed. The breakdown of the algorithm at low temperatures due to a failure of the Monte Carlo sampling is observed earlier than in the one-dimensional case. This renders the investigation of the presumed finite-temperature phase transition difficult. Furthermore, the results deviate even for temperatures larger than the critical temperature substantially from the numerical results in Ref. [6], which are obtained by the corner transfer renormalization group (CTMRG) method. This issue can, in principle, be overcome by a model-dependent cluster algorithm which hinders the increase of the rejection probability observed.

Chapter 2

Physical Framework

Before diving into the topics investigated in this study, a broader but short general introduction to this field of research is presented. It encompasses the categorization of magnetic materials and provides the rationale for focusing on the models pertinent to this study. A strict theoretical derivation starting from the most basic condensed matter Hamiltonian considering the ions, electrons and their interaction and adding to it the ideas of second quantization, is found in Ref. [12].

2.1 Magnetism - The Exchange Interaction

Magnetism is a pure quantum mechanical effect and therefore can not be explained by classical physics. This is proven by the Bohr-van Leeuwen theorem [13]. The magnetic behavior of solid states can be categorized into three main classes. The quantity these classes are distinguished by is known as the magnetic susceptibility $\chi_{\alpha\beta}(T, \mathcal{H}) = \frac{\partial M_\alpha}{\partial \mathcal{H}_\beta} (\alpha, \beta = x, y, z)$. It is a response function that describes how much a material will become magnetized (magnetization \mathbf{M}) in an applied external magnetic field \mathcal{H} and depends in general on the temperature T . Since the magnetism of solids is a many-body theory and thus can not be solved directly or analytically, the formation of models is unavoidable. A self-contained and very detailed treatment of the theory of magnetism is found in Ref. [13] while the most interesting models are reviewed in Ref. [14].

- **Diamagnetism:** Although all solid states show diamagnetic behavior, it is only observable for materials not dominated by the other two effects explained below. This is the case for models in which there exist no permanent magnetic moments. The diamagnetic behavior was first explained by an induction effect described by Lenz's law. However, this derivation contains inconsistencies and contradicts the Bohr-van Leeuwen theorem. A detailed discussion using purely quantum mechanical arguments is found in Ref. [13].

$$\chi_{dia} < 0 \quad \text{and} \quad \chi_{dia} = const. \quad (2.1)$$

- **Paramagnetism:** The opposite behavior is observed for paramagnetic materials. In this case,

permanent magnetic dipoles are prevalent and aligned by the magnetic field. This effect can be impeded by the temperature.

$$\chi_{para} > 0 \quad \text{and} \quad \chi_{para} = \chi_{para}(T) \quad (2.2)$$

The permanent magnetic dipoles can stem from unpaired electrons which means that the magnetic dipoles are localized. This is known as the temperature dependent Langevin paramagnetism. The other case, where the magnetic dipoles are due to quasi-free conducting electrons carrying a magnetic moment of μ_B , is called Pauli paramagnetism and is in first order approximation not dependent on temperature due to the Pauli principle.

$$\chi_{Pauli} \ll \chi_{Langevin} \quad (2.3)$$

- **Collective Magnetism:** The magnetic susceptibility is in this case a complicated function of the temperature T , the magnetic field \mathcal{H} , and the material and how it was treated before.

$$\chi_C = \chi_C(T, \mathcal{H}, treatment) \quad (2.4)$$

As in the paramagnetic case, magnetic dipoles need to be prevalent, however, in the case of collective magnetism, those dipoles interact with each other through the so-called exchange interaction, that leads to a magnetization underneath the critical temperature T^* . Since this is not caused by an external magnetic field, it is called a spontaneous magnetization. The exchange interaction can only be explained by quantum mechanics. As for paramagnetism, those magnetic dipoles can be local or itinerant. For temperatures $T > T^*$, the materials show paramagnetic behavior. Collective magnetism can be further divided into 3 subclasses:

- In ferromagnets, permanent magnetic dipoles align for $0 < T < T^* = T_C$ into the same direction and are fully aligned for $T = 0$. The critical temperature is in this case known by the name of Curie temperature T_C .
- The lattice of ferrimagnets is made of two ferromagnetic lattices A and B with different magnetizations $\mathbf{M}_A \neq \mathbf{M}_B$ that do not add to zero magnetization.

$$\mathbf{M} = \mathbf{M}_A + \mathbf{M}_B \neq 0 \quad \text{for} \quad T < T_C \quad (2.5)$$

- Antiferromagnets can be interpreted as a special case of ferrimagnets, where the critical temperature is here also known as the Néel temperature T_N .

$$|\mathbf{M}_A| = |\mathbf{M}_B| \neq 0 \quad \text{and} \quad \mathbf{M}_A = -\mathbf{M}_B \quad \text{for} \quad T < T_N \quad (2.6)$$

This leads to an overall magnetization of $\mathbf{M} = 0$.

- Recently, a fourth subclass of magnetic materials, known as altermagnets, was discovered. These exhibit characteristic features of both ferromagnets and antiferromagnets [15]. In this case, the opposite-spin sublattices are rotated and not connected by inversion.

The collective magnetism is the most complex category and therefore also the one that is of interest

in this work. It does not result from a direct dipole-dipole interaction between atomic magnetic moments but is a direct consequence of the Coulomb interaction between electrons and the Pauli principle. In general, this effect prevents electrons, whose spins are aligned, to get too close [13]. This can lead to a favoured magnetic arrangement of the prevalent magnetic dipoles of the presented kind and is called a spontaneous magnetization. A simple way to get a grasp of it is to look at the treatment of the H_2 -molecule by the Heitler-London theory [13]. This way, the original Hamiltonian, that contains the kinetic part and the Coulomb interaction, can be replaced by an effective Hamiltonian that only contains spin-spin interactions.

2.2 Heisenberg Model

This effective Hamiltonian can be generalized to a quantum many-body model, known as the quantum Heisenberg model [16]

$$H = J \sum_{\langle ij \rangle}^N \mathbf{S}_i \mathbf{S}_j = J \sum_{\langle ij \rangle}^N \left[\frac{1}{2} (S_i^+ S_j^- + S_i^- S_j^+) + S_i^z S_j^z \right], \quad (2.7)$$

where the sum goes over all nearest neighbors $\langle ij \rangle$, N denotes the number of sites in the model and d the local dimension of the spin operator. In the following, the local dimension of the spin operators is set to $d = 2$. These operators are denoted as $\mathbf{S} = (S^x, S^y, S^z)$ and related to the Pauli matrices $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)$

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (2.8)$$

by $\mathbf{S} = \frac{\hbar}{2} \boldsymbol{\sigma}$ [17]. For ease of notation, \hbar is set to 1. For the one-dimensional spin chain and open boundary conditions, the sum is easily expressed as

$$H = J \sum_{i=1}^{N-1} \left[\frac{1}{2} (S_i^+ S_{i+1}^- + S_i^- S_{i+1}^+) + S_i^z S_{i+1}^z \right]. \quad (2.9)$$

Certainly, the spins also respond to an external magnetic field \mathcal{H} . The resulting additional factor $\mathcal{H} \sum_i \mathbf{S}_i$ ($\mu_B = 1$) is added if the behaviour in a magnetic field is of interest. In the case of the H_2 -molecule, the parameter J can be calculated exactly by the Heitler-London theory [13]. It is dependent on the Coulomb-, the exchange- and the overlap-integral. In the general Heisenberg model, ferromagnetic materials are characterized by negative coupling constants, denoted as $J < 0$, while antiferromagnets/ferrimagnets are indicated by $J > 0$. The Heisenberg model mainly describes insulators, in which the localized magnetic moments are caused by incompletely filled electron shells. Thereby, it is essential that the shells of neighbouring atoms overlap [13]. However, in metals, there is also an indirect exchange interaction where the electric current and magnetism are governed by different groups of electrons. This interaction effect is referred to as the RKKY interaction honouring the works of M.A. Ruderman and C. Kittel [18], T. Kasuya [19] and K. Yosida

[20]. The one-dimensional Heisenberg model is realized along a specific axis in some special chemical compounds like Sr_2CuO_3 [21], copper benzoate $Cu(C_6H_5COO)_2 \cdot 3H_2O$ [1] or copper pyrazine dinitrate $CuPzN$, i.e. $Cu(C_4-H_4N_2)(NO_3)_2$ [22, 23]. To grasp how this is implemented in actual materials, the crystal structure of the copper benzoate is depicted in Fig. 2.1.

From the theoretical perspective, Heisenberg model is of additional importance since it is the first interacting model that was exactly solved in one dimension [24]. This has been achieved by H. Bethe using his famous (coordinate) Bethe ansatz [25] and led to the evolution of the term *quantum integrability*, which is an extension of the classical integrability definition to quantum physical models [26]. In a nutshell, it refers to the scattering matrix to be reducible to a product of two-body problems. This condition was formalized to the Yang-Baxter equation dating back to the works by C.N. Yang [27] and R.J. Baxter [28]. However, there are a number of slightly different formulations used and thus there is a certain disunity about the exact definition. These are compared and discussed in Ref. [29]. Later, the Heisenberg model was solved additionally to the coordinate Bethe ansatz by the famous algebraic Bethe ansatz in Refs. [26, 30, 31] and subsequently generalized to open boundary conditions in Ref. [32]. Moreover, there were non-linear integral equations found to allow for the efficient computation of thermodynamic quantities in the thermodynamic limit [33]. All in all, the Heisenberg model is in addition to its practical relevance, due to the results obtained via numerous methods and the availability of exact results, a meaningful toy model to probe quantum mechanical and quantum statistical effects. A more profound comprehension of spin chain models in a general context paves the way for numerous applications in quantum technology [34], especially in the field of quantum information and quantum computation [35, 36, 37].

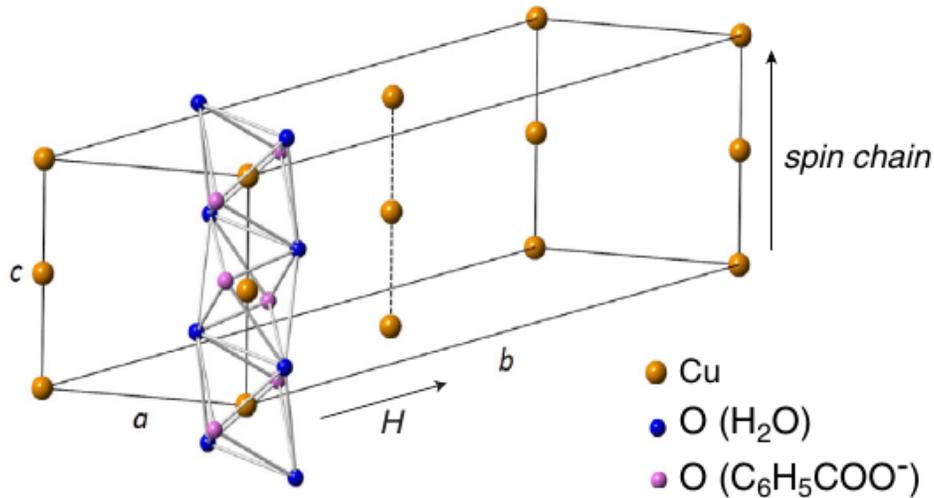


Figure 2.1: Crystal structure of copper benzoate [1]. In this arrangement, the Cu^{2+} ions possess a spin $1/2$ and organize into chains aligned along the c -axis. Each spin site occupies the center of a distorted octahedron formed by oxygen ions. Notably, the intrachain coupling strength, denoted as J in Eq. (2.7), is approximately 18 K. Additionally, there is an exceedingly weak interchain coupling, with a transition temperature (TN) of around 0.8 mK. These characteristics designate copper benzoate as a prototype for the spin- $1/2$ antiferromagnetic Heisenberg chain. Reprinted with permission from Ref. [1].

2.3 J_1 - J_2 Model

The J_1 - J_2 model adds a next-nearest neighbour interaction denoted by $\langle\langle ij \rangle\rangle$ and governed by the parameter J_2 to the nearest neighbour interaction of the Heisenberg model governed by J_1 [38]

$$H = J_1 \sum_{\langle ij \rangle} \mathbf{S}_i \mathbf{S}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \mathbf{S}_i \mathbf{S}_j . \quad (2.10)$$

For $J_1, J_2 > 0$, it is categorized within a class known as frustrated models [39]. These exhibit a more complex behavior, that is illustrated with the help of the simplest model of this kind, the Ising model

$$H = J \sum_{\langle ij \rangle} s_i s_{i+1} \quad (2.11)$$

on a triangular lattice, where s_i is a classical spin meaning $s_i \in \{-1, +1\}$. The unit cell of the triangular lattice is depicted in Fig. 2.2. To minimize the energy, two neighbouring spins are placed in opposite direction. The third spin, however, can not be orientated opposite to the two present ones leaving it frustrated [40]. Thus, the two states with the third spin up or down both have the same energy, which results in the ground state being degenerate. Since the frustration is caused by the geometry of the lattice, it is also called geometrical frustration [41]. To summarize, the interesting feature of frustrated models is that a unique configuration at low temperatures is impeded by the geometry. This leads to exciting effects like spin liquids, in which the system is in a fluid-like, highly-correlated state with no magnetic order at all [40]. For the two-dimensional J_1 - J_2 model, the consequences of frustration peak at the point $J_1 = 2J_2$ and $J_1, J_2 > 0$ [38, 42]. It was first discovered to be realized in $Li_2VO SiO_4$, which forms a two-dimensional square lattice with $\frac{J_2}{J_1} = 1.1 \pm 0.1$ [43].

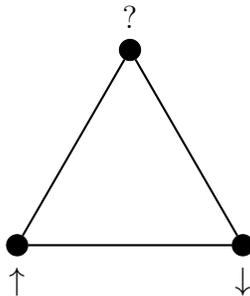


Figure 2.2: Unit cell of the Ising model on a triangular lattice depicting the frustration with respect to the upper spin.

2.4 Hubbard Model

The Hubbard model is closely related to the Heisenberg model and the latter is obtained from the former as will be discussed below. Moreover, the Hubbard model essentially emerges as the primary

model when examining electrons in a quantum many-body system with unrestricted motion, followed by the inclusion of Coulomb repulsion [12]. Hence, the Hamiltonian

$$H = -t \sum_{\langle ij \rangle, \sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_{i=1}^N \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \quad (2.12)$$

of the Hubbard model is composed of the kinetic term with the hopping parameter $t \in \mathbb{R}$ and the Coulomb interaction with the parameter $U \in \mathbb{R}$ [12]. The operator $c_{i\sigma}^\dagger$ ($c_{i\sigma}$) creates (annihilates) an electron with spin $\sigma = \{\uparrow, \downarrow\}$ at position i and the occupation number operator $\hat{n}_{i\sigma}$ yields the occupation number $n_{i\sigma} \in \{0, 1\}$. In contrast to the Heisenberg model, where the spins are localized, the electrons in the Hubbard model can hop from one site to the other. Thus, the Hubbard model can additionally describe a certain band of a conducting material [17]. Hence, it characterizes band magnets, in which the magnetic effects and the conduction are mediated by the same group of electrons in one dimension [13]. It is the simplest model to describe these metals and is exactly solvable in one dimension by the nested Bethe ansatz. This is thoroughly examined in Ref. [44]. The Heisenberg model is obtained from the Hubbard model in the low energy limit for $U \rightarrow \infty$ at half-filling $N_e = N$. This is shown in Ref. [12], as well as the derivation of those two models from second quantization and the tight-binding approach. The Heisenberg model is therein also obtained up to an additive constant to describe the low energy spectrum in the strong coupling limit $\frac{U}{t} \gg 1$.

2.5 Marshall-Peierls Sign Rule and *Stoquastic* Hamiltonians

The Marshall-Peierls sign rule for the Heisenberg Hamiltonian with periodic boundary conditions was found in a collaborative work by R.E. Peierls and W. Marshall [45] and was later extended to be applicable to a specific class of bipartite lattices in Ref. [46]. It states, that the state with the lowest eigenvalue in every sector $s^z = \sum_n s_n^z$ of a bipartite lattice, like the unfrustrated Heisenberg model in Eq. (2.7), is given by

$$|\Psi_{s^z}\rangle = \sum_n c_n(s^z) |n\rangle \quad (2.13)$$

with $c_n(s^z) > 0$, where

$$|n\rangle = (-1)^{s_A - s_A^z} |s_1^z\rangle \otimes |s_2^z\rangle \otimes \dots \otimes |s_N^z\rangle \quad (2.14)$$

the s^z -basis states and $|s_i^z\rangle$ the eigenstate of the S_i^z operator on site i , $s_A = \sum_{n \in A} |s_n^z|$ and $s_A^z = \sum_{n \in A} s_n^z$ [47]. A bipartite lattice is one which can be divided into two sublattices A and B and in which the spins on one sublattice only interact with spins on the other sublattice. This leads to the result that especially the ground state wave function is non-degenerate and the wave function components are real. However, the sign structure is rather complicated. This issue can be cured by a canonical transformation

$$\begin{aligned}
S_{i(A)}^x &\rightarrow -S_{i(A)}^x \\
S_{i(A)}^y &\rightarrow -S_{i(A)}^y \\
S_{i(A)}^z &\rightarrow S_{i(A)}^z
\end{aligned}
\tag{2.15}$$

on the x - and y - components of the spin operators on one of the two sublattices [48, 49]. This corresponds to a rotation around the z -axis by π and changes the Heisenberg Hamiltonian in one dimension to be

$$H = J \sum_{i=1}^{N-1} \left[-\frac{1}{2} (S_i^+ S_{i+1}^- + S_i^- S_{i+1}^+) + S_i^z S_{i+1}^z \right].
\tag{2.16}$$

The crucial point is that in this case the wave function components of the ground state in the s^z -basis are all real and positive implying the most simple sign structure possible [50]. This results in a great simplification of the ground state optimization in the variational Monte Carlo (VMC) method discussed in Ch. 4. The Hamiltonian is this way transformed into a so-called *stoquastic* one (a composition of the word stochastic and quantum), meaning that the off-diagonal elements in the standard product basis (s^z -basis for spin models) are real and non-positive. Hence, this Hamiltonian is also amenable to the quantum Monte Carlo (QMC) method since the matrix elements for $e^{-\beta H}$ are always positive and the sign problem is avoided for any inverse temperature $\beta \geq 0$ [51].

2.6 Phase Transitions and Critical Exponents

Phase transitions can be classified by the so-called Ehrenfest classification by observing the n -th derivatives of the free enthalpy

$$G(T, P, N) = E + PV - TS
\tag{2.17}$$

with respect to its natural variables (T and P for fluid systems and T and $\mathbf{B} = \mu_0 \mathcal{H}$ for magnets), where T the temperature, P the pressure, N the number of particles, E the energy, V the volume and S the entropy [52]. They are noted to be of n -th order if at least one of its n -th derivatives is not continuous while all lower order derivatives are. While only phase transitions of first and second order are of practical interest, the strict classification by Ehrenfest does not uphold for second order phase transitions due to real divergences in the second derivatives. This is why first order phase transitions are nowadays also called discontinuous, while second order phase transitions are rather called continuous referring to the continuity of the first derivative. Continuous phase transitions are also known as critical phenomena. They are observed by so-called response functions which are the compressibility, the specific heat capacity and the susceptibility, where the latter ones, namely

$$C_V = T \partial_T S = \frac{\langle \Delta H \rangle}{T^2} \quad \text{and} \quad \chi_T = \frac{1}{V} \left(\frac{\partial m}{\partial \mathcal{H}} \right)_T = \frac{1}{V} \left(\frac{\partial^2 G}{\partial \mathcal{H}^2} \right)_T
\tag{2.18}$$

are of special relevance in this work, where $\langle \Delta H \rangle^2 = \langle H^2 \rangle - \langle H \rangle^2$. The magnetic field strength is therein denoted by \mathcal{H} , whereas \mathbf{B} denotes the magnetic flux density. For clarity, the principal features of critical exponents are discussed here. For a more detailed derivation of the results, Ref. [52] is recommended. For second order phase transitions, the behaviour close to the critical point can be described by

$$f(\epsilon) = a\epsilon^\phi(1 + b\epsilon^{k_1} + \dots) \quad \text{with} \quad k_1 > 0, \quad (2.19)$$

where ϵ is the control parameter and ϕ the critical exponent [52]. The control parameter can be for example the temperature, with T_C being the critical temperature

$$\epsilon = \frac{T - T_C}{T_C}, \quad (2.20)$$

but also other macroscopic variables like the pressure or the magnetic field. Close to the critical value $\epsilon \rightarrow 0$, the only surviving term is ϵ^ϕ with ϕ being defined as

$$\phi = \lim_{\epsilon > 0, \epsilon \rightarrow 0} \frac{\ln |f(\epsilon)|}{\ln(\epsilon)} \quad \text{and} \quad \phi' = \lim_{\epsilon < 0, \epsilon \rightarrow 0} \frac{\ln |f(\epsilon)|}{\ln(-\epsilon)}. \quad (2.21)$$

By using ϕ and ϕ' , it is distinguished from which direction the critical point is approached. The common critical exponents (for magnets) are listed for clarity:

- α, α' : Specific heat capacity:

$$C \sim \begin{cases} \epsilon^{-\alpha} & \text{if } \epsilon > 0 \\ (-\epsilon)^{-\alpha'} & \text{if } \epsilon < 0 \end{cases}$$

- β : Magnetization: $M \sim (-\epsilon)^\beta$

- γ, γ' : Susceptibility:

$$\chi \sim \begin{cases} \epsilon^{-\gamma} & \text{if } \epsilon > 0 \\ (-\epsilon)^{-\gamma'} & \text{if } \epsilon < 0 \end{cases}$$

- ν, ν' : Correlation length:

$$\xi \sim \begin{cases} \epsilon^{-\nu} & \text{if } \epsilon > 0 \\ (-\epsilon)^{-\nu'} & \text{if } \epsilon < 0 \end{cases}$$

- η, η' : Correlation function: $O_{ij}^{\alpha\beta} = \langle S_i^\alpha S_j^\beta \rangle - \langle S_i^\alpha \rangle \langle S_j^\beta \rangle \sim \frac{1}{|i-j|^{d-2+\eta}}$ (d: dimension of the lattice)

These are so important due to the feature of universality of physical systems close to critical points. They are only dependent on three outer parameters of the system [52]:

1. Dimension d of the system
2. Range of the interaction

3. Number of relevant spin components

The critical exponents are found for a variety of models in Ref. [52]. The values for the critical exponents of the Ising model in two dimensions without external field $\mathcal{H} = 0$ are cited here:

$$\alpha, \alpha' = 0(\log), \quad \beta = 0.125, \quad \gamma, \gamma' = 1.75, \quad \nu, \nu' = 1 \quad (2.22)$$

Any phase transition with these critical exponents is referred to as a two-dimensional Ising phase transition.

The renormalization group (RG) theory allows a deeper understanding of critical phenomena [54]. It manages to explain why diverse systems exhibit similar behavior at critical points, resulting in the agreement of their critical exponents. This universality property allows to classify models into universality classes. The critical behavior of a system can be attributed to only the dimension of the system, the range of interactions and the symmetry of its order parameter [53]. For the two-dimensional Ising model, the order parameter is given by the magnetization and the Hamiltonian is invariant under \mathbb{Z}_2 transformations corresponding to a flip of all spins [12, 55]. The tendency of the system to favour rather spin-up states over spin-down states (or vice versa) for temperatures lower than the critical temperature is known as a *spontaneous symmetry breaking* [56]. The realized equilibrium state violates the \mathbb{Z}_2 symmetry. A similar situation of competing ground states, which differ by a flip of all spins, is prevalent for the two-dimensional J_1 - J_2 model in the stripe phase depicted in the phase diagram of the ground state in Fig. 2.3. The order parameter is in this case not the overall magnetization but the one measuring the difference between the magnetization in vertical and horizontal direction (see Eq. (10.5)). It is stressed that the results in between the Néel and the stripe phase remain a matter of debate. The finite-temperature phase transition in the two-dimensional J_1 - J_2 model was first predicted in Ref. [57] and later numerically investigated in Ref. [6]. This is in more detail explained in Ch. 10, where the finite-temperature phase transition is studied using a variational ansatz comprised by a NN.

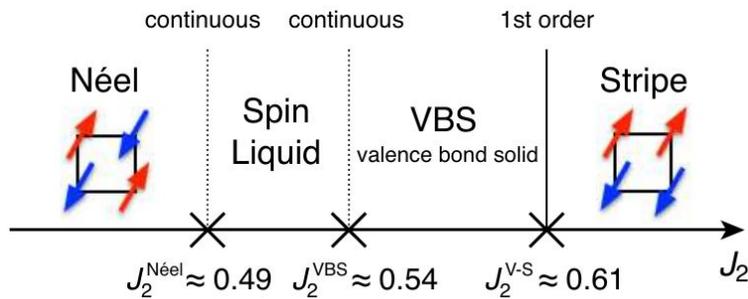


Figure 2.3: Phase diagram of the ground state of the J_1 - J_2 Heisenberg model. Reprint available under Creative Commons Attribution 4.0 International license from Ref. [5].

Chapter 3

Quantum Geometry

Quantum Geometry serves as a fundamental component for comprehending the computational processes undertaken in this study. It is essentially a reformulation of standard quantum mechanics from a geometrical perspective and is extensively covered in Refs. [58, 59, 60]. Only those parts relevant to the topic of this work are discussed in this section. There are many analogies to the description of classical mechanics in which the phase space consists of all possible states and observables are characterized by the space of real-valued functions on this symplectic manifold. Pure quantum physical states on the other hand are defined as vectors $|\Psi\rangle$ in a high-dimensional, complex Hilbert space \mathcal{H} of dimension \mathbb{C}^K with the Bloch sphere in Fig. 3.1 being the illustration of the simplest quantum physical object, the qubit or single spin-1/2-particle in the two-dimensional Hilbert space. It will serve in the following as the guiding example to depict the advanced geometrical considerations.

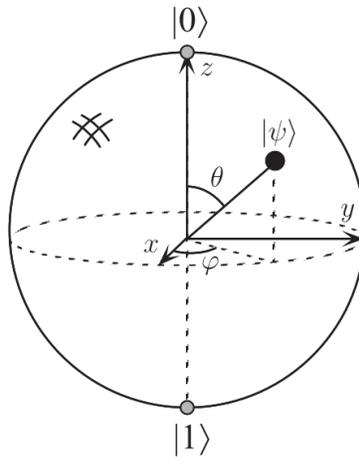


Figure 3.1: The Bloch sphere with a specific point identified by Eq. (3.13). Reprinted with permission from Ref. [2].

3.1 Projective Hilbert Space and Tangent Space

For finite dimensions, the vector space of quantum physical objects is given by \mathbb{C}^K . The dimension K will be specified later. Any vector is described in a certain basis $|k\rangle$ with $k = 1, \dots, K$ by

$$|\Psi\rangle = \sum_{k=1}^K c_k |k\rangle . \quad (3.1)$$

An initial characteristic is that rescaling a quantum state $|\Psi\rangle$ does not have any measurable effect, since physical quantities are computed by the expectation value

$$A_\Psi = \frac{\langle \Psi | A | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (3.2)$$

for arbitrary hermitian operators A . Thus, the projective Hilbert space $\mathcal{P}(\mathcal{H})$ is defined by the following equivalence relation [59]:

$$|\Psi\rangle \sim |\tilde{\Psi}\rangle \Leftrightarrow \exists \alpha \in \mathbb{C} : |\Psi\rangle = \alpha |\tilde{\Psi}\rangle \quad (3.3)$$

It is a complex projective Hilbert space that is in more mathematical terms referred to as $\mathcal{CP}^{K-1}(\mathcal{H})$ and studied in detail in Ref. [58]. The reduction in dimensionality comes from the equivalence relation in Eq. (3.3) and will become clear later on by considering the example of the Bloch sphere. However, in the following, the projective Hilbert space will consistently be denoted by $\mathcal{P}(\mathcal{H})$. It is made up of all the vectors that are related by the multiplication with a non-zero complex number. These are also known as *rays* and can be interpreted as the quantum phase space in equivalence to the classical phase space [60]. The projective Hilbert space is a real Kähler manifold, meaning it is at the same time a Riemannian, a symplectic and a complex manifold [59]. The Riemannian property allows to define a metric tensor, while a real and complex parametrization is allowed. The representative vector for the equivalence class is usually chosen to be the normalized one $\langle \Psi | \Psi \rangle = 1$. It is clear that the tangent space is reasonably defined only at an actual state $|\Psi\rangle$, which is why it is included in the notation for the tangent space states $|\delta\Psi\rangle$. To disregard changes in the direction of $|\Psi\rangle$, the following equivalence relation is used to define the tangent space [59]:

$$|\delta\Psi\rangle \sim |\widetilde{\delta\Psi}\rangle \Leftrightarrow \exists \alpha \in \mathbb{C} : |\delta\Psi\rangle - |\widetilde{\delta\Psi}\rangle = \alpha |\Psi\rangle \quad (3.4)$$

It is noted that the tangent space contains, in contrast to the projective Hilbert space, vectors of any length. This is why the tangent space is a regular vector space. As unique representatives, states with $\langle \delta\Psi | \Psi \rangle = 0$ are chosen. These can be defined for arbitrary $\mathcal{P}(\mathcal{H})$ as

$$\mathcal{H}_\Psi^\perp = \{ |\delta\Psi\rangle \in \mathcal{H} \mid \langle \delta\Psi | \Psi \rangle = 0 \} . \quad (3.5)$$

They are acquired readily if the projector

$$Q_\Psi = \left(\mathbb{1} - \frac{|\Psi\rangle\langle\Psi|}{\langle\Psi|\Psi\rangle} \right) \quad (3.6)$$

is applied to an arbitrary element of the tangent space [59]. The distance between two states or vectors $|\Psi\rangle$ and $|\Phi\rangle$ on the projective Hilbert space $\mathcal{P}(\mathcal{H})$ is given by the Fubini-Study distance

$$D_{FS}(|\Psi\rangle, |\Phi\rangle) = \arccos \sqrt{\frac{|\langle\Psi|\Phi\rangle|^2}{\langle\Phi|\Phi\rangle\langle\Psi|\Psi\rangle}} \quad (3.7)$$

and is used in Refs.[58, 61] to define infinitesimal distances $D_{FS}(|\Psi\rangle, |\Psi\rangle + \delta|\Psi\rangle)$ and thus, effectively, the Fubini-Study metric

$$\mathcal{F}_{\mu\nu} = \frac{\langle\delta\Psi|\delta\Psi\rangle\langle\Psi|\Psi\rangle - \langle\delta\Psi|\Psi\rangle\langle\Psi|\delta\Psi\rangle}{(\langle\Psi|\Psi\rangle)^2}. \quad (3.8)$$

The name was coined to honour the first contributions to this topic made by G. Fubini [62] and E. Study [63]. The argument of the square root in Eq. (3.7) is also known as the quantum fidelity

$$f[|\Psi\rangle, |\Phi\rangle] = \frac{|\langle\Psi|\Phi\rangle|^2}{\langle\Phi|\Phi\rangle\langle\Psi|\Psi\rangle}. \quad (3.9)$$

The Fubini-Study metric is closely related to the quantum Fisher information matrix (QFIM) for pure states and its generalization to mixed states, the Bures metric. This is extensively covered in Ref. [64, 65]. Hence, the quantity in Eq. (3.8) is referred to interchangeably as the Fubini-Study metric or the QFIM [66].

3.2 Variational Manifold

A variational submanifold of the Hilbert space \mathcal{H} or the projective Hilbert space $\mathcal{P}(\mathcal{H})$ is distinguished in the following by denoting the former by $\mathcal{V}(\mathcal{H})$ whereas the latter is denoted by $\mathcal{V}(\mathcal{P}(\mathcal{H}))$. Although it seems irrelevant from the discussion around Eq. (3.2) to address the variational submanifold of the Hilbert space as well, in the finite-temperature computations, it will be worthwhile due to the fact that the norm of the wave function holds relevant physical information in that specific treatment. However, for now, we limit ourselves to the discussion of the variational submanifold of the projective Hilbert space.

3.2.1 Variational Manifold of the Projective Hilbert Space

The tiny difference to the submanifold of the Hilbert space is investigated in the imaginary-time evolution in Sec. 4.2.1. It is also specified if the manifold is defined with respect to the field of real parameters $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{R})$ or complex parameters $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$. For large dimensions K , it is aimed to

approximate the general vectors of $\mathcal{P}(\mathcal{H})$ in Eq. (3.1) by a variational state

$$|\Psi_{\mathbb{P}}\rangle = \sum_{k=1}^K \Psi_{\mathbb{P}}(k)|k\rangle \quad \text{or} \quad |\Psi_{\mathbb{W}}\rangle = \sum_{k=1}^K \Psi_{\mathbb{W}}(k)|k\rangle, \quad (3.10)$$

that is parameterized by an ansatz $\Psi_{\mathbb{P}}(k)$ or $\Psi_{\mathbb{W}}(k)$ using a set of real parameters \mathbb{P} or complex parameters \mathbb{W} of size P . The input can be comprised by a single input value or a vector of input values (see Sec. 4.1). Furthermore, the expectation value with respect to an operator in Eq. (3.2) can thus also be viewed as a functional

$$\langle A \rangle_{\Psi_{\mathbb{P}}} = \frac{\langle \Psi_{\mathbb{P}} | A | \Psi_{\mathbb{P}} \rangle}{\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle} \quad \text{or} \quad \langle A \rangle_{\Psi_{\mathbb{W}}} = \frac{\langle \Psi_{\mathbb{W}} | A | \Psi_{\mathbb{W}} \rangle}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle}, \quad (3.11)$$

which is parametrized by the set of parameters \mathbb{P} or \mathbb{W} . It is stressed that this notation will be used throughout this work to denote the difference between the actual expectation value with respect to a certain state $|\Psi\rangle$, as in Eq. (3.2), and the functional in Eq. (3.11). An example for such an ansatz and its parametrization is given for a two-dimensional Hilbert space by the Bloch sphere S^2 , which is in the more mathematical notation, mentioned underneath Eq. (3.3), denoted as the complex projective line $\mathbb{C}\mathcal{P}^1(\mathcal{H}, \mathbb{R})$ [58]. It is parameterized in spherical coordinates $\mathbb{P} = \{\Theta, \varphi\}$ as

$$|\Psi_{\mathbb{P}}\rangle = \cos\left(\frac{\Theta}{2}\right)|0\rangle + e^{i\varphi}\sin\left(\frac{\Theta}{2}\right)|1\rangle, \quad (3.12)$$

where Θ is the polar angle, φ the azimuthal angle and $|0\rangle, |1\rangle$ the two basis vectors or states [2]. The angles define a point on the three-dimensional unit sphere

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \sin(\Theta) \cos(\varphi) \\ \sin(\Theta) \sin(\varphi) \\ \cos(\Theta) \end{pmatrix}, \quad (3.13)$$

which is depicted in Fig. 3.1. In this case, the variational manifold is just a parametrization of the projective Hilbert space $\mathcal{P}(\mathcal{H})$.

For the upcoming discussion, it is important if the underlying field is given by the real or complex numbers. The quantum phase space is in both cases considered a differentiable manifold with respect to its underlying field [67]. Additionally, they are also Kähler manifolds, which means they have a Riemannian, symplectic and a complex structure. This is an interesting topic of Riemannian geometry that goes far beyond the topic of this work but can be extensively studied using Refs. [59, 67, 68]. For the purpose of this work, it is sufficient to know that a proper metric tensor can be defined in both cases and that they are related, as will be shown below. The following derivation of the tangent space and the metric tensor focuses on a more illustrative derivation. Before proceeding, it is imperative to provide additional clarification on the significance of both constructs $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{R})$ and $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$. While the latter seems more natural, there exist two primary rationales for why the former also holds significance. Firstly, the authors in Ref. [60] use the variational projective Hilbert space over the field of real numbers $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{R})$ to derive quantum mechanical theorems in a geometric fashion, obtain some results that are analogous to their classical counterparts (e.g. both

are Kähler manifolds) and work out the differences (e.g. probabilistic aspects related to the metric tensor). Secondly, the real-valued ansatz is of practical importance for quantum computational applications, in which only real parameters can be used as classical input to quantum operators in a hybrid algorithm [69]. Thus, it is used in the following to introduce the most important features of the projective Hilbert space using the most illustrative example, the Bloch sphere. Additionally, the differences to the projective Hilbert space over the field of complex numbers $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$ are pointed out and the respective quantities derived in that case as well. Throughout this work, the computations are performed using complex parametrizations.

For the variational manifolds, there is also a tangent space $\mathcal{T}_\Psi \mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{R})$ and $\mathcal{T}_\Psi \mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$ defined. It describes the space of changes at the actual state $|\Psi\rangle$. For the space of unique representatives of the tangent space, a local basis is defined with $\partial_\mu = \frac{\partial}{\partial \mu}$ and $\mu \in \{1, \dots, P\}$ by

$$|V_{\mu, \mathbb{P}}\rangle = \mathbb{Q}_{\Psi_{\mathbb{P}}} |\partial_\mu \Psi_{\mathbb{P}}\rangle \quad \text{and} \quad |V_{\mu, \mathbb{W}}\rangle = \mathbb{Q}_{\Psi_{\mathbb{W}}} |\partial_\mu \Psi_{\mathbb{W}}\rangle, \quad (3.14)$$

where the derivatives are defined as

$$|\partial_\mu \Psi_{\mathbb{P}}\rangle := \sum_{k=1}^K \frac{\partial \Psi_{\mathbb{P}}(k)}{\partial \mathbb{P}_\mu} |k\rangle \quad \text{and} \quad |\partial_\mu \Psi_{\mathbb{W}}\rangle := \sum_{k=1}^K \frac{\partial \Psi_{\mathbb{W}}(k)}{\partial \mathbb{W}_\mu} |k\rangle, \quad (3.15)$$

making use of the projector defined in Eq. (3.6). The local basis induces a metric tensor that is, as usually, deduced from the inner product[58]. It simplifies in the case of normalized states to the quantum geometric tensor (QGT) for complex parameters

$$G_{\mu\nu}(\mathbb{W}) = \frac{\langle V_{\mu, \mathbb{W}} | V_{\nu, \mathbb{W}} \rangle}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle} = \frac{\langle \partial_\mu \Psi_{\mathbb{W}} | \partial_\nu \Psi_{\mathbb{W}} \rangle}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle} - \frac{\langle \partial_\mu \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle \langle \Psi_{\mathbb{W}} | \partial_\nu \Psi_{\mathbb{W}} \rangle}{(\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle)^2}, \quad (3.16)$$

and the QGT for real parameters

$$g_{\mu\nu}(\mathbb{P}) = \frac{\text{Re}[\langle V_{\mu, \mathbb{P}} | V_{\nu, \mathbb{P}} \rangle]}{\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle} = \frac{\text{Re}[\langle \partial_\mu \Psi_{\mathbb{P}} | \partial_\nu \Psi_{\mathbb{P}} \rangle]}{\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle} - \frac{\langle \partial_\mu \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle \langle \Psi_{\mathbb{P}} | \partial_\nu \Psi_{\mathbb{P}} \rangle}{(\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle)^2}, \quad (3.17)$$

since $\mathbb{Q}_\Psi^2 = \mathbb{Q}_\Psi$ [59, 70]. This is basically the variational formulation of the Fubini-Study metric in Eq. (3.8). For the real-valued parametrization, the metric in Eq. (3.17) has to be restricted to be real-valued. The second term itself is already inherently real and has a special meaning that is discussed below. However, the QGT can be computed in both cases and is investigated for the specific parametrization of the Bloch sphere in the following. First, it is divided into its real and imaginary part

$$\frac{\langle V_{\mu, \mathbb{P}} | V_{\nu, \mathbb{P}} \rangle}{\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle} = g_{\mu\nu} + iw_{\mu\nu}, \quad (3.18)$$

where

$$g_{\mu\nu} = \frac{\text{Re}[\langle V_{\mu, \mathbb{P}} | V_{\nu, \mathbb{P}} \rangle]}{\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle} \quad \text{and} \quad w_{\mu\nu} = \frac{\text{Im}[\langle V_{\mu, \mathbb{P}} | V_{\nu, \mathbb{P}} \rangle]}{\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle}. \quad (3.19)$$

These quantities define the Kähler space, which means they are related in a specific way as is explained for the real-valued case in Refs. [59, 71]. The respective equations for the mentioned parametrization of the Bloch sphere in Eq. (3.12) are noted

$$|V_1\rangle = -\frac{1}{2} \sin\left(\frac{\Theta}{2}\right) |0\rangle + \frac{e^{i\varphi}}{2} \cos\left(\frac{\Theta}{2}\right) |1\rangle \quad (3.20)$$

$$|V_2\rangle = -\frac{i}{2} \sin\left(\frac{\Theta}{2}\right) \sin(\Theta) |0\rangle + \frac{ie^{i\varphi}}{2} \cos\left(\frac{\Theta}{2}\right) \sin(\Theta) |1\rangle \quad (3.21)$$

and the components of the metric tensor computed to be [59]

$$g_{\mu\nu} = \frac{1}{4} \begin{pmatrix} 1 & 0 \\ 0 & \sin^2(\Theta) \end{pmatrix} \quad \text{and} \quad w_{\mu\nu} = \frac{1}{4} \begin{pmatrix} 0 & -\sin(\Theta) \\ \sin(\Theta) & 0 \end{pmatrix}. \quad (3.22)$$

The real part of the inner product in equation (3.18) is identified as the standard metric tensor of the unit sphere $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ with radius $r = \frac{1}{2}$ using the Einstein sum convention. The QGTs in Eqs. (3.16) and (3.17) can be used to define the projectors

$$\mathbf{P}_{\Psi_{\mathbb{P}}} = \frac{(\mathbf{g}^{-1})^{\nu\mu}}{\langle \Psi_{\mathbb{P}} | \Psi_{\mathbb{P}} \rangle} |V_{\mathbb{P},\nu}\rangle \text{Re}\langle V_{\mathbb{P},\mu}| \quad \text{and} \quad \mathbf{P}_{\Psi_{\mathbb{W}}} = \frac{(\mathbf{G}^{-1})^{\nu\mu}}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle} |V_{\mathbb{W},\nu}\rangle \langle V_{\mathbb{W},\mu}| \quad (3.23)$$

onto the tangent spaces of the variational manifolds $\mathcal{T}_{\Psi}\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{R})$ and $\mathcal{T}_{\Psi}\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$, where the states $|V_{\mathbb{P},\mu}\rangle$ and $|V_{\mathbb{W},\mu}\rangle$ are defined in Eq. (3.14) and the Einstein sum convention applies. This is a crucial quantity for the time evolution described in Sec. 4.2. Furthermore, it is stressed that the metric tensor of $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{R})$ can be also obtained by a computation considering the distance between two states that differ by a small shift [72]. The quantities

$$\beta_{\mu}(\mathbb{P}) = i\langle \Psi_{\mathbb{P}} | \partial_{\mu, \mathbb{P}} \Psi_{\mathbb{P}} \rangle \quad \text{and} \quad \beta_{\mu}(\mathbb{W}) = i\langle \Psi_{\mathbb{W}} | \partial_{\mu, \mathbb{W}} \Psi_{\mathbb{W}} \rangle, \quad (3.24)$$

which are part of the QGTs in Eqs.(3.16) and (3.17), are known as the Berry connections, respectively. These quantities are of physical relevance. When integrated around a closed curve \mathcal{C} , the respective Berry phases

$$\gamma_{\mu} = \int_{\mathcal{C}} \beta_{\mu}(\mathbb{P}) d\mathbb{P} \quad \text{and} \quad \gamma_{\mu} = \int_{\mathcal{C}} \beta_{\mu}(\mathbb{W}) d\mathbb{W}. \quad (3.25)$$

are obtained [73]. They are gauge invariant and thus define an observable that can be measured in experiments. Its most prominent manifestation is the Aharonov-Bohm phase [74]. The Berry phases are manifestly real, which is already true for the Berry connections in Eq. (3.24). This is proven by taking the derivative of the norm of a state $|\Psi\rangle$

$$\partial_{\mu} \langle \Psi | \Psi \rangle = \langle \Psi | \partial_{\mu} \Psi \rangle + \langle \partial_{\mu} \Psi | \Psi \rangle = 0. \quad (3.26)$$

Thus, $\langle \Psi_{\mathbb{P}} | \partial_{\mu, \mathbb{P}} \Psi_{\mathbb{P}} \rangle$ and $\langle \Psi_{\mathbb{W}} | \partial_{\mu, \mathbb{W}} \Psi_{\mathbb{W}} \rangle$ are purely imaginary rendering the Berry connection purely real. The link to the classical Fisher information, which is the metric tensor for classical probabilities,

is attained by parametrizing the wave function components by

$$|\Psi_{\mathbb{P}}\rangle = \sum_{k=1}^K \sqrt{p_{\mathbb{P}}(k)} |k\rangle, \quad (3.27)$$

where $\sum_{k=1}^K p_{\mathbb{P}}(k) = 1$ is a classical probability distribution [70]. Completing the discussion in Sec. 3.1 about the mentioned metric tensors and their respective names, the classical Fisher information matrix F can be related to the QGT by using the parametrization in Eq. (3.27) (see Ref. [70]).

$$g_{\mu\nu}(\mathbb{P}) = \frac{1}{4} F_{\mu\nu}(\mathbb{P}) = \frac{1}{4} \sum_{k=1}^K p_{\mathbb{P}}(k) (\partial_{\mu} \log \Psi_{\mathbb{P}}(k)) (\partial_{\nu} \log \Psi_{\mathbb{P}}(k)) \quad (3.28)$$

It is stressed that the mixed QGT for real and complex parameters has to be computed with caution and is derived step by step in Ref. [75].

3.2.2 Variational Manifold of the Hilbert Space

The variational manifold of the Hilbert space $\mathcal{V}(\mathcal{H}, \mathbb{C})$ only differs from the one of the projective Hilbert space $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$ such that the equivalence relation in Eq. (3.3) does not hold. The states of the variational manifold of the Hilbert space $\mathcal{V}(\mathcal{H}, \mathbb{C})$ are thus not normalized and allowed to change their length. The local basis of the tangent space thus changes from Eq. (3.14) to

$$|\hat{V}_{\mu, \mathbb{W}}\rangle = |\partial_{\mu} \Psi_{\mathbb{W}}\rangle, \quad (3.29)$$

where the derivatives are defined as in Eq. (3.15). The local basis induces a metric tensor that is again deduced from the inner product [58]. The quantum geometric tensor (QGT) for complex parameters

$$\hat{G}_{\mu\nu}(\mathbb{W}) = \frac{\langle \hat{V}_{\mu, \mathbb{W}} | \hat{V}_{\nu, \mathbb{W}} \rangle}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle} = \frac{\langle \partial_{\mu} \Psi_{\mathbb{W}} | \partial_{\nu} \Psi_{\mathbb{W}} \rangle}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle} \quad (3.30)$$

differs from the one for the variational manifold of the projective Hilbert space $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$ in Eq. (3.16) only by the second term. The projector onto the tangent space defined in Eq. (3.23) changes according to the updated basis of the tangent space in Eq. (3.29), but keeps its form

$$\hat{P}_{\Psi_{\mathbb{W}}} = \frac{(\hat{\mathbf{G}}^{-1})^{\nu\mu}}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle} |\hat{V}_{\mathbb{W}, \nu}\rangle \langle \hat{V}_{\mathbb{W}, \mu}|. \quad (3.31)$$

3.3 Quantum Natural Gradient (QNG)

The expectation values of quantum physical operators (see Eq. (3.2)) can be interpreted as real-valued functionals $\langle A \rangle_{\Psi_{\mathbb{P}}}$ or $\langle A \rangle_{\Psi_{\mathbb{W}}}$ (see Eq. (3.11)) over the field of real or complex numbers parametrized by the set of parameters \mathbb{P} or \mathbb{W} . Thus, the gradient with respect to these parameters can be

computed in its basis representation (Eq. (3.1)) using Eq. (3.15). An important practical implication is that according to the general theory of manifolds, the steepest descent direction is modified by the metric tensor in the following way

$$\widetilde{\nabla}\langle A\rangle_{\Psi_{\mathbb{P}}} = g^{-1} \cdot (\nabla_{\mathbb{P}}\langle A\rangle_{\Psi_{\mathbb{P}}}) , \quad \widetilde{\nabla}\langle A\rangle_{\Psi_{\mathbb{W}}} = G^{-1} \cdot (\nabla_{\mathbb{W}^*}\langle A\rangle_{\Psi_{\mathbb{W}}}) , \quad (3.32)$$

where $\nabla = (\partial_1, \dots, \partial_P)$ is the gradient [76]. Thus, it is also referred to as the natural gradient in general and the quantum natural gradient in the specific case of quantum physics [70]. For complex parameters, it is stressed that the Euclidean gradient is given by the derivative with respect to the complex conjugate of the parameters [77]. The quantities resulting in the application of the gradient are also referred to as the respective forces

$$F(\mathbb{P}) = \nabla_{\mathbb{P}}\langle A\rangle_{\Psi_{\mathbb{P}}} \quad \text{and} \quad F(\mathbb{W}) = \nabla_{\mathbb{W}^*}\langle A\rangle_{\Psi_{\mathbb{W}}} . \quad (3.33)$$

3.4 Quantum Velocity

In the following, the distance between a quantum state $|\Psi\rangle$ at time t and its time evolved version at time $t + dt$ is computed to derive the quantum velocity. The evolution is described by the Schrödinger equation with $\hbar = 1$ and $d_t = \frac{d}{dt}$.

$$i d_t |\Psi(t)\rangle = H(t) |\Psi(t)\rangle \quad (3.34)$$

The Taylor expansion for small dt can thus be obtained

$$|\Psi(t + dt)\rangle = |\Psi(t)\rangle - i H(t) |\Psi(t)\rangle dt - \frac{1}{2} [i(d_t H(t)) |\Psi(t)\rangle + H^2(t) |\Psi(t)\rangle] dt^2 + \mathcal{O}(dt^3) \quad (3.35)$$

and used to compute the overlap between those mentioned states

$$|\langle \Psi(t) | \Psi(t + dt) \rangle|^2 = 1 - \Delta H_{\Psi(t)}^2 dt^2 + \mathcal{O}(dt^3) , \quad (3.36)$$

where $\Delta H_{\Psi(t)}^2 = (H^2)_{\Psi(t)} - (H_{\Psi(t)})^2$ the variance of the Hamiltonian [78]. For pure states, the Fubini-Study distance in Eq. (3.7) is used to expand the overlap for small dt

$$f[|\Psi(t)\rangle, |\Psi(t + dt)\rangle] = \cos^2(D_{FS}[|\Psi(t)\rangle, |\Psi(t + dt)\rangle]) = 1 - (d_t D_{FS}[|\Psi(t)\rangle, |\Psi(t + dt)\rangle])^2 \cdot dt^2 + \mathcal{O}(dt^3) , \quad (3.37)$$

where $f[|\Psi(t)\rangle, |\Psi(t + dt)\rangle]$ is the fidelity defined in Eq. (3.9). By comparing Eq.(3.36) and (3.37) and setting the infinitesimal distance to $s = D_{FS}[|\Psi(t)\rangle, |\Psi(t + dt)\rangle]$, the quantum or Hilbert space velocity

$$\frac{ds}{dt} = \Delta H_{\Psi(t)} \quad (3.38)$$

is determined [79, 58]. The speed of quantum evolution is thus governed by the energy fluctuation.

The relation is sometimes also called Anandan-Aharonov theorem since it was first discovered by J. Anandan and Y. Aharonov in Ref. [78]. It will be of great importance in Sec. 4.4 to quantify the error arising from the approximate time evolution on the variational manifold.

Chapter 4

Variational Monte Carlo Method

The variational Monte Carlo (VMC) method builds upon the variational principle and a Monte Carlo sampling of the basis states of the Hilbert space \mathcal{H} , that is in quantum many-body physics generally exponentially growing with system size. Additionally, the wave function components of physical states are approximated in a certain basis by a variational ansatz $\Psi_{\mathbb{P}}$ or $\Psi_{\mathbb{W}}$ that is specified by a set of real variational parameters \mathbb{P} or complex variational parameters \mathbb{W} . In this section, the focus will be on complex parameters, however, the difference to real parameters is noted whenever it is crucial. The equations for real parameters are slightly different and are given in Sec. 3. While the concrete nature of the variational ansatz is the topic of Sec. 5 and 6, the main goal of this section is to explain the principal algorithm and to derive the basic equations that are used with only small modifications throughout this work. To do so, first, the most evident application of the VMC method, the ground state optimization is discussed step by step. Thereafter, the imaginary- and real-time evolution are explained. The explanations of the adaptive Heun method in Sec. 4.3 and the β modification in Sec. 4.4 have been published in Ref. [80] (Wagner, Klümper, Sirker) already.

4.1 Ground State Optimization

The variational Monte Carlo (VMC) method to find an approximation of the ground state and thus also the ground state energy of quantum physical models is based on the variational principle, which generally states that \forall states $|\Psi\rangle \in \mathcal{H}$, the expectation value with respect to the Hamiltonian H of a quantum model is bounded from below by the ground state energy [81]

$$E_0 \leq E_{\Psi} = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \quad (4.1)$$

It is mentioned that the expectation value of the Hamiltonian H_{Ψ} is most commonly denoted by E_{Ψ} . This notation is adapted from here on. A general quantum state $|\Psi\rangle$ can be expressed in any basis of the Hilbert space \mathcal{H} . For the spin-1/2 chains of interest, that were discussed in Sec. 2, the s^z -basis $|\mathbf{s}\rangle$ with configuration $\mathbf{s} = \{s_1^z, \dots, s_N^z\}$, where $s_i^z = \pm \frac{1}{2}$ defines the direction of the s^z -component of

the spin at site i , is the most evident one and used throughout this work [82]. Any quantum state is expanded in this basis as

$$|\Psi\rangle = \sum_{\mathbf{s}} \mathbf{F}_{\mathbf{s}} |\mathbf{s}\rangle , \quad (4.2)$$

where $\mathbf{F}_{\mathbf{s}} \in \mathbb{C}^{2^N}$ are called the wave function components of the spin configuration \mathbf{s} . The wave function components are in general complex-valued. The number of configurations is given by 2^N and therefore grows exponentially in system size. Hence, it is in principle not possible to diagonalize the Hamiltonian and work with the eigenstates for system sizes larger than a few spins. This is why the many-body state is approximated by a variational ansatz

$$|\Psi_{\mathbb{W}}\rangle = \sum_{\mathbf{s}} \Psi_{\mathbb{W}}(\mathbf{s}) |\mathbf{s}\rangle \quad (4.3)$$

with $\Psi_{\mathbb{W}}(\mathbf{s})$ the variational wave function that is parameterized by a set of complex-valued parameters \mathbb{W} . The quality of the ansatz is judged by how accurate the ground state can be approximated for the appropriate choice of parameters. The number of parameters should clearly grow considerably slower with system size than the dimension of the Hilbert space does. The concrete design is usually a compromise between the computational time and the accuracy of the approximation. The exact nature of the ansatz wave functions that are used in this work are discussed later in Sec. 5 and 6. By inserting the variational wave function in Eq. (4.1), a compact expression for the energy of the state is obtained

$$\langle H \rangle_{\Psi_{\mathbb{W}}} = \sum_{\mathbf{s}} \underbrace{\frac{|\Psi_{\mathbb{W}}(\mathbf{s})|^2}{\sum_{\mathbf{s}''} |\Psi_{\mathbb{W}}(\mathbf{s}'')|^2}}_{p(\mathbf{s})} \underbrace{\sum_{\mathbf{s}'} \langle \mathbf{s} | H | \mathbf{s}' \rangle \frac{\Psi_{\mathbb{W}}(\mathbf{s}')}{\Psi_{\mathbb{W}}(\mathbf{s})}}_{E_{\text{loc}}(\mathbf{s})}, \quad (4.4)$$

where $p(\mathbf{s})$ is the probability distribution and $E_{\text{loc}}(\mathbf{s})$ the so-called local energy. Since the number of basis states of a quantum system grows exponentially in system size, it is essential to approximate the probability distribution $p(\mathbf{s})$ as well. This is achieved by the Metropolis-Hastings algorithm, in which one starts by choosing the initial configuration \mathbf{s} randomly from a uniform distribution and adding it as the initial entry to the list of sampled configurations of size N_{Met} [83]. Additionally, the initial parameters are also chosen in a specific manner, usually randomly from a normal distribution. For the probability distribution $\pi(\mathbf{s})$, given by the occurrence of the sampled configurations, to converge to the correct probability distribution $p(\mathbf{s})$, two conditions need to be fulfilled. One is the stationary requirement

$$p_{\mathbf{s}'} = \sum_{\mathbf{s}} T_{\mathbf{s} \rightarrow \mathbf{s}'} p_{\mathbf{s}} , \quad (4.5)$$

where $T_{\mathbf{s} \rightarrow \mathbf{s}'}$ the transition probability for the configuration \mathbf{s} to change to the configuration \mathbf{s}' [3]. A sufficient condition for the stationary requirement is the detailed balance condition

$$T_{\mathbf{s} \rightarrow \mathbf{s}'} \pi(\mathbf{s}) = T_{\mathbf{s}' \rightarrow \mathbf{s}} \pi(\mathbf{s}') . \quad (4.6)$$

Since it can be directly used to define various possible transitions T , the detailed balance condition is sometimes used interchangeably with the stationary requirement in Eq. (4.5), however, it is stressed that it is actually only a sufficient condition for the stationary requirement [84]. The second

condition is known by the name of ergodicity, requiring any state to be visited by a finite number of steps when starting from any other state. The configuration \mathbf{s}_l is changed in a Monte Carlo chain algorithm, known as the Metropolis-Hastings algorithm, according to a proposed transition T and accepted with an acceptance probability of

$$Acc(\mathbf{s}_l \rightarrow \mathbf{s}_{l+1}) = \min \left\{ 1, \frac{|\Psi_{\mathbb{W}}(\mathbf{s}_{l+1})|^2}{|\Psi_{\mathbb{W}}(\mathbf{s}_l)|^2} \right\}, \quad (4.7)$$

where l denotes the number of the Monte Carlo steps and \min simply takes the minimum of its two arguments [11]. This is implemented into an algorithm by choosing N_{Met} random numbers r_l from a uniform distribution in the range $(0, 1]$. For every sampling step l , the acceptance probability is compared to the respective random number r_l . If the acceptance probability is larger than r_l , the new configuration is accepted and written into the sampling list at position l , otherwise it is rejected and the former configuration is once again written into the sampling list at position l . This is repeated N_{Met} times leading to a sampling list of configurations of the same length. The probabilities of the configurations that can not be neglected is thus approximated by the number of occurrences in the sampling list

$$p(\mathbf{s}_l) \approx \frac{occ(\mathbf{s}_l)}{N_{\text{Met}}}. \quad (4.8)$$

In the case of spin-1/2-chains, the local change corresponds, for example, to a simple spin flip at a random position. In the ground state optimization, one is in certain models interested to stay in a specific s^z -sector, which can be achieved by proposing an exchange of two spins at random positions as the transition T . By this algorithm, the energy of a certain state parameterized by a set of variational parameters can be approximated by considering only the most probable configurations. The Monte Carlo step of the algorithm is illustrated in blue in the flowchart in Fig. 4.1.

4.1.1 Stochastic Gradient Descent (SGD)

In order to minimize the energy of the variational state to approximate the ground state and its energy, the simple derivative with respect to the energy functional (see Eq. (3.11)) of every single parameter \mathbb{W}_μ is taken. Combined with the stochastic approximation of the probability distribution, this is known as the stochastic gradient descent (SGD) approach [3]. The gradient with respect to the complex parameters

$$F_\mu(\mathbb{W}) = \partial_{\mathbb{W}_\mu^*} \langle H \rangle_{\Psi_{\mathbb{W}}} = \langle O_\mu^* H \rangle_{\Psi_{\mathbb{W}}} - \langle O_\mu^* \rangle_{\Psi_{\mathbb{W}}} E_{\Psi_{\mathbb{W}}}, \quad (4.9)$$

also known as the energy forces (see Eq. (3.33)), differs from the one for real parameters

$$F_\mu(\mathbb{P}) = \partial_{\mathbb{P}_\mu} \langle H \rangle_{\Psi_{\mathbb{P}}} = 2 \text{Re} \left[\langle O_\mu^* H \rangle_{\Psi_{\mathbb{P}}} - \langle O_\mu^* \rangle_{\Psi_{\mathbb{P}}} E_{\Psi_{\mathbb{P}}} \right]. \quad (4.10)$$

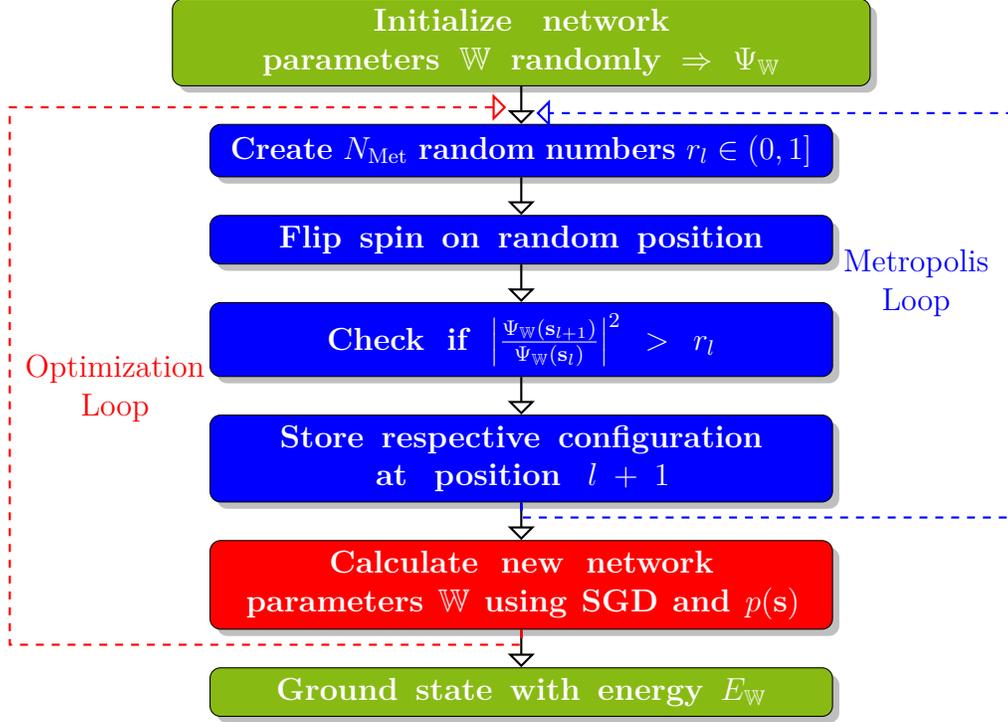


Figure 4.1: Flowchart of the VMC algorithm. The iteration comprises two nested loops, the Metropolis loop in blue and the optimization loop in red.

The respective second term originates from the derivative of the norm in Eq. (4.4). The notation $\langle \cdot \rangle$ denotes the average with respect to the probability distribution $p(\mathbf{s})$.

$$\langle A \rangle_{\Psi_{\mathbb{W}}} = \sum_{\mathbf{s}} p(\mathbf{s}) \langle \mathbf{s} | A | \mathbf{s} \rangle \quad (4.11)$$

The logarithmic derivatives are defined by

$$O_{\mu}(\mathbf{s}) = \frac{\partial_{\mathbb{W}_{\mu}} \Psi_{\mathbb{W}}(\mathbf{s})}{\Psi_{\mathbb{W}}(\mathbf{s})}. \quad (4.12)$$

The parameters are optimized step by step with the simplest approach being the Euler method

$$\mathbb{W}_{\mu}[l+1] = \mathbb{W}_{\mu}[l] - \epsilon F_{\mu}(\mathbb{W}[l]), \quad (4.13)$$

where ϵ is the step size and l denotes the l -th step. The Taylor expansion of the energy for a variation of the parameters $\delta \mathbb{W}_{\mu} = \epsilon F_{\mu}(\mathbb{W})$ is given by

$$E_{\Psi_{\mathbb{W}+\delta \mathbb{W}}} = E_{\Psi_{\mathbb{W}}} + \sum_{\mu} \partial_{\mathbb{W}_{\mu}} \langle H \rangle_{\Psi_{\mathbb{W}}} \delta \mathbb{W}_{\mu} + O(\epsilon^2) = E_{\Psi_{\mathbb{W}}} - \epsilon \sum_{\mu} F_{\mu}^2(\mathbb{W}) + O(\epsilon^2) \quad (4.14)$$

with $\partial_{\mathbb{W}_\mu^*} \langle H \rangle_{\Psi_{\mathbb{W}}} = -F_\mu(\mathbb{W})$ and thus shows that the energy is always minimized for small enough values of the parameter ϵ [3].

$$\Delta E_{\Psi_{\mathbb{W}}} = E_{\Psi_{\mathbb{W}+\delta\mathbb{W}}} - E_{\Psi_{\mathbb{W}}} = -\epsilon \sum_{\mu} F_\mu^2(\mathbb{W}) \leq 0 \quad (4.15)$$

The SGD method adds another separate loop to the flowchart, as is depicted in red in Fig. 4.1. Finally, all ingredients to understand the entire algorithm, which might be slightly changed according to the model at hand and the accuracy desired, are available. It is thus summed up for clearance: Starting with random initial parameters, the Metropolis loop computes a list of N_{Met} configurations, which allows to define the probability of each configuration. It is given by the number of times it occurs in the sampling list. This enables the computation of the energy according to Eq. (4.4) and, subsequently, to compute the gradients with respect to every parameter \mathbb{W}_μ . The parameters are altered accordingly and the next iteration step starts with the Monte Carlo sampling again. After all the iterations are completed, the final variational state approximates the actual ground state to an accuracy that is predefined by the ability of the variational ansatz to represent this state and the sampling accuracy.

If one is only interested in the ground state energy, the only demand the variational ansatz needs to fulfill is to represent the ground state accurately. Thus, if there are any features of the ground state known, for example its symmetry of some kind or the magnetization, the ansatz can be adjusted accordingly. For instance, in the case of the antiferromagnetic Heisenberg chain, it is advisable to confine the proposed configurations to the regime where $s^z = \sum_i^N s_i^z = 0$, since the ground state is, according to the Marshall-Lieb-Mattis theorem, part of this regime [85]. Furthermore, it is emphasized that the energy is only one example of an operator that can be optimized. The VMC method and the presented notation stem from quantum physics. However, the Monte Carlo algorithm is used in various fields to approximate probability distributions in a high-dimensional space.

4.1.2 Quantum Natural Gradient (QNG)

The SGD approach is improved by using the quantum natural gradient (QNG) explained in Sec. 3.3. The variational parameters not necessarily form a Euclidean but a curved space and thus the metric tensor known as the QGT is used to define the steepest descent. The QGT for complex parameters in Eq. (3.16) is expressed using the notation in Eq. (4.12) as

$$G_{\mu\nu}(\mathbb{W}) = \langle O_\mu^* O_\nu \rangle_{\Psi_{\mathbb{W}}} - \langle O_\mu^* \rangle_{\Psi_{\mathbb{W}}} \langle O_\nu \rangle_{\Psi_{\mathbb{W}}} , \quad (4.16)$$

whereas the one for real-valued parameters in Eq. (3.17) is expressed as

$$G_{\mu\nu}(\mathbb{P}) = \text{Re}[\langle O_\mu^* O_\nu \rangle_{\Psi_{\mathbb{P}}} - \langle O_\mu^* \rangle_{\Psi_{\mathbb{P}}} \langle O_\nu \rangle_{\Psi_{\mathbb{P}}}] . \quad (4.17)$$

Since all quantities are subjected to the sampling error, it is ensured that the second term on the right hand side is also real-valued. According to Eq. (3.32), to obtain the QNG

$$\tilde{\mathbf{F}}(\mathbb{W}) = -\mathbf{G}^{-1}(\mathbb{W}) \cdot \mathbf{F}(\mathbb{W}) , \quad (4.18)$$

the inverse of the QGT needs to be computed. This is, in principle, an ill-posed problem also known as the inverse problem. In Sec. 4.5, it is discussed how this issue can be dealt with. The minus sign is added since the energy is of course supposed to be minimized. It is stressed that Eq. (4.16) and Eq. (4.17) have been first proposed by S. Sorella in Ref. [86] introducing the name *stochastic reconfiguration method* in this field of physics. This is discussed in detail in Ref. [3], where Eq. (4.18) is derived for real parameters using Lagrange multiplier. Additionally, it is noted that the QNG can be viewed as a type of 2nd-order method, contrary to the SGD being a 1st-order method [87]. The Fisher information metric in the classical case or the QGT in the quantum case is considered a substitute for the Hessian. However, in a strict mathematical sense, this can not be correct since second derivatives of the objective function do not show up, but only second derivatives of the variational wave function. The most accurate way to put this is, that second order effects are taken into account by the respective metric tensor [61]. The approach is of course applicable on any Riemannian manifold [67]. For the quantum physical computations, we adopt to the more general designation of (Q)NG.

4.2 Time Evolution

The approach for the ground state optimization can be generalized to arbitrary evolution operators. The ones of distinguished physical interest are those of the real- and imaginary-time evolution. The Schrödinger equation describes the time evolution of a certain state of a quantum mechanical system.

$$\frac{d|\Psi(t)\rangle}{dt} = -\omega H|\Psi(t)\rangle, \quad (4.19)$$

where $\omega = i$ for the real-time evolution and $\omega = 1$ for the imaginary-time evolution. For notational clarity, the variable τ will be used in the following for the imaginary time, while the real time is denoted by t .

$$d_\tau|\Psi(\tau)\rangle = -H|\Psi(\tau)\rangle \quad d_t|\Psi(t)\rangle = -iH|\Psi(t)\rangle \quad (4.20)$$

The Schrödinger equation for a time-independent Hamiltonian and an explicitly normalized state

$$|\Psi(\tau)\rangle = \frac{1}{\sqrt{\langle\Psi(0)|e^{-2\tau H}|\Psi(0)\rangle}} e^{-\tau H}|\Psi(0)\rangle \quad (4.21)$$

on the projective Hilbert space $\mathcal{P}(\mathcal{H})$ at all times τ is altered in the following way [69]:

$$d_\tau |\Psi(\tau)\rangle = -(H - E_{\Psi(\tau)}) |\Psi(\tau)\rangle . \quad (4.22)$$

The additional term on the right hand side comes from the derivative of the denominator in Eq. (4.21).

4.2.1 Imaginary-Time Evolution

The imaginary time evolution is first derived for quantum states on the variational manifold of the projective Hilbert space $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$ before it is allowed to extend to the variational manifold of the Hilbert space $\mathcal{V}(\mathcal{H}, \mathbb{C})$. It is emphasized that by the first case it is essentially meant that the norm of the variational state is not allowed to change its length during the evolution. Hence, the projected Schrödinger equation is for this purpose restricted to be

$$\begin{aligned} \mathbf{Q}_{\Psi_{\mathbb{W}}(\tau)} \frac{d}{d\tau} |\Psi_{\mathbb{W}}(\tau)\rangle &= \mathbf{Q}_{\Psi_{\mathbb{W}}(\tau)} \sum_{\alpha} \frac{\partial \mathbb{W}_{\alpha}}{\partial \tau} |\partial_{\mathbb{W}_{\alpha}} \Psi_{\mathbb{W}}(\tau)\rangle \\ &= \sum_{\alpha} \frac{\partial \mathbb{W}_{\alpha}}{\partial \tau} |V_{\mathbb{W},\alpha}(\tau)\rangle = -\mathbf{P}_{\Psi_{\mathbb{W}}(\tau)} H |\Psi_{\mathbb{W}}(\tau)\rangle , \end{aligned} \quad (4.23)$$

where the two projectors $\mathbf{Q}_{\Psi_{\mathbb{W}}(\tau)}$ and $\mathbf{P}_{\Psi_{\mathbb{W}}(\tau)}$ were defined in Eq. (3.6) and Eq. (3.23), respectively. By multiplying the last line in Eq. (4.23) with the vector $\langle V_{\mathbb{W},\gamma}(\tau) |$ from the left, the following matrix equation describing the time evolution of the state $|\Psi_{\mathbb{W}}(\tau)\rangle$ via a change of the variational parameters $\mathbb{W}[\tau]$ is obtained

$$\frac{\partial \mathbb{W}}{\partial \tau} = -\mathbf{G}^{-1}(\mathbb{W}) \mathbf{F}(\mathbb{W}) \quad \Rightarrow \quad \mathbb{W}[\tau + \Delta\tau] \approx \mathbb{W}[\tau] - \Delta\tau \mathbf{G}^{-1}(\mathbb{W}) \mathbf{F}(\mathbb{W}), \quad (4.24)$$

where $\mathbf{G}(\mathbb{W})$ is the metric tensor from Eq. (4.16) and $\mathbf{F}(\mathbb{W})$ the energy force from Eq. (4.9). The computation is step by step explained in appendix A. The respective equation for the variational manifold of the Hilbert space $\mathcal{V}(\mathcal{H}, \mathbb{C})$ is readily adjusted by considering the distinction with respect to the local basis defined in Eq. (3.14). This was derived in Sec. 3.2.2. The QGT in Eq. (4.16) and the energy forces in Eq. (4.9) are altered by leaving out the respective second term

$$\hat{G}_{\mu\nu}(\mathbb{W}) = \langle O_{\mu}^* O_{\nu} \rangle , \quad \hat{F}_{\mu}(\mathbb{W}, \mathcal{V}) = \langle O_{\mu}^* H \rangle . \quad (4.25)$$

The calculation is in detail performed in appendix B. This way, the imaginary-time evolution on the Hilbert space and the projective Hilbert space are very similar and the respective second term of the QGT and the energy forces controls the change of the norm of the state. All in all, it is stressed that Eq. (4.18), originally derived as the QNG of the ground state optimization, is recovered to be equal to the one for the imaginary-time evolution on the variational manifold of the projective Hilbert space $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$. Due to its significance, it is stated once again at this point.

$$\tilde{\mathbf{F}}(\mathbb{W}) = -\mathbf{G}^{-1}(\mathbb{W}) \cdot \mathbf{F}(\mathbb{W}) \quad (4.26)$$

The similarity of those equations becomes clear by viewing the imaginary-time evolution as a projection technique onto the ground state [3].

4.2.2 Real-Time Evolution

The equations for the real-time evolution naturally differ from the ones for the imaginary-time evolution only by a factor of i . However, due to their significance, they should be noted here at least once [11]:

$$\tilde{\mathbf{F}}(\mathbb{W}) = -i \mathbf{G}^{-1}(\mathbb{W}) \cdot \mathbf{F}(\mathbb{W}) \quad (4.27)$$

This is also known as the time-dependent variational principle [88]. It is emphasized that the real-time evolution alters neither the energy nor the norm of a quantum state [3].

4.3 Adaptive Heun Method

While for the energy optimization, the Euler method is usually already sufficient, for the real- or imaginary-time evolution, the trajectory needs to be approximated accurately and thus it is of interest to reduce the error of the time evolution. Typically, one would start with a Runge-Kutta method of an order that seems sufficient for the simulation at hand. However, the propagation step of the Runge-Kutta method is constant, no matter what the landscape of the manifold at hand looks like at a specific point in time. Loosely speaking, the Runge-Kutta method traverses landscapes in which the parameters change drastically at the same speed as it traverses landscapes in which the parameters are barely changing. This issue is overcome by an adaptive algorithm, like the adaptive Heun method. This is an iterative method, where in every time step the error of the integration is estimated and the length of the time step $\Delta\tau$ or Δt adjusted such that the error stays constant. It can be used for the real- and imaginary-time evolution and is explained in the following for imaginary time. We describe this predictor-corrector method in the following for a second order integration, for which the standard notation for the derivative $\dot{\mathbb{W}} = f(\mathbb{W})$ and the exact solution $\mathbb{W}(\tau + \Delta\tau)$ at time $\tau + \Delta\tau$, where \mathbb{W} is a P -component vector and P the number of parameters, is used [89]. We define

$$\begin{aligned} \mathbf{k}_1 &= f(\mathbb{W}[l]) , \\ \mathbf{k}_2 &= f(\mathbb{W}[l] + \Delta\tau \cdot \mathbf{k}_1) , \\ \mathbb{W}[l + 1] &= \mathbb{W}[l] + \frac{\Delta\tau}{2} (\mathbf{k}_1 + \mathbf{k}_2) , \end{aligned} \quad (4.28)$$

where l denotes the l -th step evolution. This leads to a cubic error in the integration step

$$\mathbb{W}[l + 1] = \mathbb{W}(\tau + \Delta\tau) + \boldsymbol{\alpha}(\Delta\tau)^3, \quad (4.29)$$

where α is a constant vector. This is now compared to the error for two time steps of length $\frac{\Delta\tau}{2}$

$$\mathbb{W}'[l+1] = \mathbb{W}(\tau + \Delta\tau) + \underbrace{2\alpha \left(\frac{\Delta\tau}{2}\right)^3}_{\delta\mathbf{1}}, \quad (4.30)$$

where δ is the integration error and $\mathbf{1}$ the unity vector. The difference of these two results

$$\|\mathbb{W}[l+1] - \mathbb{W}'[l+1]\| = \frac{3}{4}\|\alpha\|(\Delta\tau)^3 \equiv 3\delta \quad (4.31)$$

allows the computation of the integration error δ and thus to adjust the time step

$$\Delta\tau' = \Delta\tau \left(\frac{\varepsilon}{\delta}\right)^{\frac{1}{3}} \quad (4.32)$$

to provide an accuracy of ε for the time evolution. The norm in Eq. (4.31) is calculated for any metric space by $\|\mathbf{x}\| = \frac{1}{P}\sqrt{\mathbf{x}^\dagger\mathbf{G}\mathbf{x}}$, where \mathbf{G} the metric tensor, which is in the variational manifold given by the QGT in Eq. (3.16). Note that the tolerance ε and the initial step size $\Delta\tau$ are in principle parameters which need to be specified at the beginning of the algorithm. The adjusted step size $\Delta\tau'$ is always used as the proposed step size $\Delta\tau$ in the next step. By applying a pretempering at $\tau = 0$ of a few steps in which parameters are not changed, $\Delta\tau'$ converges to a value that is used as the initial step size. Thus, the only remaining parameter to be fixed is the desired accuracy ε , which defines the accuracy of the method. In this work, the number of iterative steps in the Heun algorithm is denoted by R .

4.4 Imaginary-Time Evolution on the Manifold and β Modification

The variational time evolution in Eqs. (4.26) and (4.27) needs a more thorough consideration that takes into account the projection onto the variational manifold $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$ of the projective Hilbert space. In the following, the imaginary-time evolution is focused upon, where the imaginary time τ is modified to $\beta(\tau)$. The notation will make sense for the finite-temperature calculations implemented in Sec. 8, but is applicable for any evolution operator. First, from the action of the projector $\mathbb{P}_{\Psi_{\mathbb{W}}(\tau)}$ (Eq. (3.23)), the computation in Eq. (4.23) is extended by

$$\mathbb{P}_{\Psi_{\mathbb{W}}(\tau)}H|\Psi_{\mathbb{W}}(\tau)\rangle = \mathbb{P}_{\Psi_{\mathbb{W}}(\tau)}(H - E_{\Psi_{\mathbb{W}}(\tau)})|\Psi_{\mathbb{W}}(\tau)\rangle, \quad (4.33)$$

since $\mathbb{P}_{\Psi_{\mathbb{W}}(\tau)}|\Psi_{\mathbb{W}}(\tau)\rangle = 0$ [59]. The energy at imaginary time τ is denoted by $E(\tau)$. Due to the projection, the time traveled in the manifold and the actual imaginary time in the algorithm are not

equal. This is quantified by the following ratio

$$\begin{aligned}\gamma_1 &= \frac{\|\mathbf{P}_{\Psi_{\mathbb{W}}(\tau)}(H - E_{\Psi_{\mathbb{W}}(\tau)})|\Psi_{\mathbb{W}}(\tau)\rangle\|^2}{\|(H - E_{\Psi_{\mathbb{W}}(\tau)})|\Psi_{\mathbb{W}}(\tau)\rangle\|^2} \\ &= -\frac{2}{(\Delta H)_{\Psi_{\mathbb{W}}(\tau)}^2} \left(\frac{dE_{\Psi_{\mathbb{W}}(\tau)}}{d\tau} \right),\end{aligned}\tag{4.34}$$

where the two quantities of interest on the right hand side

$$E_{\Psi_{\mathbb{W}}(\tau)} = \frac{\langle \Psi_{\mathbb{W}}(\tau) | H | \Psi_{\mathbb{W}}(\tau) \rangle}{\langle \Psi_{\mathbb{W}}(\tau) | \Psi_{\mathbb{W}}(\tau) \rangle}\tag{4.35}$$

$$(\Delta H)_{\Psi_{\mathbb{W}}(\tau)}^2 = \frac{\langle \Psi_{\mathbb{W}}(\tau) | (H - E_{\Psi_{\mathbb{W}}(\tau)})^2 | \Psi_{\mathbb{W}}(\tau) \rangle}{\langle \Psi_{\mathbb{W}}(\tau) | \Psi_{\mathbb{W}}(\tau) \rangle}\tag{4.36}$$

are directly calculated at every time step [90]. For greater clarity, the action of the projector in Eq. (4.34) is calculated step by step

$$\begin{aligned}\frac{dE_{\Psi_{\mathbb{W}}(\tau)}}{d\tau} &= \frac{\langle \Psi_{\mathbb{W}}(\tau) | H \left(\frac{d}{d\tau} | \Psi_{\mathbb{W}}(\tau) \rangle \right)}{\langle \Psi_{\mathbb{W}}(\tau) | \Psi_{\mathbb{W}}(\tau) \rangle} + h.c. \\ &= -\frac{\langle \Psi_{\mathbb{W}}(\tau) | H \left(\mathbf{P}_{\Psi_{\mathbb{W}}(\tau)}(H - E_{\Psi_{\mathbb{W}}(\tau)}) | \Psi_{\mathbb{W}}(\tau) \rangle \right)}{\langle \Psi_{\mathbb{W}}(\tau) | \Psi_{\mathbb{W}}(\tau) \rangle} - h.c. \\ &= -2 \frac{\langle \Psi_{\mathbb{W}}(\tau) | (H - E_{\Psi_{\mathbb{W}}(\tau)}) \mathbf{P}_{\Psi_{\mathbb{W}}(\tau)}^2 (H - E_{\Psi_{\mathbb{W}}(\tau)}) | \Psi_{\mathbb{W}}(\tau) \rangle}{\langle \Psi_{\mathbb{W}}(\tau) | \Psi_{\mathbb{W}}(\tau) \rangle}.\end{aligned}\tag{4.37}$$

In the final line, we employed Eq. (4.33) and leveraged a property of projection operators, namely, $\mathbf{P}_{\Psi_{\mathbb{W}}(\tau)}^2 = \mathbf{P}_{\Psi_{\mathbb{W}}(\tau)}$. The quantities in Eq.(4.36) can be computed at every time step in the algorithm. The derivative of the energy is afterwards computed numerically in a second order method from the energy results $E_{\Psi_{\mathbb{W}}(\tau)}$. The imaginary time τ is therefore adjusted and the modified inverse temperature $\beta(\tau)$ computed by integrating Eq. (4.34).

$$\gamma_1 = \frac{d\beta(\tau)}{d\tau} \quad \Rightarrow \quad \beta(\tau) = - \int_0^\tau \frac{1}{(\Delta H)_{\Psi_{\mathbb{W}}(\tau')}^2} \frac{dE_{\Psi_{\mathbb{W}}(\tau')}}{d\tau'} \Big|_{\mathbb{W}(\tau')} d\tau' \tag{4.38}$$

We emphasize that the temperature adjustment does not fully correct the errors due to the restriction of the time evolution to the variational manifold $\mathcal{V}(\mathcal{P}(\mathcal{H}), \mathbb{C})$. The proper state in the full Hilbert space is not accessible and, in general, different from the variational state. However, this modification clearly signifies whenever the length of the tangent vector of the imaginary-time evolution compared to the quantum velocity defined in Eq. (3.38) decreases drastically and takes that into account by modifying the imaginary time. Similarly, a β modification can be computed for the imaginary-time evolution on the variational manifold of the Hilbert space $\mathcal{V}(\mathcal{H}, \mathbb{C})$ as well. It is emphasized that the variational state is now also allowed to change its length at every iteration step. The modification

factor is similarly to Eq. (4.34) computed to be

$$\gamma_2 = \frac{||\hat{P}_{\Psi_{\mathbb{W}}(\tau)} H |\Psi_{\mathbb{W}}(\tau)\rangle||^2}{||H |\Psi_{\mathbb{W}}(\tau)\rangle||^2} = -\frac{E_{\Psi_{\mathbb{W}}(\tau)}}{(H^2)_{\Psi_{\mathbb{W}}(\tau)}} \left(\frac{dE_{\Psi_{\mathbb{W}}(\tau)}}{d\tau} \right). \quad (4.39)$$

4.5 Regularization

The matrix inversion in Eqs. (4.18), (4.26) and (4.27) is a crucial part of the QNG approach. It needs to be treated carefully, since it is a well-known discrete ill-posed problem also known as the inverse problem [91]. Therefore, it requires some sort of regularization. The matrix dimensions are given by the number of parameters \mathbb{W} , which are denoted as before by P . In this section, there is a more mathematical notation used to underline the universality of the statements. Let $\mathbf{b} \in \mathbb{C}^P$ and $\mathbf{G} \in \mathbb{C}^{P \times P}$ be given and $\mathbf{x} \in \mathbb{C}^P$ the quantity of interest. Then the following system of equations

$$\mathbf{G}\mathbf{x} = \mathbf{b} \quad (4.40)$$

has to be solved. This can be formulated as a minimization problem known as the least-squares problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} ||\mathbf{G}\mathbf{x} - \mathbf{b}||^2, \quad (4.41)$$

where $||\cdot||$ is the Euclidean norm. If Monte Carlo sampling is used, it has to be taken into account that the quantities $\tilde{\mathbf{G}}$ and $\tilde{\mathbf{b}}$ contain statistical errors that are limited by $||\tilde{\mathbf{G}} - \mathbf{G}|| \leq \delta_{\mathbf{G}}$, $||\tilde{\mathbf{b}} - \mathbf{b}|| \leq \delta_{\mathbf{b}}$ [92]. The Singular Value Decomposition (SVD) of $\tilde{\mathbf{G}} \in \mathbb{C}^{P \times P}$

$$\tilde{\mathbf{G}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger, \quad (4.42)$$

where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_P] \in \mathbb{C}^{P \times P}$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_P] \in \mathbb{C}^{P \times P}$ are unitary matrices and the diagonal matrix $\mathbf{\Sigma} = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_P] \in \mathbb{R}^{P \times P}$ with $\sigma_j \geq 0 \forall j$ the singular values σ_j of $\tilde{\mathbf{G}}$ ordered in ascending order, is used to write $\tilde{\mathbf{G}}$ in the following way [93, 91]:

$$\tilde{\mathbf{G}} = \sum_{j=1}^P \sigma_j \mathbf{u}_j \mathbf{v}_j^\dagger \quad (4.43)$$

To solve the inverse problem in Eq. (4.40), the inverse of the matrix $\tilde{\mathbf{G}}$

$$\tilde{\mathbf{G}}^{-1} = \sum_{j=1}^P \mathbf{v}_j \sigma_j^{-1} \mathbf{u}_j^\dagger \quad (4.44)$$

needs to be computed. This is unstable if the singular values are very small or even zero and leads to rounding errors [91]. Additionally, the errors $\delta_{\mathbf{G}}$ and $\delta_{\mathbf{b}}$ are pronounced. Thus, there is a regularization introduced that is of course more subtle in the case of real- or imaginary-time evolution. For the ground state optimization, the regularization is in principle allowed to alter the

trajectory of the minimization, as long as the correct ground state is reached and the number of steps does not grow substantially.

4.5.1 Ground State Optimization

The most straightforward approach is to eliminate singular values by truncating those that are smaller than a cutoff parameter ϵ_{pinv} . This is achieved by redefining the inverse of the singular values as

$$\sigma_j^+ = \begin{cases} 0, & \text{if } \left| \frac{\sigma_j}{\sigma_{max}} \right| < \epsilon_{pinv} \\ \sigma_j^{-1}, & \text{if } \left| \frac{\sigma_j}{\sigma_{max}} \right| \geq \epsilon_{pinv} \end{cases}. \quad (4.45)$$

Essentially, this corresponds to the Moore-Penrose pseudo-inverse with a finite cutoff parameter, specified by

$$\tilde{\mathbf{G}}^+ = \sum_{j=1}^P \mathbf{v}_j \sigma_j^+ \mathbf{u}_j^*. \quad (4.46)$$

The Moore-Penrose pseudo-inverse is for singular matrices the solution of the least-squares problem of Eq. (4.41) with respect to the Euclidean norm [94, 93]. This regularization method is used in the ground state optimization and also in case of time evolution [11]. In the ground state optimization methods, in which the correct explicit path to the ground state is not of interest, additional regularization is often applied to further mitigate potential issues. It consists of adding a constant factor to the diagonal of the matrix $\tilde{\mathbf{G}}$

$$\min_{\mathbf{x} \in \mathbb{C}^n} \{ \|\tilde{\mathbf{G}}\mathbf{x} - \tilde{\mathbf{b}}\|^2 + \|\epsilon_T \mathbf{1}\mathbf{x}\|^2 \}, \quad (4.47)$$

which amounts to a shift of all singular values [95]. The parameter ϵ_T is the Tikhonov regularization parameter and can be chosen to be step dependent [11]. This approach is a specific version of the more general Tikhonov regularization, where a penalizing term is added to control the size of the solution vector

$$\min_{\mathbf{x} \in \mathbb{C}^n} \{ \|\tilde{\mathbf{G}}\mathbf{x} - \tilde{\mathbf{b}}\|^2 + \|\mathbf{L}_{\epsilon_T} \mathbf{x}\|^2 \}, \quad (4.48)$$

where the matrix \mathbf{L}_{ϵ_T} is called the regularization matrix [91]. However, to make a reasonable choice for \mathbf{L}_{ϵ_T} , some knowledge about the nature of the solution is required. Thus, the most popular choice for \mathbf{L}_{ϵ_T} is given by the identity matrix $\mathbf{L}_{\epsilon_T} = \epsilon_T \mathbf{1}$ in Eq. (4.47). The most common way to derive the regularization parameter is by the generalized discrepancy principle [92]. To obtain ϵ_T the following function has to be minimized:

$$f(\epsilon_T) = \|\tilde{\mathbf{G}}\mathbf{x} - \tilde{\mathbf{b}}\|^2 - (\delta_{\tilde{\mathbf{b}}} + \delta_{\tilde{\mathbf{G}}} \cdot \epsilon_T \|\mathbf{L}_{\epsilon_T} \mathbf{x}\|)^2 = 0 \quad (4.49)$$

Since the function $f(\epsilon_T)$ is convex for small ϵ_T but concave for large ones, global and monotone convergence of Newton's method can not be guaranteed [96, 93]. There are ideas to transform this equation in a way that two free parameters are obtained [97, 92]. Under some smoothness assumptions for $\tilde{\mathbf{G}}$ and \mathbf{L}_{ϵ_T} , these parameters can be chosen in a way that convergence is guaranteed.

Those parameters are computed by an iterative algorithm [97]. Given its advanced nature, this algorithm is not typically constrained by a specific number of iterations, and its assumptions may not always be fulfilled in variational problems. As a result, it is not commonly employed in such applications. However, it proves to be a valuable method when one seeks to examine the parameter update at a particular step in the algorithm. Nevertheless, in the majority of practical scenarios, regularization utilizing the unity matrix (Eq. (4.47)) is applied, where ϵ_T is chosen to be step-dependent [11], or it may be contingent on the appearance of parameters in the ansatz [95].

4.5.2 Time Evolution

If one is also interested in the time evolution and not only in the final result the algorithm converges to, a more cautious regularization becomes necessary. First, it is noted that the formulation of the adjusted Moore-Penrose pseudo inverse in Eq. (4.46) can be formulated for hermitian matrices in a similar way using the eigenvalue decomposition

$$\tilde{\mathbf{G}} = \mathbf{V}\mathbf{D}\mathbf{V}^+, \quad (4.50)$$

instead of the singular value decomposition in Eq. (4.42), where $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_P] \in \mathbb{C}^{P \times P}$ and \mathbf{v}_k the eigenvectors of $\tilde{\mathbf{G}}$, $\mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_N) \in \mathbb{R}^{P \times P}$ and $\lambda_1, \dots, \lambda_N$ the eigenvalues of $\tilde{\mathbf{G}}$. These are real because hermitian matrices are unitarily similar to a real diagonal matrix [93, 98]. Thus, the pseudo-inverse can also be defined as

$$\tilde{\mathbf{G}}^+ = \sum_{j=1}^w \mathbf{v}_j \lambda_j^+ \mathbf{v}_j^\dagger, \quad (4.51)$$

where

$$\lambda_j^+ = \begin{cases} 0, & \text{if } \left| \frac{\lambda_j}{\lambda_{max}} \right| < \epsilon_{pinv} \\ \lambda_j^{-1}, & \text{if } \left| \frac{\lambda_j}{\lambda_{max}} \right| \geq \epsilon_{pinv} \end{cases}, \quad (4.52)$$

and $|\lambda_{max}|$ is the maximum eigenvalue in terms of the modulus. This enables the establishment of a so-called *soft cutoff*

$$\lambda_j^+ = \left[\lambda_j \left(1 + \left(\frac{\epsilon_{pinv}}{\left| \frac{\lambda_j}{\lambda_{max}} \right|} \right)^6 \right) \right]^{-1} \quad (4.53)$$

to allow parameters to slowly cross the line to being taken into account and vice versa [98]. It is stressed that the form of Eq. (4.53) is proposed in Ref. [98] as one possibility. It is aimed to prohibit discontinuities. These appear, if parameters, that have fallen into the first class in Eq. (4.52), change at a certain time step to fulfill the conditions of the second class. Therefore they contribute suddenly while being neglected before. Additionally, small inaccuracies due to the sampling error might be enhanced by taking the inverse. Hence, there is another cutoff introduced that is controlled by the

signal-to-noise ratio (SNR) of the transformed force vector $\mathbf{F}(\mathbf{V}) = \mathbf{V}^\dagger \mathbf{b}$ [89].

$$\lambda_j^+ = \left[\lambda_j \left(1 + \left(\frac{\epsilon_{SNR}}{SNR(\rho_j)} \right)^6 \right) \right]^{-1} \quad (4.54)$$

This leaves the method with two cutoff parameters ϵ_{pinv} and ϵ_{SNR} , which need to be tuned for the model, respectively. The SNRs for the relevant quantities are derived in Ref. [89]. The SNRs of the eigenvalues of the matrix $\tilde{\mathbf{G}}$ do not depend on their magnitude

$$SNR(\lambda_j) = \sqrt{\frac{N_{Met}}{2}} \quad \forall j, \quad (4.55)$$

while the SNR of the forces F_j do.

$$SNR(F_j(\mathbf{V})) = \sqrt{\frac{N_{Met}}{1 + \frac{\lambda_j}{|F_j(\mathbf{V})|^2} \Delta H}} \quad (4.56)$$

Chapter 5

Neural Network (NN)

While in the previous section the general VMC algorithm was explained, the most important ingredient for this method, namely the ansatz wave function, has been set aside for this section. Historically, the first ideas for a variational ansatz were based on physical observations regarding small systems. This is thoroughly explained in Sec. 6. However, in a seminal paper by G. Carleo and M. Troyer in 2017, the idea of using (artificial) neural networks (NNs) to approximate wave functions in quantum many-body models was proposed [11]. We want to derive why NNs are essentially, in general, a promising tool to approximate functions in a high-dimensional space. To do so, there are in principle two main roads to go along. One is to focus on the practical aspects and the link to biological neural circuits using examples of models in low-dimensional space to illustrate the expressive capabilities of NNs. The other one is to keep on the theoretical side and to elaborate on the theorems that motivate why NNs are the perfect candidate to approximate functions accurately without the need of specific knowledge about the nature of this function at hand. We decided for obvious reasons to go with the latter approach and complement every now and then some illustrative examples of the former one. This will lead to a combined ansatz of using mathematical theorems and the depictions of NNs. But first, we want to introduce this topic by a short historical reminiscence whose connection to NNs becomes clear in the proceeding of this section.

5.1 Hilbert's 13-th Problem

D. Hilbert published his famous 23 mathematical problems at the conference of the international congress of mathematicians in Paris in 1900 [99]. While the general motivation for the 13-th problem comes from the field of nomography and leads too far into a mathematical field which is not treated in this work, Hilbert summarized the problem in the headline he used: “Impossibility of the Solution of the General Equation of the 7-th Degree by Means of Functions of only Two Arguments” [99]. Although referring originally to algebraic functions, he assumes in a later publication that the 7-th degree equation can not be solved by arbitrary continuous functions of two arguments [100]. The reason why Hilbert referred to the 7-th degree equation in specific, will be clear by studying his

calculations. The general n-th degree equation

$$x^n + a_1x^{n-1} + \dots + a_n = 0 , \quad (5.1)$$

with $x \in \mathbb{R}$ and $a_1, \dots, a_n \in \mathbb{R}$ the respective coefficients, can be transformed by a Tschirnhaus transformation [101, 102]. In general, this is achieved by using the ansatz

$$y = x^{n-1} + b_1x^{n-2} + \dots + b_{n-1} \quad (5.2)$$

for the variable transformation, where b_1, \dots, b_{n-1} are coefficients that need to be fixed [100]. The n-th degree equation for y is then obtained as

$$y^n + c_1y^{n-1} + \dots + c_n = 0 , \quad (5.3)$$

where $c_i = c_i(a_1, \dots, a_n, b_1, \dots, b_{n-1})$. Comparing Eq. (5.1) and Eq. (5.3), it seems like the transformation did not have any effect. However, the crucial point is that the coefficients b_j in Eq. (5.2) can be chosen in a way that some of the coefficients c_j in Eq. (5.3) are actually 0 [100].

$$y^n + c_4x^{n-4} + c_3x^{n-3} + \dots + c_n = 0 \quad (5.4)$$

The simplest example of a Tschirnhaus transformation is considered for the 3rd degree equation, for which the quadratic term can be chosen to vanish [101, 102]. This yields the following equations for degree 5 to 7

$$\begin{aligned} y^5 + c_1y + 1 &= 0 \\ y^6 + c_2y^2 + c_1y + 1 &= 0 \\ y^7 + c_4^3 + c_3^2 + c_2y + 1 &= 0 , \end{aligned} \quad (5.5)$$

where it becomes apparent that in the 7-th degree equation, three coefficients occur for the first time. D. Hilbert stated the problem in his famous list of 23 mathematical problems by wondering if the 7-th degree equation is solvable with the help of any continuous functions of only two arguments [99]. However, the true meaning of Hilbert's 13-th problem is to this day the topic of a heated debate, basically evolving around the question if these functions have to be algebraic or only continuous [103]. Although the outline of the problem is clear, there is evidence to suggest that D. Hilbert simply forgot the term *analytic* or *algebraic* [104]. Fortunately, in this work, only the application of continuous functions matters, allowing for the enjoyment of the historical controversy without delving further into it.

5.2 Kolmogorov-Arnold Representation Theorem

While Hilbert's 13-th problem for algebraic functions is still not answered, V.I. Arnold solved it in 1957 for continuous functions by using a theorem of his supervisor A. Kolmogorov in Ref. [105]. The original paper was published in Ref. [106] and translated in Ref. [107]. The theorem has been

refined and is often referred to as the Kolmogorov-Arnold representation theorem to honour the two founders.

Theorem 1. *Kolmogorov-Arnold Representation Theorem [108, 109]*

Let $f : I_N \rightarrow \mathbb{R}$ be a continuous function on the unit cube $I_N = [0, 1]^N$. Then f can be represented in the following form

$$f(x_1, x_2, \dots, x_N) = \sum_{m=1}^{2N+1} \Theta_m \left[\sum_{n=1}^N \Phi_{mn}(x_n) \right], \quad (5.6)$$

where Θ_m and Φ_{mn} are continuous functions. The functions Φ_{mn} do not depend on the nature of f .

The last sentence already hints to a possible formidable applicability of this theorem and the intriguing idea to use it to approximate wave functions in the VMC algorithm (see Sec. 4). It basically gives rise to the idea of a universal ansatz wave function.

A nice survey describing the aftermath of this theorem and the upcoming refinements is found in Ref. [104]. The sole property attributed to the functions Φ_{mn} , as mentioned in Ref. [105], was that they are not smooth. Hence, and due to the lack of knowledge about the nature of the functions Θ_m , the direct practical applicability of this theorem was doubted. The theorem was even judged to be irrelevant for NNs [110, 111]. However, later works revealed that the operations of addition, multiplication by a number, superposition and a continuous nonlinear function are sufficient for the approximation of an arbitrary function $f(\mathbf{x})$ [112]. Thus, it proved successful to take a step back and investigate to what extent continuous functions can be approximated by a superposition of a similar form as in Eq. (5.6). This led to theorems that are directly connected to what is known as a special type of NN, namely the feed-forward neural network (FFNN).

5.3 Single-Layer Feed-Forward Neural Network (SLFFNN)

Instead of aiming to represent a continuous function f of interest, we are for the moment content with approximating this particular function. In this sense, there were similar theorems obtained that are all considered to be a form of the universal approximation theorem (see Refs. [113, 114, 115]). This universal behavior, already shining through in Th. 1, was first proven in Ref. [114], where the name was coined as well. However, since they focus on Borel measurable functions, which leads deep into a mathematical field that is not that relevant here, we will cite a theorem by G. Cybenko that considers, as before, continuous functions on the unit cube.

Theorem 2. *Universal Approximation Theorem [115]*

Let σ be any continuous sigmoidal function defined as

$$\sigma(x) = \begin{cases} 1, & \text{for } x \rightarrow \infty \\ 0, & \text{for } x \rightarrow -\infty \end{cases} . \quad (5.7)$$

Then, finite sums of the form

$$G(\mathbf{x}) = \sum_{m=1}^M a_m \sigma \left(\sum_{n=1}^N W_{mn} x_n + b_m \right) \quad (5.8)$$

are dense in the space of continuous functions $C(I_N)$ on the N -dimensional unit cube $I_N = [0, 1]^N$. In other words, given any $f \in C(I_N)$ and $\epsilon > 0$, there is a sum, $G(\mathbf{x})$, of the above form, for which

$$|G(\mathbf{x}) - f(\mathbf{x})| < \epsilon \quad \forall \mathbf{x} \in I_N .$$

This theorem shows that sums of sigmoidal functions can approximate high-dimensional real and continuous functions to any desired degree of accuracy. The meaning of being a universal approximator becomes immediately clear since the principle ansatz of the sigmoidal functions is not dependent on the specific function f , but universal. The most prominent specific choice for a sigmoidal function is the sigmoid function, also known as logistic function, which is defined as

$$\sigma(x) = \frac{1}{2} \left(1 + \tanh \left(\frac{x}{2} \right) \right) . \quad (5.9)$$

However, it is in principle not clear how large M and thus the number of overall parameters $\mathbf{a}, \mathbf{b} \in \mathbb{R}^M$ and $\mathbf{W} \in \mathbb{R}^{M \times N}$ needs to be, where \mathbf{W} is known as the weight matrix and \mathbf{b} as the bias. It is stressed that the actual size of M is crucial for the practical relevance of this theorem and is thus discussed in a separate section, namely Sec 5.9. The restriction on f to be continuous on the unit cube can be loosened to be continuous on a compact subset $M \subset \mathbb{R}^N$ if the activation function is only continuous, bounded and non-constant [116]. The distance is in that case measured by the L^p -norm.

To understand the term activation function, the connection between the mathematical considerations discussed above and the more illustrative depictions of NNs is discussed in the following. Detailed explanations can be found in Ref. [113]. Firstly, the NNs studied in this work are so-called feed-forward neural networks (FFNN) meaning that there is one direction of information flow producing an output for every input. The right hand side of Eq. (5.8) is directly expressed in a graphical notation in Fig. 5.1 defining the input, the hidden and the output layer. Its elements are often referred to as nodes or units. The term *hidden layer* was coined because the state or value of a hidden node is not considered explicitly, leaving its actual state or value hidden. The number of input nodes is denoted throughout this work by the letter N , while the number of hidden nodes is denoted by M . The sigmoid function used in Eq. (5.8) is in general referred to as the activation

function governing the activation of the proceeding nodes. This stems from the notion of an NN being a model for the functioning of a human brain, in which the neurons are modelled as binary nodes and fire or are activated if the value of the activation function exceeds a certain value. The idea of using sigmoidal functions is motivated by this idea too, with output values ranging from 0 to 1 in a smooth way. To gain a practical understanding, it is recommended to study the visual proof of Th. 2, in Ref. [117] in which the effect of the weight matrix \mathbf{W} and the bias \mathbf{b} is outlined for a minimal FFNN of one input node and two hidden nodes. Additionally, it is very helpful to go through the example of learning the XOR function in Ref. [113] to understand why linear activation functions are not sufficient in some cases, especially when the domain of the function to approximate consists of a certain number of discrete values.

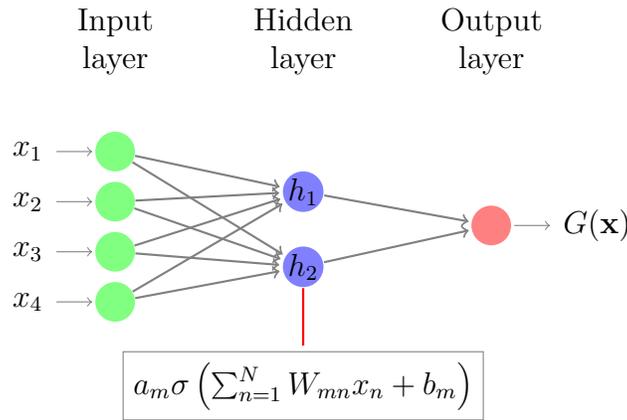


Figure 5.1: Depiction of the SLFFNN described by Eq. (5.8). The state of the hidden nodes is depicted in the box. The number of hidden nodes is in principle arbitrary and determines the accuracy of the approximation.

5.4 Multi-Layer Feed-Forward Neural Network (MLFFNN)

Naturally, one can think of adding another hidden layer to the SLFFNN depicted in Fig. 5.1 to improve the accuracy of the approximation or, to speak in terms of the respective Eq. (5.8), one can add another function and compose it with the existing activation function in Eq. (5.8). The question is, if this can in principle, as is indicated by Th. 1, not only lead to an approximation but a representation of the desired function f . Unfortunately, the Kolmogorov-Arnold representation theorem provides limited insights into the precise characteristics of the inner functions Φ_{mn} while the outer functions Θ_m are even dependent on the desired function f . The latter inhibits a universality feature of the representation contrary to the one that is achieved with one hidden layer by Th. 2. This property would be of course highly desirable. Unfortunately, the Kolmogorov-Arnold representation theorem in general has lacked the ability to be the direct building block of other mathematical theorems as well as the ability of being of direct practical relevance due to the dependence of the functions Θ_m on f [118, 119]. However, over the years, the Kolmogorov-Arnold representation

theorem has been refined in numerous ways, see for example Refs.[120, 121, 122]. A nice and extensive review on this topic can be found in Ref. [110]. However, a refined version of the theorem leads us to the definition of multi-layer feed-forward neural networks (MLFFNN).

Theorem 3. *Theorem by J.Braun [122]*

Let $f : I_N \rightarrow \mathbb{R}$ be a continuous function on the unit cube $I_N = [0, 1]^N$. Then f can be represented in the following form

$$f(\mathbf{x}) = \sum_{m=1}^{2N+1} \Theta \left[\sum_{n=1}^N a_n \Phi(x_n + mb) + c_m \right], \quad (5.10)$$

where $\Theta : \mathbb{R} \rightarrow \mathbb{R}$ and $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ are continuous functions and $\mathbf{a} \in \mathbb{R}^N$, $\mathbf{b} \in \mathbb{R}^N$ and $\mathbf{c} \in \mathbb{R}^{2N+1}$. The function Φ does not depend on the nature of f , while there is an algorithm described in Ref. [122] to compute the function Θ .

The right hand side of Eq. (5.10) is conveniently depicted as a MLFFNN with two hidden layers and $N = 2$ as was done originally already by the author in Ref. [122]. It is transferred to our presentation format and depicted in Fig. 5.2 for the example of two input nodes. A NN with more than one hidden layer is also referred to as a deep neural network (DNN). Furthermore, it is stressed that every hidden layer is linked to a specific non-linear activation function. To achieve complete universality in both activation functions, one needs to be content with approximating the desired function rather than representing it. This way, a universal approximation theorem as in Th. 2 can be naturally formulated for a MLFFNN as well. This can be achieved by using sigmoidal activations functions

$$\sum_{m=1}^M a_m \sigma \left(\sum_{n=1}^N W_{mn} x_n + b_m \right) \quad (5.11)$$

for the activation functions of the first hidden layer Φ and the second hidden layer Θ [123]. Moreover, an upper bound is derived for the number of hidden nodes as a function of f . All in all, there is a universal approximation theorem for the SLFFNN and the MLFFNN and a certain variety of activation functions is available. This leaves a lot of room in practical applications to try different architectures of FFNNs. A huge number of examples with architectures varying in depth (number of hidden layers) and width (number of hidden nodes in a certain layer) can be found in Refs. [124, 113]. It is emphasized that this derivation of FFNNs in general coming from the Kolmogorov-Arnold representation theorem is not the historically correct one meaning that there have been proposals of FFNNs in specific and NNs in general before this theorem. One can argue that the starting point of this field of research was with the suggestion of the McCulloch-Pitts neuron to simulate the behavior of a biological neuron [125, 126, 113]. This general approach of explaining the functioning of FFNNs by their similarity to biological neurons is a very prominent and illustrative one. It was especially in the early stages the most essential one. However, it is for research in mathematics and physics also a vague one relying on vivid examples rather than theorems. Thus, and since the general aim in this work is to approximate wave functions, the derivation presented so far with a focus on the theorems starting with the Kolmogorov-Arnold theorem was preferred. Nevertheless, it is important to emphasize that, based on the derivation and illustration of FFNNs, there are

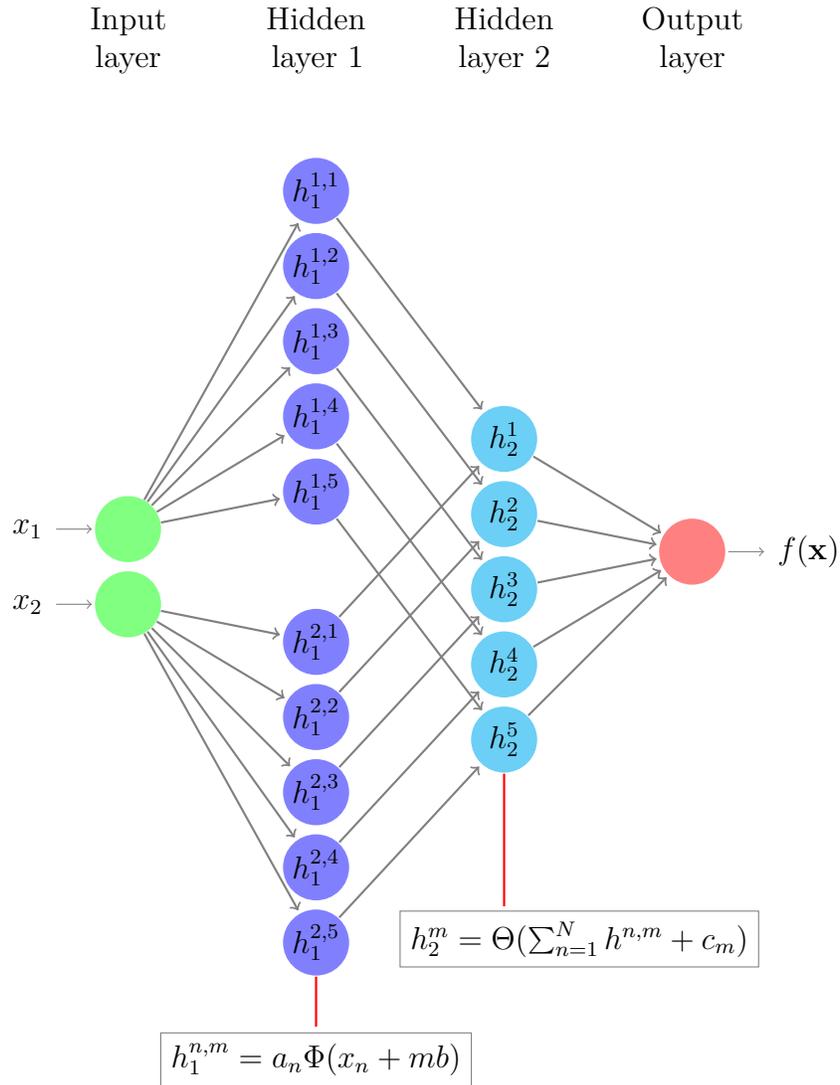


Figure 5.2: Depiction of the MLFFNN described by Eq. (5.10) for two input nodes $N = 2$ [122].

essentially no constraints on imagination when contemplating the width and depth of hidden layers, as well as the nature of their activation functions. The general N_H -Layer FFNN can accordingly be written as

$$f_{\mathbb{W}}(\mathbf{x}) = g_{N_H}(\mathbb{W}_{N_H} \dots (g_1(\mathbb{W}_1, \mathbf{x}))) , \quad (5.12)$$

where $g_j(\mathbb{W}_j)$ is the activation function of the j -th hidden layer with the respective parameter set \mathbb{W}_j [127]. The huge success of several FFNN architectures has led to a very practical approach of testing new architectures and measuring their success in certain fields of applications ranging from speech recognition, image classification and modification to medical diagnosis and finance [113, 128, 129]. Thus, in closing, there is a very general MLFFNN depicted in Fig. 5.3 using sigmoidal functions as the activation function for both hidden layers. It is described by the following equation

$$f(\mathbf{x}) \approx G(\mathbf{x}) = \underbrace{\sum_{m_2=1}^{M_2} c_{m_2} \sigma \left[\sum_{m_1=1}^{M_1} V_{m_2 m_1} \left(\underbrace{a_{m_1} \sigma \left[\sum_{n=1}^N W_{m_1 n} x_n + b_{m_1} \right]}_{h_{m_1}^1} \right) + d_{m_2} \right]}_{h_{m_2}^2} , \quad (5.13)$$

where the values of the hidden nodes are explicitly marked and the parameters of the first hidden layer are $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{M_1}$, $\mathbf{W} \in \mathbb{R}^{M_1 \times N}$ and the ones of the second hidden layer are $\mathbf{c}, \mathbf{d} \in \mathbb{R}^{M_2}$, $\mathbf{V} \in \mathbb{R}^{M_2 \times M_1}$. This provides a picture of how much the number of parameters increases with an additional hidden layer. This can in some applications be circumvented by using parameter sharing, which is directly established by the operation of convolution. Accordingly, these FFNNs are called convolutional neural networks (CNN) and are explained in detail in Refs. [113, 130]. To get an idea of how parameter sharing works, the operation to compute the hidden nodes of the first hidden layer is replaced by a convolution

$$h_{m_1}^1 = a_{m_1} \sigma [(\mathbf{W} * \mathbf{x})_{m_1} + b_{m_1}] \quad (5.14)$$

that is defined as

$$(\mathbf{W} * \mathbf{x})_{m_1} = \sum_{n=1}^N W_n x_{\text{mod}[n-m_1, N]+1} , \quad (5.15)$$

where mod the modulo operator ensuring that the argument is always a correct index of the vector $\mathbf{W} \in \mathbb{C}^N$. This way, the weights are not exclusive for every connection between input and first hidden layer in Fig. 5.3, but one column of the former weight matrix \mathbf{W} is shared, reducing the weight matrix effectively to a weight vector.

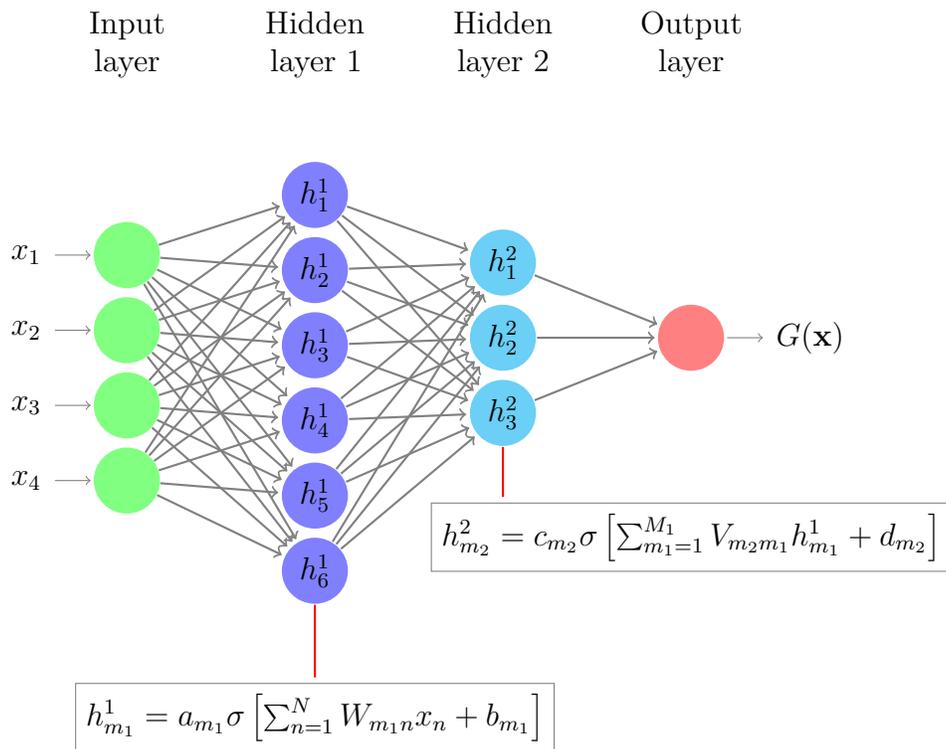


Figure 5.3: Depiction of the MLFFNN described by Eq. (5.13) with two hidden layers. The values of the hidden nodes are depicted in the boxes.

5.5 Feed-Forward Neural Network (FFNN) over the Complex Field

While the discussion so far was limited to real-valued continuous functions f , a FFNN can only be a good approximator for quantum physical wave functions, if it is extended to the field of complex numbers. Contrary to the field of real-valued FFNNs, which have a wide variety of real world applications and are thus well studied, the field of complex-valued FFNNs (or NNs in general) is rather new and the literature thus limited. Another real-world application is besides quantum physical computations given by electromagnetic waves in signal transmission [124]. In the complex case, it is more subtle to find a reasonable activation function. This is due to the fact that according to Liouville's theorem, every analytic (or entire) and bounded function $f(z)$ with $z \in \mathbb{C}$ must be constant $f(z) = c$ with $c \in \mathbb{C}$. An analytic function is one that is holomorphic and thus differentiable in the neighbourhood of every point in the complex plane. Therefore, one has to pick for the activation function to be either bounded and thus not holomorphic on a certain domain, that might be irrelevant, or analytic but unbounded. The property of being holomorphic at a certain point $z = x + iy$ restricts the derivatives of the function $f(z) = u(x, y) + iv(x, y)$ to follow the Cauchy-Riemann equations [131]

$$\partial_x u = \partial_y v \quad \text{and} \quad \partial_y u = -\partial_x v , \quad (5.16)$$

which is equivalently written as

$$\partial_{\bar{z}} f = 0 , \quad (5.17)$$

where $\bar{z} = x - iy$ and

$$\partial_z f = \frac{1}{2} (\partial_x f - i\partial_y f) , \quad \partial_{\bar{z}} f = \frac{1}{2} (\partial_x f + i\partial_y f) . \quad (5.18)$$

Thus, there is a wide range of activation functions proposed, ranging from real-valued split functions for either both the real and imaginary part or the absolute value and the phase of $f(z)$, respectively, to so-called fully complex activations functions that try to avoid singular points or deal with the issue of the parameters potentially increasing to magnitudes that can not be handled anymore [131, 132]. However, since this is a wide field of research by itself, it is not the aim of this work to investigate the expressive power of different architectures of complex-valued FFNNs in general. Conversely, the mathematical motivation presented so far is used to indicate why the following ideas for a complex-valued FFNN ansatz for the variational wave function in the VMC algorithm described in Sec. 4 is reasonable.

5.6 Neural Network Quantum State (NNQS)

The respective ansatz for the wave function $\Psi_{\mathbb{W}}(\mathbf{s})$ for complex parameters \mathbb{W} is the quantity of interest in the following. Whenever it is given by a FFNN, it is naturally called a neural network

quantum state (NNQS). It is stressed again that, unlike the wave functions discussed in Sec. 6, the following are motivated by the expressive power of FFNNs and not tailored explicitly towards a certain quantum many-body model. The following ansatz wave functions are in principle valid for any spin models, however, in this work, it is focused upon the models introduced in Ch. 2.

5.6.1 Sigmoidal Ansatz (SA)

The results discussed so far in this chapter are used in the following to motivate a variational ansatz wave function in the s^z -basis, as introduced in Ch. 4. The most evident and simple ansatz derived from the analysis in this chapter would be a SLFFNN as in Th. 2 for the absolute value and the phase of the wave function separately

$$\Psi_{\mathbb{W}}(\mathbf{s}) = \left[\frac{1}{M_1} \sum_{m=1}^{M_1} a_m \sigma \left(\sum_{n=1}^N W_{mn} s_n^z + b_m \right) \right] \cdot \exp \left[\frac{i\pi}{M_2} \sum_{m=1}^{M_2} \tilde{a}_m \tanh \left(\sum_{n=1}^N \tilde{W}_{mn} s_n^z + \tilde{b}_m \right) \right], \quad (5.19)$$

where the parameter arrays of the first SLFFNN are $\mathbf{a}, \mathbf{b} \in \mathbb{R}^{M_1}$, $\mathbf{W} \in \mathbb{R}^{M_1 \times N}$ and the ones for the second one are $\tilde{\mathbf{a}}, \tilde{\mathbf{b}} \in \mathbb{R}^{M_2}$, $\tilde{\mathbf{W}} \in \mathbb{R}^{M_2 \times N}$. One could leave out the parameters \mathbf{a} and $\tilde{\mathbf{a}}$ to allow the absolute value to be only in the range of $[0, 1]$ and the phase to be only in the range $[-i\pi, i\pi]$. The number of hidden units is often phrased as a multiple of the input units, which is denoted by $\alpha = \frac{M}{N}$. It is stressed that, contrary to the theorems provided in this section, the function to approximate is not defined on the unit cube but only has a binary input rendering the approximation even easier. A similar ansatz as in Eq. (5.19) was proposed in Ref. [133]. However, the first ansatz based on FFNNs was proposed by the authors in Ref. [11] and is based on the classical Boltzmann machine (CBM).

5.6.2 Boltzmann Machine (BM)

In the following, the motivation behind the ansatz wave function known by the name restricted Boltzmann machine (RBM) is shortly discussed. It has originated from its classical counterpart with real-valued parameters. Therefore, we take a step back to the classical Boltzmann machine (CBM) with binary input and real-valued parameters before it is extended to the complex space and used as a variational ansatz wave function.

Classical Boltzmann Machine (CBM)

The classical Boltzmann machine (CBM) is categorized within the realm of energy-based models, which, in turn, falls under the broader category of graphical or structured probabilistic models [113]. These are probabilistic models that describe a probability distribution using a graph, thereby connecting to the mathematical field of graph theory. The probability $p(\mathbf{v}, \mathbf{h}) \in [0, 1]$ and the partition function Z

$$p(\mathbf{v}, \mathbf{h}) = \frac{1}{Z} e^{-H(\mathbf{v}, \mathbf{h})} \quad Z = \sum_{\mathbf{v}, \mathbf{h}} e^{-H(\mathbf{v}, \mathbf{h})} \quad (5.20)$$

of an energy-based model are solely given by the energy

$$E(\mathbf{v}, \mathbf{h}) = \sum_n^N a_n v_n + \sum_m^M b_m h_m + \sum_{n,m} h_m W_{mn} v_n + \sum_{n_1 < n_2} v_{n_1} C_{n_1 n_2} v_{n_2} + \sum_{m_1 < m_2} h_{m_1} D_{m_1, m_2} h_{m_2} \quad (5.21)$$

of its state defined by its binary input or visual nodes $\mathbf{v} \in \{0, 1\}^N$, its binary hidden nodes $\mathbf{h} \in \{0, 1\}^M$ and respective parameter vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$ and weight matrices $\mathbf{W} \in \mathbb{R}^{M \times N}$, $\mathbf{C} \in \mathbb{R}^{N \times N}$, $\mathbf{D} \in \mathbb{R}^{M \times M}$ [113]. These are called classical Boltzmann machines (BMs), which stems from the fact that the probability distribution in Eq. (5.20) is given by the Boltzmann distribution. Without the hidden nodes, the interaction between the input nodes is solely given by the matrix \mathbf{C} and is thus linearly dependent on the other input nodes. Therefore, the benefit of introducing hidden nodes, whose state is not known and not of interest, becomes immediately clear. They allow for interaction terms of higher order and can even be complemented by more than one hidden layer [113]. The CBM is not a FFNN as the ones used so far but classified as a Hopfield network, which is an undirected network that is extensively discussed in Ref. [134]. For the restricted classical Boltzmann machine (RCBM), the intra-layer interactions are not allowed leading to the energy function being

$$E(\mathbf{v}, \mathbf{h}) = \sum_n^N a_n v_n + \sum_m^M b_m h_m + \sum_{n,m} h_m W_{mn} v_n. \quad (5.22)$$

This allows for the computation of the probability distribution of a realization of the visual nodes

$$p(\mathbf{v}) = \sum_{\mathbf{h}=\{-1,1\}} p(\mathbf{v}, \mathbf{h}) \propto e^{\sum_{n=1}^N a_n v_n} \prod_{m=1}^M \cosh \left(\sum_{n=1}^N W_{mn} v_n + b_m \right) \quad (5.23)$$

by summing over the hidden nodes. A notable difference of the ansatz, compared to the ones discussed in this section so far, is the usage of the product in the output node rather than the sum. A FFNN of this kind is labelled as a product of experts [135]. The RCBM was successfully used to reduce the dimensionality of data [136]. Its expressive power is for boolean functions comparable to the ones of SLFFNN with sigmoid functions as their activation functions and can be classified again as a FFNN for the input \mathbf{v} [137]. It is additionally proven to be a universal approximator for probability distributions on the discrete set of binary numbers $\{0, 1\}^N$ with $k+1$ hidden units, where k is the number of input vectors with non-zero probability [138]. A similar theorem is obtained for deep Boltzmann machines (DBM) in Ref. [139].

Translationally-Invariant Classical Restricted Boltzmann Machine (TICRBM)

The CRBM can be forced to adhere to certain symmetries generated by general transformation matrices $T_l : \mathbb{R}^N \rightarrow \mathbb{R}^N$ ($l = 1, \dots, L$). In the following, the translational invariance is considered with T_l being the translation operator shifting the spins by l sites. The energy-function for the

translationally-invariant CRBM (TICRBM) is then given by

$$E(\mathbf{v}, \mathbf{h}) = a \sum_{l=1}^N v_l + \sum_{m=1}^M \sum_{l=1}^N b_{lm} h_{lm} + \sum_{m=1}^M \sum_{l=1}^N \sum_{n=1}^N h_{lm} W_{mn} (T_l \mathbf{v})_n, \quad (5.24)$$

where $a \in \mathbb{R}$, $\mathbf{b} \in \mathbb{R}^{N \times M}$ and $\mathbf{W} \in \mathbb{R}^{M \times N}$ [140]. Since the action of the translation operator basically defines a convolution as explained in Eq. (5.15), the TICRBM is classified as a CNN [141]. The probability for a visual configuration is obtained by summing up all configurations of the hidden nodes [11]:

$$p(\mathbf{v}) \propto e^{a \sum_{l=1}^N v_l} \prod_{m=1}^M \prod_{l=1}^N \cosh \left(\sum_{n=1}^N W_{mn} (T_l \mathbf{v})_n + b_{m,l} \right) \quad (5.25)$$

Restricted Boltzmann Machine (RBM)

Building on the idea of the CRBM, the first ansatz that was proposed by the authors in Ref. [11] for spin models is the so-called restricted Boltzmann machine (RBM)

$$\Psi_{\mathbb{W}}(\mathbf{s}) = e^{\sum_{n=1}^N a_n s_n^z} \prod_{m=1}^M \underbrace{\cosh \left(\sum_{n=1}^N W_{mn} s_n^z + b_m \right)}_{\Theta_m}, \quad (5.26)$$

where the parameters $\mathbf{a} \in \mathbb{C}^N$, $\mathbf{b} \in \mathbb{C}^M$, $\mathbf{W} \in \mathbb{C}^{M \times N}$ are now complex-valued. The input is given by the s_n^z -values of the actual configuration $\mathbf{s} = \{s_1^z, \dots, s_N^z\}$. Thus, the input is also binary and the square of the wave function components also follow a probability distribution. The RBM is due to its significance depicted in Fig. 5.4. The connections between the input layer and the hidden layer can also be restricted to be local up to a certain range. A RBM of this kind is called a short-range RBM, in which the second dimension of the weight matrix W is effectively limited. The factors of the product in Eq. (5.26) are thus altered to

$$\Theta_{m|\underline{n}} = \cosh \left(\sum_{n \in \underline{n}} W_{mn} s_n^z + b_m \right), \quad (5.27)$$

where

$$\underline{n} = \{l : |l - n| < R\} \quad (5.28)$$

with $R > 0$ a local subregion [142]. For clarity of notation, the standard RBM is thus sometimes also referred to as the fully-connected RBM. In order for the expectation value of any operator to be correct, the wave function components of the state of interest itself need to be approximated correctly. The theoretical foundation of this ansatz is not strictly proven in the complex domain, however, motivated by the theorems in this chapter. Similar to other FFNN applications, it has been used and tested in several fields in the recent past. In Section 7, the state-of-the-art methods for the ground state optimization using FFNNs are discussed in detail and the reasoning for the RBM to be the FFNN of choice is explained. Finally, it is used in Ch. 8 and 10 for the computation of finite

temperature results in one and two dimensions, whereas the real-time evolution is investigated in Ch. 9.

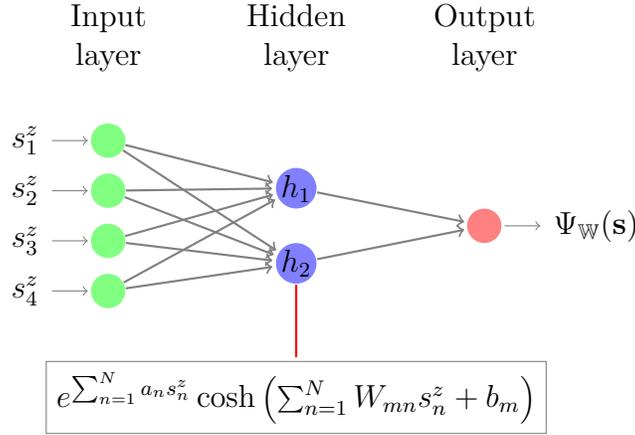


Figure 5.4: Depiction of the RBM described by Eq. (5.26) with the input given by the s^z -values of the configurations and the parameters being complex-valued.

5.7 Effective Update Scheme

The Monte Carlo sampling is of course a time consuming part of the VMC algorithm presented in Ch. 4, in which at every sampling step the wave function component of the proposed configuration needs to be computed. Thus, it is crucial that this can be done efficiently when using the RBM ansatz. This is achieved by storing the arguments of the wave function as

$$\Theta_m = \sum_{n=1}^N W_{mn} s_n^z + b_m \quad \text{and} \quad \zeta = \sum_{n=1}^N a_n s_n^z. \quad (5.29)$$

The new storage parameters after a spin flip in a configuration \mathbf{s} at site l are then directly calculated to be

$$\Theta'_m = \Theta_m - W_{ml} s_l^z \quad \text{and} \quad \zeta' = \zeta - a_l s_l^z. \quad (5.30)$$

The new values of the hidden nodes are subsequently obtained by applying the activation function in Eq. (5.26). By saving the arguments of the activation functions in Eq. (5.29), the logarithmic derivatives

$$\begin{aligned} O_{a_n} &= s_n \\ O_{b_m} &= \tanh(\Theta_m) \\ O_{W_{mn}} &= s_n O_{b_m} \end{aligned} \quad (5.31)$$

defined in Eq. (4.12) are effectively computed as well.

5.8 Details of the Learning Algorithm

The expressive power of SLFFNNs and MLFFNNs was established in this chapter. In the following, the details of general so-called learning algorithms in NN theory are discussed and the various fields of applications mentioned. Additionally, the specific learning algorithm, that is basically used throughout this work, is explained. It is emphasized that this is essentially just a reformulation of the VMC algorithm presented in Ch. 4, utilizing the classifications and terminologies commonly employed in the NN community. Thus, it comes with a short summary of important parts of NN theory. However, this does not claim to be to any extent complete and the reader is referred to follow Ref. [113] for more details.

The importance of NNs in general originates from their ability to be applicable to a wide variety of tasks T , for example classification, regression, translation or structuring an output. The accuracy is measured by a performance measure P [113]. In principle, NN learning algorithms are divided into supervised, unsupervised and reinforcement learning algorithms. The algorithms examined in this work are all classified as reinforcement learning algorithms, because it is a Markov decision process which can be viewed as an interplay between an *agent* and an *environment* [143, 11]. The *agent* is in our case comprised by the FFNN and the update algorithm, while the *environment* is comprised by the set of configurations obtained by the Monte Carlo sampling. The action taken by the *agent* to minimize the reward or loss function, for example the energy, changes the *environment*, in our case the configurations that are chosen by the Monte Carlo sampling. The reinforcement algorithm can be further classified as a regression or optimization algorithm in a high-dimensional space. To unite the notation in the NN framework with the one from the VMC framework presented in Sec. 4, the ground state optimization algorithm is paraphrased in the NN language used in Ref. [113].

In an NN algorithm, the performance measure P is given by the so-called loss function, also known as cost function or reward function. In our guiding example of the ground state optimization, this loss function is given by the energy in Eq.(4.4). The energy is minimized according to Eq. (4.14), or, as it is famously denoted in the NN community, the algorithm learns the representation of the ground state. While it seems in this case awkwardly exaggerating to rename the optimization algorithm into learning algorithm, the illusion of learning is in general applications hard to disassemble into its algorithmic parts. Additionally, the difference between optimization and learning is discussed in an own chapter in Ref. [113]. Therein, optimization is considered a direct optimization of the performance measure P , whereas learning specifies a task in which P can be intractable and is only optimized indirectly by actually optimizing the loss function.

To this point, the VMC algorithm was explained to be composed of the variational ansatz wave function and the Monte Carlo sampling algorithm, which was illustrated in blue in the flowchart in Fig. 4.1. While they are of course intertwined, as is immediately clear for example from Eq. (4.4), they are viewed in NN theory in a more unified way being two different features of the NN itself. In the previous sections of this chapter, the accuracy with which the target function can be approximated by the FFNN of choice was discussed in detail. This is referred to as the expressive power of the FFNN, or simply its expressibility [144]. On the other hand, the number of Monte Carlo samples necessary to approximate the target probability distribution is also referred to as the generalization property of this specific FFNN [145]. The generalization error, being the difference between the training and the test error, is thereby the quantity of interest. In NN theory applications, the

generalization property thus illustrates to what extent the optimized NN is applicable to data sets outside the training data. In the VMC algorithm of quantum many-body physics, it is thus related to the number of Monte Carlo samples and consequently, the number of configurations, or basis vectors of the Hilbert space, that need to be traversed in the Monte Carlo sampling. The smaller the generalization error, the smaller the number of samples needed to ensure that the target probability distribution is accurately approximated, even for those configurations or basis states that may not directly appear in the Monte Carlo sampling. These considerations will turn out to be of crucial importance in Ch. 10. Therein it will be obvious, that the generalization feature is not a distinct feature of the FFNN at hand, but also the data at hand, given in our study by the configurations or basis states of the Hilbert space. Thus, the notion with respect to the generalization feature is not adopted but the distinction between peculiarities of the variational ansatz and the Monte Carlo sampling are distinguished whenever possible.

While the optimization with respect to a SLFFNN is evident and numerically effective, since it consists only of derivatives of the one activation function with respect to the parameters, the optimization of the MLFFNN is more advanced and numerically more elaborate. The respective and effective algorithm is known as the backpropagation algorithm [146, 113].

5.9 Curse of Dimensionality

The relevance of the mathematical theorems provided in Sec. 5.3 and 5.4 for actual computations depends heavily on how good the approximations are for a manageable number of parameters. In numerical calculations, this coincides of course with the computational resources available. All in all, there is unfortunately no clear theorem that can provide a decisive connection between the approximation accuracy of a generic FFNN and its network architecture [147]. Still, there are some theorems for various FFNN architectures limiting the width for arbitrary depth [148] and relating depth and width for some concrete examples [149]. In Ref. [123], there is a limit for the number of hidden nodes for a MLFFNN, as described by Eq. (5.13) and depicted in Fig. 5.3, computed. It depends on the modulus of continuity of the desired function f .

Naturally, the problem becomes inherently difficult if the number of input data is large. This is referred to as the curse of dimensionality [113]. However, in the application to quantum physics, the sheer size of the Hilbert space growing exponentially with system size 2^N is misleading, since the actual physical space for local Hamiltonians, that is comprised by the states that are physically relevant, is definitely smaller. This is quantified in Ref. [150] by computing all those states that can be reached by the real-time evolution of a product state with a local Hamiltonian, leading to the size of the physical space to be exponentially small compared to the volume of the Hilbert space. Another result supporting this notion is the area-law for one-dimensional gapped Hamiltonians, which states that the entanglement entropy of their ground states is bounded by a term that is proportional to the boundary area [151, 152]. While the accuracy of the universal approximators presented increases with the number of parameters, the number of dimensions in the parameter space grows at the same time. This can lead in general to a more subtle optimization prone to issues for example due to the regularization discussed in Sec. 4.5. However, one can certainly state that, in principle, the number of hidden units necessary in a FFNN to approximate a variety of states

increases with the entanglement entropy those states possess. This is numerically verified in Sec. 8.3 and Ch. 9.

All in all, in most applications, there is no clear indication that the number of hidden nodes in a FFNN architecture and thus the number of parameters is numerically feasible. The huge success of NN in general has lead to a very practical *modus operandi*.

5.10 Entanglement Entropy and Expressive Power of RBMs

The bipartite entanglement entropy

$$S(\rho_A) = -\text{Tr}[\rho_A \log \rho_A] = -\text{Tr}[\rho_B \log \rho_B] = S(\rho_B) \quad (5.32)$$

of a pure state, given by its density matrix ρ , of a system divided into parts A and B , quantifies the quantum entanglement between these two parts. Here, the reduced density matrices $\rho_A = \text{Tr}_B[\rho]$ and $\rho_B = \text{Tr}_A[\rho]$ are computed by tracing out the respective complementary part of the system. For tensor networks (TN) in general and matrix product states (MPS) in specific, the entanglement entropy is inherently connected to the bond dimension D , which itself defines the accuracy of the algorithm and thus also the computational effort [153, 154]. The method, which allows for the optimization of MPS states, the density-matrix renormalization group (DMRG), can also be interpreted as a variational method [8]. It has been very successful for the ground state optimization of models which adhere to the area-law for ground states [151].

In principle, all variational wave functions are built up from local correlations. For the presented NNQSs in Sec. 5.6, namely the SA in Eq. (5.19) and the RBM in Eq. (5.26), this is evident as well as for the ones that will be presented later in Ch. 6. However, for the RBM, which is capable of efficiently representing volume-law entangled states, the link between the number of variational parameters and the entanglement entropy is not that simple [142]. From the discussion in this chapter, one can already conclude that, in principle, fully-connected NNQS should even be able to approximate states with a large entanglement entropy effectively. It is stressed, that this is the more physical reasoning for the enhanced interest in using FFNNs for the approximation of wave functions. The RBM ansatz, originally proposed in Ref. [11], was proven to efficiently allow to describe volume-law entangled states contrary to the MPS ansatz [9]. While the word efficiently often remains vague, it is concretized in Ref. [142] to denote a linear increase in the number of non-zero parameters. A comprehensive examination of the entanglement entropy of RBMs, along with an in-depth analysis of the entanglement entropy of short-range RBMs, is thoroughly explored in Ref. [142]. The fully-connected RBM is compared to a variety of known TN states in Ref. [9] and is classified as being a non-local string-bond state, meaning that in the RBM there are many strings that overlap on the entire lattice. A detailed work on the expressive power of (deep) RBMs is found in Ref. [155], while Ref. [156] provides an algorithm to transform the RBM to a TN and states the conditions under which the inverse transformation can be achieved. In this manner, the authors also introduce the idea of inversely being able to detect a limit for the number of variational parameters of the RBM by the link to the bond dimension of the TN the RBM is transformed into. It is recalled that a similar direct theorem is lacking in the NN theory as described in Sec. 5.9.

It should be noted, that the second important advantage of the RBM over the DMRG method is its simple generalization to more than one dimension. This does of course not mean that there are no approaches to modify the DMRG method for two-dimensional systems [157]. Another interesting feature about the entanglement entropy of random RBM states is illustrated in Ref. [142] and will be of importance in the quantum typicality approach investigated in Sec. 8.3. It shows that the entanglement entropy of a random RBM ansatz state still obeys volume-law entanglement, however, it is noticeably lower than the one for random pure states, which is also known as the Page entropy.

Chapter 6

Variational Wave Functions for Strongly Correlated Systems

Historically, the first ideas for a variational ansatz wave function were based on physical observations regarding small systems. Those that are, to some extent, used later in this work, are described in the following. The variational wave function, expressed in the s^z -basis in Eq. (4.2), can be formulated in any chosen basis. In the following, it will be the one given by the fermionic occupation numbers $n_{i\sigma} = \{0, 1\}$, $\sigma \in \{\uparrow, \downarrow\}$. The wave function of a system of N_e fermions with $N_{e,\uparrow}$ fermions with spin up and $N_{e,\downarrow}$ with spin down is given by

$$|\Psi\rangle = \sum_{\mathbf{n}_\uparrow, \mathbf{n}_\downarrow} \mathbf{F}_{\mathbf{n}_\uparrow, \mathbf{n}_\downarrow} |\mathbf{n}_\uparrow\rangle |\mathbf{n}_\downarrow\rangle, \quad (6.1)$$

where $\mathbf{F}_{\mathbf{n}_\uparrow, \mathbf{n}_\downarrow} \in \mathbb{C}^{2^N \times 2^N}$ and $\mathbf{n}_\sigma = (n_{1\sigma}, \dots, n_{N\sigma})$. If the overall particle number N_e is limited to be equal to the system size N and double occupation is prohibited, the number of parameters in $\mathbf{F}_{\mathbf{n}_\uparrow, \mathbf{n}_\downarrow}$ is reduced to 2^N . The objective is here, as well as it was to this point for the s^z -basis, to approximate $\mathbf{F}_{\mathbf{n}_\uparrow, \mathbf{n}_\downarrow}$ by a variational wave function. The ideas proposed in this regard can be divided into two groups that go by the name of single-particle orbitals and paired orbitals. The general difference is best introduced by looking at the treatment of the H_2 -molecule as described in Ref. [3].

6.1 H_2 -Molecule

The first ansatz for a quantum mechanical wave function of the H_2 -molecule was proposed by W. Heitler and F. London for strongly correlated electrons [158]. Later, R. Hartree and W.A. Fock introduced an ansatz that is suited for independent electrons [159].

6.1.1 Heitler-London Ansatz

In the Heitler-London ansatz, there is no contribution by electrons on the same atom and the original version is written in second quantization as

$$|\Psi_{HL}\rangle = \frac{1}{\sqrt{2}}(c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{1\downarrow}^\dagger) |0\rangle, \quad (6.2)$$

where $|0\rangle = |\mathbf{0}_\uparrow\rangle|\mathbf{0}_\downarrow\rangle$ is the vacuum state and

$$c_{i\sigma}^\dagger = \int d\mathbf{r} \phi(\mathbf{r} - \mathbf{R}_i) \Psi_\sigma^\dagger(\mathbf{r}) \quad c_{i\sigma} = \int d\mathbf{r} \phi^\dagger(\mathbf{r} - \mathbf{R}_i) \Psi_\sigma(\mathbf{r}) \quad (6.3)$$

the fermionic operator that creates (annihilates) an electron with spin σ in an orbital $\phi(\mathbf{r} - \mathbf{R}_i)$ that is centered around the atom $i = 1, 2$ at position \mathbf{R}_i and $\Psi_\sigma^\dagger(\mathbf{r})$ ($\Psi_\sigma(\mathbf{r})$) the fermionic field operator. The anti-commutation relations for the fermionic creation and annihilation operators follow from the orthogonality condition for the orbitals of different atoms. The wave function becomes accurate in the atomic limit $|\mathbf{R}_1 - \mathbf{R}_2| \rightarrow \infty$.

6.1.2 Hartree-Fock Ansatz

Another ansatz proposed in the early days of quantum mechanics is the one by R. Hartree and W.A. Fock [159]. It considers the electrons to be independent of each other and can also be written as a Slater determinant [160]

$$|\Psi_{HF}\rangle = \Phi_\uparrow^\dagger \Phi_\downarrow^\dagger |0\rangle. \quad (6.4)$$

The operator Φ_σ^\dagger creates single particle orbitals. The simplest realization for a two-particle system like the H_2 -molecule

$$\Phi_\sigma^\dagger = \frac{1}{\sqrt{2}}(c_{1\sigma}^\dagger + c_{2\sigma}^\dagger) \quad (6.5)$$

leads to the Hartree-Fock wave function

$$|\Psi_{HF}\rangle = \frac{1}{2}(c_{1\uparrow}^\dagger c_{1\downarrow}^\dagger + c_{1\uparrow}^\dagger c_{2\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{1\downarrow}^\dagger + c_{2\uparrow}^\dagger c_{2\downarrow}^\dagger) |0\rangle. \quad (6.6)$$

In contrast to the Heitler-London wave function, the Hartree-Fock approximation becomes exact whenever the electron-electron repulsion is neglected. Therefore, the correct solution lies in between the Hartree-Fock and the Heitler-London ansatz [3].

6.2 Many-Body Ansatz Wave Function

Starting from these first considerations regarding the H_2 -molecule, there were a number of variational fermionic wave functions proposed for a range of models. In the following, brief discussions will be

provided for those variational wave functions that are, to some extent, related to this work or serve as a comparison for the NNQS proposed in Ch. 5.

6.2.1 Hartree-Fock Ansatz

The Hartree-Fock ansatz is extended to many-body fermionic models by defining the wave function ansatz

$$|\Psi_{HF}\rangle = \prod_{\alpha=1}^{N_e} \Phi_{\alpha}^{\dagger} |0\rangle \quad (6.7)$$

with single-particle orbitals

$$\Phi_{\alpha}^{\dagger} = \sum_i^N W_{\alpha i \uparrow} c_{i \uparrow}^{\dagger} + \sum_i^N W_{\alpha i \downarrow} c_{i \downarrow}^{\dagger}, \quad (6.8)$$

that are chosen to be some parametrization of the fermionic creation operators with $W_{\alpha j \sigma} \in \mathbb{C}$ being complex coefficients [3]. The orthogonality condition is in this case given by the following equation

$$\sum_i^N \left(W_{\alpha i \uparrow} W_{\beta i \uparrow}^* + W_{\alpha i \downarrow} W_{\beta i \downarrow}^* \right) = \delta_{\alpha \beta}. \quad (6.9)$$

The wave function in equation (6.7) can be expressed as a Slater determinant that respects fermionic commutation relations [160]. However, using the Hartree-Fock ansatz, only reasonable results are obtained in the weak interaction regime [3].

6.2.2 Gutzwiller and Jastrow Factor

The Gutzwiller factor

$$P_G(g) = e^{-g \sum_{j=1}^N \hat{n}_{j \uparrow} \hat{n}_{j \downarrow}} \quad (6.10)$$

was introduced in Ref. [161] to attach a weight $g \in \mathbb{R}$ depending on the number of double occupied sites. The weight g can be used as a variational parameter. On the other hand, for spin models, it is used to prohibit double occupancy by choosing $g = \infty$ and thus also referred to as a projector in this case. The notation is adjusted to distinguish the occupation number operator \hat{n}_i from its expectation value n_i . For the H_2 -molecule the Gutzwiller factor allows to interpolate between the Heitler-London wave function in Eq. (6.2) with $g = \infty$ and the Hartree-Fock one in Eq. (6.6) with $g = 0$ [3].

Furthermore, adding the Jastrow factor

$$P_J(\mathbf{v}) = e^{-\frac{1}{2} \sum_{i,j=1}^N v_{i,j} (\hat{n}_i - n)(\hat{n}_j - n)} \quad (6.11)$$

which was introduced in Ref. [162], allows for density fluctuations the Gutzwiller factor can not grasp [3, 163]. The density is given by $n_i = n_{j \uparrow} + n_{j \downarrow}$ and the density average by n , whereas $v_{i,j} \in \mathbb{R}$ are variational parameters. The Gutzwiller and Jastrow factor are the two most basic ingredients

that are usually combined with other variational wave functions by multiplying them all together [163].

6.3 Pair-Product (PP) Wave Function

A particularly important ansatz is known as the pair-product (PP) or Pfaffian wave function and is closely related to the theory of superconductivity. This link is explained first before the difference to the more general PP wave function is outlined. In the end, the overlap with a real space configuration is computed and the Pfaffian obtained.

6.3.1 Bardeen-Cooper-Schrieffer (BCS) Wave Function

The Bardeen-Cooper-Schrieffer (BCS) wave function

$$|\Psi_{BCS}\rangle = \exp\left(\sum_k f_k c_{k\uparrow}^\dagger c_{-k\downarrow}^\dagger\right) |0\rangle, \quad (6.12)$$

is the ground state of the respective translationally-invariant BCS Hamiltonian in mean-field approximation

$$H_{BCS} = \sum_{k,\sigma} (\epsilon_k - \mu_0) c_{k,\sigma}^\dagger c_{k,\sigma} + \sum_k \Delta_k c_{k,\uparrow}^\dagger c_{-k,\downarrow}^\dagger + h.c., \quad (6.13)$$

where f_k is the pairing function, ϵ_k the dispersion, which gives the band structure, μ_0 the chemical potential and $\Delta_k = \Delta_{-k}$ a singlet pairing amplitude [3]. This is proven in detail in Ref. [164]. The BCS theory was originally introduced in Ref. [165] to describe the effective attractive interaction between electrons with opposite spin and momenta $(\mathbf{k}, -\mathbf{k})$ mediated by virtual phonons. These bound states are known as Cooper pairs and explain the effect of low-temperature superconductivity in several solids [17]. The real space BCS wave function is obtained by the Fourier transform of Eq. (6.12) and expanding the exponential function

$$|\Psi_{BCS}\rangle = \exp\left(\sum_{i,j=1}^N f_{ij} c_{i\uparrow}^\dagger c_{j\downarrow}^\dagger\right) |0\rangle = \frac{1}{\left(\frac{N_e}{2}\right)!} \left(\sum_{i,j=1}^N f_{ij} c_{i\uparrow}^\dagger c_{j\downarrow}^\dagger\right)^{\frac{N_e}{2}} |0\rangle, \quad (6.14)$$

where $f_{ij} = \sum_k \frac{1}{\sqrt{N}} f_k e^{-ik(i-j)}$ [166]. The parameters f_{ij} are well-suited to be the starting point of the variational method discussed in Ch. 4, especially in the following modified ansatz.

6.3.2 Resonating Valence Bond (RVB) Wave Function

The deficiency of the BCS theory to describe high-temperature superconductivity was addressed by P.W. Anderson in Ref. [167] by an alternative ansatz for the wave function

$$|\Psi_{RVB}\rangle = P_G(\infty) \frac{1}{\left(\frac{N_e}{2}\right)!} \left(\sum_{i,j=1}^N f_{ij} c_{i\uparrow}^\dagger c_{j\downarrow}^\dagger \right)^{\frac{N_e}{2}} |0\rangle, \quad (6.15)$$

known as the resonating valence bond (RVB) ansatz. The Gutzwiller projector $P_G(\infty)$ removes all states, in which doubly occupied sites are present. This is why it was first used to obtain the ground state of spin chains like the Heisenberg model [167, 168, 3]. The name RVB originates from earlier works by L. Pauling [169] and P.W. Anderson [170]. The RVB describes an overlap of all singlet pairs possible and is depicted in Fig. 6.1.

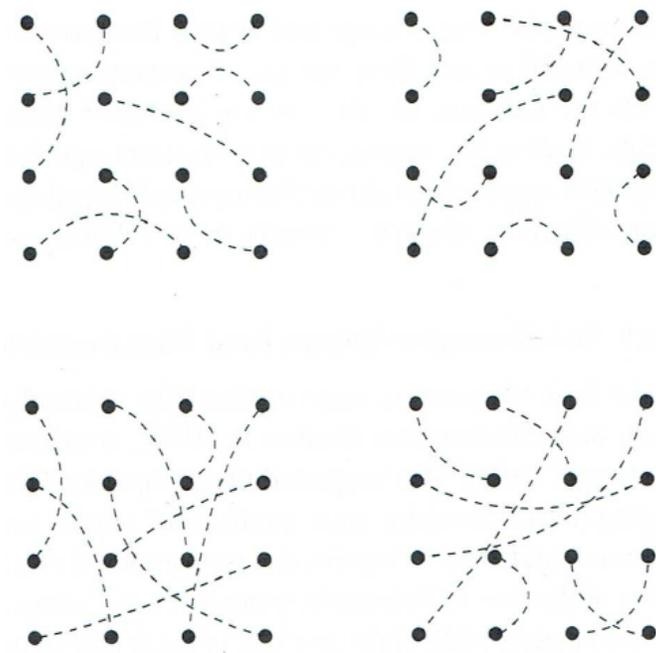


Figure 6.1: Illustration of four singlet pair patterns of the RVB wave function ansatz in two dimensions. Reprinted with permission from Ref. [3].

6.3.3 Pair-Product (PP) Wave Function

The BCS and the RVB wave function can be generalized even further by allowing any pair or bond, also between fermions whose spin points into the same direction. This results in a wave function that is known as pair-product (PP) wave function [163]. It is sometimes also referred to by Pfaffian or simply pairing wave function [171, 10]. At the end of this section, it will become clear where the term Pfaffian stems from. However, from now on, the shortcut PP for pair-product is used. We

will adhere to the derivation of the overlap of this wave function with a real space state with N_e particles $|\mathbf{n}_\uparrow\rangle|\mathbf{n}_\downarrow\rangle = c_{r_1\sigma_1}^\dagger c_{r_2\sigma_2}^\dagger \cdots c_{r_{N_e}\sigma_{N_e}}^\dagger |0\rangle$ provided in Ref. [10]. The PP wave function is stated in the general case

$$|\Psi_{PP}\rangle = \left(\sum_{i,j=1}^N \sum_{\sigma,\sigma'=\uparrow,\downarrow} F_{ij}^{\sigma,\sigma'} c_{i\sigma}^\dagger c_{j\sigma'}^\dagger \right)^{\frac{N_e}{2}} |0\rangle \quad (6.16)$$

and the anti parallel case

$$|\Psi_{AP-PP}\rangle = \left(\sum_{i,j=1} f_{ij} c_{i\uparrow}^\dagger c_{j\downarrow}^\dagger \right)^{\frac{N_e}{2}} |0\rangle, \quad (6.17)$$

in which $s^z = \sum_i s_i^z = 0$ holds [163]. First, we want to look at the general PP wave function and rewrite the sum in terms of permutations \mathcal{P} of the N_e indices

$$|\Psi_{PP}\rangle = \sum_{\mathcal{P}} \left(\frac{N_e}{2} \right)! \prod_{l=1}^{N_e/2} \left(F_{r_{\mathcal{P}(2l-1)}r_{\mathcal{P}(2l)}}^{\sigma_{\mathcal{P}(2l-1)}\sigma_{\mathcal{P}(2l)}} - F_{r_{\mathcal{P}(2l)}r_{\mathcal{P}(2l-1)}}^{\sigma_{\mathcal{P}(2l)}\sigma_{\mathcal{P}(2l-1)}} \right) \prod_{l=1}^{N_e/2} \left(c_{r_{\mathcal{P}(2l-1)}\sigma_{\mathcal{P}(2l-1)}}^\dagger c_{r_{\mathcal{P}(2l)}\sigma_{\mathcal{P}(2l)}}^\dagger \right) |0\rangle, \quad (6.18)$$

where the permutations fulfill the following two conditions for the creation operators to be in ascending order with respect to the lattice sites [10]:

$$\mathcal{P}(2l-1) < \mathcal{P}(2l) \quad \text{and} \quad \mathcal{P}(2l-1) < \mathcal{P}(2l+1) \quad (6.19)$$

Since in numerical calculations the projection onto a real space state $|\mathbf{n}_\uparrow\rangle|\mathbf{n}_\downarrow\rangle$ is of importance, it is considered in the following. Due to the circumstance that only the creation operators act non-trivially and the fermionic anti-commutation relations apply, the second part is heavily simplified

$$\langle \mathbf{n}_\uparrow | \langle \mathbf{n}_\downarrow | \prod_{l=1}^{N_e/2} \left(c_{r_{\mathcal{P}(2l-1)}\sigma_{\mathcal{P}(2l-1)}}^\dagger c_{r_{\mathcal{P}(2l)}\sigma_{\mathcal{P}(2l)}}^\dagger \right) |0\rangle = (-1)^{\mathcal{P}}, \quad (6.20)$$

where $(-1)^{\mathcal{P}}$ is the parity of \mathcal{P} [10]. The remaining formula coincides with the definition of the PP wave function and is recognized to be a Pfaffian [172]

$$\langle \mathbf{n}_\uparrow | \langle \mathbf{n}_\downarrow | \Psi_{PP}\rangle = \left(\frac{N_e}{2} \right)! \sum_{\mathcal{P}} (-1)^{\mathcal{P}} \prod_{l=1}^{N_e/2} \left(F_{r_{\mathcal{P}(2l-1)}r_{\mathcal{P}(2l)}}^{\sigma_{\mathcal{P}(2l-1)}\sigma_{\mathcal{P}(2l)}} - F_{r_{\mathcal{P}(2l)}r_{\mathcal{P}(2l-1)}}^{\sigma_{\mathcal{P}(2l)}\sigma_{\mathcal{P}(2l-1)}} \right) = \left(\frac{N_e}{2} \right)! \text{Pf}[\mathbf{X}] \quad (6.21)$$

of the skew-symmetric matrix

$$X_{ij} = F_{r_i r_j}^{\sigma_i \sigma_j} - F_{r_j r_i}^{\sigma_j \sigma_i} \quad (6.22)$$

of size $N_e \times N_e$. The particles are labelled by i and j and the diagonal is thus zero. The Pfaffian itself is defined as:

$$\text{Pf}[\mathbf{X}] = \frac{1}{2^{\frac{N_e}{2}} \left(\frac{N_e!}{2} \right)} \sum_{\sigma \in S_{N_e}} \prod_{i=1}^{N_e/2} X_{\sigma_{2i-1} \sigma_{2i}} \quad (6.23)$$

The set of variational parameters of this specific ansatz is reduced to be comprised of four parameter arrays $\mathbf{X}^{\uparrow\uparrow}$, $\mathbf{X}^{\uparrow\downarrow}$, $\mathbf{X}^{\downarrow\uparrow}$ and $\mathbf{X}^{\downarrow\downarrow}$ with each containing $(N_e^2 - N_e)/2$ complex parameters. These are

combined to a large variational parameter array denoted as \mathbb{X} . The two-particle PP wave function is a direct generalization of the one-particle Hartree-Fock ansatz (see equations (6.7) and (6.8)), which means that the latter is obtained by a specific choice of the variational parameters of the former [163].

$$F_{r_i r_j}^{\sigma_i \sigma_j} = \sum_{\alpha}^{N_e/2} \left(W_{(2\alpha-1)i\sigma_i} W_{2\alpha j \sigma_j} - W_{(2\alpha-1)j\sigma_j} W_{2\alpha i \sigma_i} \right) \quad (6.24)$$

$$f_{ij} = \sum_{\alpha}^{N_e/2} W_{\alpha i \uparrow} W_{\alpha j \downarrow} \quad (6.25)$$

For the application onto the spin models considered in this work, the number of fermions equals the number of sites $N_e = N$, which can also be denoted by adding the Gutzwiller projector $P_G(\infty)$ to the definition of the PP wave function in Eq. (6.16) or (6.17).

6.3.4 Efficient Computation of the Pfaffian and its Derivative

Employing the PP wave function in the VMC algorithm in Ch. 4 requires the computation of the Pfaffian in every Monte Carlo step. Thus, it is crucial to compute it in an efficient way. In general, the computational cost of the Pfaffian of a $N \times N$ matrix is of order $\mathcal{O}(N^3)$ [10]. Luckily, one can make use of the fact that in the Monte Carlo sampling, the updated matrix $\tilde{\mathbf{X}}$ differs from the one in the previous step (\mathbf{X}) only by one row and one column. Therefore, Cayley's identity [173]

$$\det \begin{pmatrix} 0 & X_{12} & \dots & X_{1N} \\ b_{12} & 0 & \dots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ b_{1M} & -X_{2N} & \dots & 0 \end{pmatrix} = Pf \begin{pmatrix} 0 & X_{12} & \dots & X_{1N} \\ -X_{12} & 0 & \dots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -X_{1N} & -X_{2N} & \dots & 0 \end{pmatrix} \cdot Pf \begin{pmatrix} 0 & b_{12} & \dots & b_{1N} \\ -b_{12} & 0 & \dots & X_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ -b_{1N} & -X_{2N} & \dots & 0 \end{pmatrix} \quad (6.26)$$

leads to the update equation

$$Pf[\tilde{\mathbf{X}}] = \frac{\det[\tilde{\mathbf{X}}]}{Pf[\mathbf{X}]} = Pf[\mathbf{X}] \sum_{m=1}^N X_{\alpha m}^{-1} b_m, \quad (6.27)$$

where the general identity $Pf[\mathbf{X}]^2 = \det[\mathbf{X}]$ was used. This allows to compute the Pfaffian of the updated matrix $\tilde{\mathbf{X}}$ by using the matrix \mathbf{X} of the previous step. The two matrices differ only by the column and row indexed by α . The new row α is given by \mathbf{b} [10]. Through this approach, the computational cost reduces to $\mathcal{O}(N)$ since the inverse of the matrix \mathbf{X} needs to be computed for the Pfaffian derivative anyhow and thus contributes no additional cost. The computation of the inverse itself is executed efficiently ($\mathcal{O}(N^2)$) as well using the following update rule [10]:

$$\tilde{X}_{ij}^{-1} = X_{ij}^{-1} + \frac{1}{\sum_m X_{\alpha m}^{-1} b_m} \left[- \left(\sum_m X_{im}^{-1} b_m \right) X_{\alpha j}^{-1} + \left(\sum_m X_{jm}^{-1} b_m \right) X_{\alpha i}^{-1} + \delta_{i\alpha} X_{\alpha j}^{-1} - \delta_{j\alpha} X_{\alpha i}^{-1} \right] \quad (6.28)$$

To update the variational parameters, the logarithmic derivative is computed via

$$\frac{1}{Pf[\mathbf{X}_{\mathbb{X}}]} \partial_{\mathbb{X}_k} Pf[\mathbf{X}_{\mathbb{X}}] = \text{Tr} \left[\mathbf{X}_{\mathbb{X}}^{-1} \partial_{\mathbb{X}_k} \mathbf{X}_{\mathbb{X}} \right] , \quad (6.29)$$

where the parameter dependency of the matrix \mathbf{X} in Eq. (6.22) on the set of complex parameters \mathbb{X} is explicitly denoted [163].

6.4 Transformation from Fermions to Spins

Since we are in this work interested in spin models and the NNQSs in Ch. 5 have been derived in the s^z -basis, it is of crucial importance to rewrite the fermionic wave functions in this section in the s^z -basis. The fermionic creation and annihilation operators $c_{i\sigma}^\dagger$ and $c_{i\sigma}$ are transformed to the spin operators by the following transformation rule

$$\mathbf{S}_i = \frac{1}{2} c_{i\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{i\beta} , \quad (6.30)$$

where the matrix of spin and Pauli operators was defined in Eq. (2.8). It is written in components to enhance clarity [174, 44]:

$$\begin{aligned} S_i^x &= \frac{1}{2} (c_{i\uparrow}^\dagger c_{i\downarrow} + c_{i\downarrow}^\dagger c_{i\uparrow}) \\ S_i^y &= \frac{i}{2} (-c_{i\uparrow}^\dagger c_{i\downarrow} + c_{i\downarrow}^\dagger c_{i\uparrow}) \\ S_i^z &= \frac{1}{2} (c_{i\uparrow}^\dagger c_{i\uparrow} - c_{i\downarrow}^\dagger c_{i\downarrow}) = \frac{1}{2} (n_{i\uparrow} - n_{i\downarrow}) \end{aligned} \quad (6.31)$$

Thus, for every s^z -basis state, there are several fermionic configurations possible. However, since double occupancy is prohibited in spin models ($P_G(\infty)$), there is a one-to-one correspondence between the spin and fermionic formulation that can be denoted by

$$|s_1^z \dots s_N^z\rangle \leftrightarrow |\mathbf{n}_\uparrow\rangle |\mathbf{n}_\downarrow\rangle = \prod_{i=1}^N c_{i\sigma_i}^\dagger |\mathbf{0}_\uparrow\rangle |\mathbf{0}_\downarrow\rangle , \quad (6.32)$$

where the left and right hand side are linked by $s_i^z = \frac{\sigma_i}{2}$ [5]. Hence, the pairing between the electrons can be understood as well by a pairing between the spins in the spin model.

6.5 PP-NN Ansatz

It is in general always possible to combine those in this section presented variational wave functions among each other. However, this is of course limited by the number of variational parameters that can be efficiently updated in the VMC method on the computational device available. In this work,

it is for several reasons, debated in Ch. 7 and Sec. 8.3, of interest to combine the RBM ansatz in Eq. (5.26) (or the SA ansatz in Eq. (5.19)) with the one in Eq. (6.21). The latter one is written in the s^z -basis using the identification in Eq. (6.32). Thus, the combined ansatz is implemented as

$$|\Psi_{\mathbb{w},\mathbb{x}}\rangle = \sum_{\mathbf{s}} \Psi_{\mathbb{w}}(\mathbf{s})\Phi_{\mathbb{x}}(\mathbf{s})|\mathbf{s}\rangle , \quad (6.33)$$

where $\Phi_{\mathbb{x}}(\mathbf{s}) = \langle \mathbf{s} | \Psi_{PP} \rangle$ is given by the Pfaffian (see Eq. (6.21) and Eq. (6.32)) [95]. Since the RBM adheres to the universal approximation theorem of FFNNs (Th. 2) and the RBM is viewed as just a representative of this class (see discussion in Ch. 7), this combined ansatz is called PP-NN ansatz.

Chapter 7

Ground State Optimization with Feed-Forward Neural Networks (FFNNs)

The motivation to use a FFNN ansatz for the wave function in the VMC method has been outlined in Ch. 5. The ground state optimization is the most straightforward application. The finite-temperature computations, that are the topic of Ch. 8, can only be applied to models for which the ground state and its energy are actually obtained accurately in the ground state optimization. Otherwise, low temperatures can not be reached in finite-temperature computations. Loosely speaking, the ground state optimization corresponds to a finite-temperature computation, in which the initial state and the specific route taken are not important as long as the correct ground state is reached. Hence, the requirements for the accuracy of the evolution of the algorithm are notably reduced in the context of ground state optimization, a characteristic that is evident in the regularization applied, as explained in Sec. 4.5, for instance. Nevertheless, since the VMC approach is a high-dimensional optimization algorithm, in principle, there is always the possibility that the algorithm traverses an unfavourable landscape and gets stuck. Another principal issue possible is that the ground state might not be accurately approximated in the variational ansatz. The latter issue can, however, simply be mitigated in any FFNN ansatz by increasing the number of parameters which corresponds to enhancing the width or depth of the network.

In general, the RBM has successfully been applied to various quantum many-body models [175] ranging from spin models, fermionic [176] and bosonic models [177] to specific applications like learning topological states [178] and phases of matter [179]. The RBM is able to represent those states efficiently. For the spin models of interest in this work, the state of the art, regarding the ground state optimization, is outlined in this chapter. It is also intended to give a short review about the progress in this field, the main obstacles and our findings. Moreover, this approach helps elucidate the rationale behind opting for the RBM as the principal ansatz in conducting the finite-temperature computations in this work.

7.1 Barren Plateaus

In the seminal work introducing complex-valued RBMs [11], this ansatz was used to obtain the ground state energy for the one- and two-dimensional Heisenberg model. The algorithm used explicitly the translational invariance and thus the TIRBM ansatz, that was mentioned in Eq. (5.25). Unfortunately, it became clear that this could only be achieved after making use of the canonical transformation of the x - and y - components of the spin operators of one of the two sublattices, that was introduced in Sec. 2.5. This leads to the wave function components of the ground state being all real and positive. Hence, the TIRBM failed to learn the complicated sign structure the ground state of the Hamiltonian possesses if this transformation is not applied. This is not a failure due to the Monte Carlo sampling, but also occurs if the full basis of the subspace is used [133]. For the Heisenberg model, it is always advisable to restrict the states in the ground state optimization to the subspace with $s^z = \sum_i s_i^z = 0$, which is in the Monte Carlo algorithm accomplished by starting with a random state in that subspace and suggesting new states by exchanging two spins. Unfortunately, there could not be found a way out by altering the TIRBM slightly or using the SA ansatz in Sec. 5.6.1. A similar result was obtained in Ref. [133] for the J_1 - J_2 model, where the expressivity of the TIRBM and other MLFFNNs could be ruled out as a possible reason as well. It was instead attributed to what the authors termed the *rugged variational manifold landscape* and is also known as a *barren plateau* [180], which the algorithm can not escape easily. This issue is well-known and usually always a possible pitfall of any high-dimensional optimization algorithm that is hard to circumvent. The term pitfall can thus be understood in a literal way. Nevertheless, it is emphasized that deep architectures have been discovered capable of discerning the correct sign structure without any prior implementation. Unfortunately, this comes with the disadvantage that a huge number of 6 hidden layers only for the FFNN of the amplitude must be considered [181, 182]. The concrete shape, as well as the different activation functions used, look extremely arbitrary and do not follow from a transparent reasoning. Furthermore, the intended computational advantage of saving variational parameters through symmetry utilization is counteracted by the more intricate architecture and consequently a more time-consuming parameter update algorithm.

Fortunately, overcoming this failure is at last also possible for SLFFNNs by either refraining from applying translational symmetry altogether or by applying translational symmetry in an alternative form using the sum instead of the product of the shifted configurations in Eq. (5.25) [183]. Incorporating additional features and symmetries of the ground state, such as momentum or spin parity, can contribute to a further improvement of the accuracy of the ground state energy [184]. Restricting the variational wave function this way is known as the quantum-number projection method [185].

7.2 Improvement by the PP-NN Ansatz

The ground state optimization making use of the RBM ansatz was applied to the two-dimensional Heisenberg model with periodic boundary conditions in Ref. [95]. It is stressed that the authors

make use of the translational invariance and take the variational parameters of the RBM to be real, which is sufficient as long as the transformation in Sec. 2.5 is deployed. Thus, the results are not directly comparable to the ones in Sec. 7.1. The results in this reference are of interest for a different reason that will become clear in a moment. To obtain results on the 8×8 square lattice, the number of hidden units has to be increased to $\alpha = \frac{M}{N} = \frac{N}{2}$ to reach an accuracy that is equivalent to the one obtained in a comparable variational ansatz, which is itself known as the many-variable VMC method (mVMC) [10]. This ansatz is supplemented by a spin ($s^z = \sum_i s_i^z = 0$) and momentum ($K = 0$) projection. The full ansatz is thus written in its full glory as

$$|\Psi_{mVMC}\rangle = \mathcal{L}^{S^z} \mathcal{L}^K P_G(\infty) |\Psi_{PP}\rangle, \quad (7.1)$$

where the respective projections are denoted by \mathcal{L} and $|\Psi_{PP}\rangle$ is the projected-pair (PP) ansatz state discussed in Sec. 6.3.3. Due to the substantial number of parameters necessary in the RBM ansatz, surpassing the count of $MN = \frac{N^3}{2}$, the ansatz was alternatively combined with the PP ansatz itself, as described in Sec. 6.5. Additionally, the momentum projection and the one onto the subspace $s^z = 0$ were used as well. This combined ansatz yields more accurate results for $\alpha = 2$ already [95]. After that, the authors applied this ansatz to investigate the J_1 - J_2 model on the square lattice [5]. In that way, they managed to validate the ground state phase diagram for the J_1 - J_2 model in Fig. 2.3.

7.3 Exact Representations with Deep Boltzmann Machines (DBMs)

MLFFNNs like the deep Boltzmann machine (DBM) or deep CNNs, discussed in Sec. 5.4, can be used as just another ansatz for a variational wave function in the sense of Th. 2. However, due to the success of RBMs and similar FFNNs, their usage as universal approximators is, as discussed so far especially in Sec. 7.1, not imperative for the spin models of interest in this work. A related, but still very different approach including DBMs is left out and has not been investigated in this study, since it is fundamentally different. Instead of optimizing a loss function, the idea of this method is to exactly represent ground states by changing the architecture of the DBMs in every step of the optimization [186]. This is to be understood in the sense of Th. 3. With this procedure, the action of the imaginary-time evolution operator is directly translated into an adaption of the DBM architecture itself. It is emphasized that this action needs to be calculated separately for each model. This way, an alternative to the path-integral formulation is developed [186]. Correlation functions like the energy are then also obtained by a sampling that is similar to the one performed in QMC. The sampling error thus solely comprises the error of this approach. However, this method is also reliant on the Hamiltonian to be stoquastic (see Sec. 2.5) and is thus not applicable for real-time evolution nor the imaginary-time evolution of frustrated models.

The representational power of DBMs is investigated thoroughly in Ref. [155]. Therein, it is shown that DBMs can efficiently represent most physical states including the ground states of k -local Hamiltonians with k referring to the k -body interactions being, for example, $k = 2$ for the Heisenberg or the J_1 - J_2 model.

7.4 Wave Function for Finite-Temperature Computations

All in all, the reasoning for using the RBM as the main ansatz for the finite-temperature computations are explained from the findings of the ground state optimization. First of all, there is no clear reasoning for MLFFNNs being a substantially more efficient approximator, neither in general NN theory in Sec. 5.9 nor in the specific application of wave functions to find the ground state of spin chains. Since the SLFFNNs are easier to handle, we choose them over the MLFFNNs. For the question of making use of symmetries and especially translational invariance, there are two reasons why we chose to not apply any of them at all. The first was due to the long-lasting uncertainty about the issue of barren plateaus for models whose ground states possess a complicated sign structure. It is stressed that this prevailed for a variety of SLFFNNs and MLFFNNs. The second reason was that we were not only interested in translationally-invariant systems, but also intended to look at open boundary conditions. Finally, we were left to choose between the fully-connected RBM and the SA ansatz. Although we did not observe any evident difference between these two approaches in the ground state optimization of the one-dimensional Heisenberg model, the decision was made to adopt the RBM ansatz in Eq. (5.26). It is the more popular and well studied one and due to its universality in the following simply referred to as the NN ansatz.

$$\Psi_{\mathbb{W}}(\mathbf{s}) = e^{\sum_{n=1}^N a_n s_n^z} \prod_{m=1}^M \cosh \left(\sum_{n=1}^N W_{mn} s_n^z + b_m \right) \quad (7.2)$$

Since the PP ansatz is of interest from a quantum typicality perspective, the PP-NN ansatz is investigated as an ansatz for a variational thermal pure quantum state in Sec. 8.3.

Chapter 8

Finite-Temperature Results for the One-Dimensional Heisenberg Model

Understanding the ground states of quantum lattice Hamiltonians with precision is crucial for comprehending various states of matter and quantum phase transitions between them. As described in Ch. 7, this has been investigated in detail for NNQS based on various FFNN ansatz wave functions. Additionally, there is of course good reason to study the efficient computation of finite-temperature results as well.

One is the existence of finite-temperature phase transitions in models with dimension $d \geq 2$, which, for example, has been suggested to occur for the J_1 - J_2 model in Ref. [57]. This is the topic of Ch. 10. Another one is to provide comparable results for physical experiments, for example with cold atoms, which are always conducted at finite temperatures [187, 188, 189]. Since the QMC method is in principle not available for non-stoquastic Hamiltonians, as mentioned in Sec. 2.5, the Heisenberg Hamiltonian (see Eq. (2.16)) has been chosen to be the toy model of choice. It allows for the comparison with the QMC results obtained by the stochastic series expansion code available at [190] and described in Ref. [191]. It is emphasized, that the DMRG method is, besides the VMC approach, the method of choice for models for which the QMC method is not applicable, at least for one-dimensional models [192, 193].

In essence, the numerical objective is to approximate the partition function and the expectation value of quantum physical operators A

$$Z = \text{Tr} [e^{-\beta H}], \quad f = -\frac{T}{N} \ln Z, \quad \langle A \rangle = \frac{\text{Tr} [A e^{-\beta H}]}{Z}. \quad (8.1)$$

To the best of my knowledge, there are three approaches available in a variational setting: introducing ancilla sites in the purification method, sampling the trace using, for example, minimally entangled thermal states (METS), and employing quantum typicality. In this chapter, these three approaches are studied and the numerical results for large spin chains are compared. We note that the first approach was already considered for deep Boltzmann machines in Refs. [194, 195]. However, the method to obtain the initial, infinite-temperature state is different from our approach. We will discuss this point in more detail below. The second approach has also very recently been studied in

the neural network context in Ref. [90]. One of the main goals of this work is to compare these three different approaches instead of focusing on only one of them. It is emphasized that the numerical results of this chapter have been published already in Ref. [80] (Wagner, Klümper, Sirker).

8.1 Purification Method

The first approach to compute thermodynamic quantities is the so called purification or ancilla method. It allows to relate mixed states to pure states in an extended system. It is the inherently most appealing procedure presented here, since the infinite-temperature state is easily represented exactly in the variational approach and the results are valid for any system size N .

8.1.1 Purification

Purification is a general feature of arbitrary mixed quantum states. Take an arbitrary mixed state represented by the density matrix

$$\rho = \sum_k p_k |k\rangle_1 \langle k|_1, \quad (8.2)$$

where $\{|k\rangle_1\}_{k=1,\dots,K}$ is an orthonormal basis of \mathcal{H}_1 and $\sum_k p_k = 1$ [2]. Take \mathcal{H}_2 to be a copy of \mathcal{H}_1 and define the pure state $|\Psi\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$ as follows

$$|\Psi\rangle = \sum_k \sqrt{p_k} |k\rangle_1 \otimes |k\rangle_2, \quad (8.3)$$

where $\{|k\rangle_2\}_{k=1,\dots,K}$ is an orthonormal basis of \mathcal{H}_2 . The original mixed state is recovered by tracing out the second Hilbert space [196].

$$\rho = \text{Tr}_2 |\Psi\rangle \langle \Psi| \quad (8.4)$$

Thus, for the spin models of interest ($K = 2^N$), a purified state can always be obtained by adding an ancilla spin for every spin of the system [197].

$$|\{\mathbf{s}\}\rangle \rightarrow |\{\mathbf{s}\}\rangle \otimes |\{\mathbf{s}'\}\rangle \equiv |\{\mathbf{s}, \mathbf{s}'\}\rangle \quad (8.5)$$

This results in an overall spin chain of length $2N$. The size of the spin chain, that is actually worked with in the algorithm, is thus double the size of the physical spin chain. To obtain the correct infinite-temperature state of the system, a purified state has to be chosen that reduces to the infinite-temperature density matrix after tracing out the Hilbert space of the ancillas (see equation (8.4)). Due to the fundamental equation for the partition function

$$Z(\beta) = \text{tr}[e^{-\beta H}], \quad (8.6)$$

the infinite-temperature state is in any basis given by the state with all wave function components

having equal weight [198], or, to put it differently, by maximally entangling each spin with its respective ancilla spin [193].

$$|\Psi(\beta = 0)\rangle = \sum_{\mathbf{s}=-\mathbf{s}'} |\mathbf{s}\rangle \otimes |\mathbf{s}'\rangle \quad (8.7)$$

However, it is stressed that the choice of this state is not unique but all states

$$|\Psi(\beta = 0)\rangle = \sum_{\mathbf{s}=-\mathbf{s}'} |\mathbf{s}\rangle \otimes U|\mathbf{s}'\rangle \quad (8.8)$$

with U a unitary matrix only acting on the ancilla spins, are a valid choice for the infinite-temperature state in the combined system. We can think of the ancillas as an ‘index’ keeping track of the initial state of the real sites before we have acted on this state with some operator. In the purification method, it needs to be respected that the operators of interest only act on the real spins while keeping the ancilla spins unchanged. The respective state at any temperature $T = \frac{1}{\beta}$ is obtained by evolving the infinite-temperature state in imaginary time τ by

$$|\Psi(\tau)\rangle = (\exp(-\tau H/2) \otimes \mathbb{1})|\Psi(0)\rangle . \quad (8.9)$$

If the imaginary-time evolution can be performed error-free, the correct partition function in Eq. (8.6) is obtained as the norm of the wave function for $\beta = 2\tau$. However, it is emphasized that the imaginary-time evolution is performed on the variational manifold of the Hilbert space $\mathcal{V}(\mathcal{H}, \mathbb{C})$. This was described in detail in Sec. 3.2.2. The approximation causes an error that depends on how accurate the wave function components can be approximated by the variational wave function. The actual temperature is obtained after the time evolution via the β modification using the ratio γ_1 in Eq. (4.38). Unfortunately, this modification only holds for correlation functions. The norm of the state is allowed to change here because it carries physical information that is used to compute the partition function and the free energy (see Sec. 8.1.2). The expectation value of an arbitrary operator A is computed by

$$\langle A \rangle_\tau = \frac{\langle \Psi(\tau) | (A \otimes \mathbb{1}) | \Psi(\tau) \rangle}{\langle \Psi(\tau) | \Psi(\tau) \rangle} \quad (8.10)$$

where $\langle \Psi(0) | \Psi(0) \rangle = d^N$ with $d = 2$ the local dimension for spin-1/2 models. It is noteworthy that the two sheets, comprised of the real spins and the ancilla spins, respectively, become interconnected upon performing a partial trace over the ancillas, thereby forming a cylindrical geometry. In the numerical calculations, one might therefore find it advantageous to move half of the imaginary time evolution to the ancilla spins and to apply $\exp(-\tau H/4)$ on both sheets, which makes the problem more symmetric. Likewise, the expectation value of an arbitrary operator is then derived in a more symmetric manner as

$$\langle A \rangle_\tau = \frac{1}{2} \frac{\langle \Psi(\tau) | (A \otimes \mathbb{1} + \mathbb{1} \otimes A) | \Psi(\tau) \rangle}{\langle \Psi(\tau) | \Psi(\tau) \rangle} . \quad (8.11)$$

However, for large system sizes, this procedure doubles the computation time for the energy, which is crucial whenever $(H^2)_{\Psi(\tau)}$ is computed in order to calculate the β modification described in Sec. 4.4. Hence, it is not used in those computations. The description of the purification algorithm so far is general and can be performed with any variational ansatz wave function, provided that the initial

infinite-temperature state (8.5) can be represented accurately. The question remaining is how the state in equation (8.7) can be represented by a variational ansatz in general and the NN ansatz in Eq. (7.2) in specific. We note that the performance of the algorithm discussed below might possibly be improved by using deep neural networks (DNN) as, for example, convolutional neural networks (CNN), along similar lines as in the ground state optimization algorithms [89]. Here, the focus is on demonstrating how thermodynamic calculations can be performed in principle and what the fundamental limitations are. For convenience, we label both the real and the ancilla spins as s_i with $i = 1, \dots, N$ for the real spins and $i = N + 1, \dots, 2N$ for the ancillas. For the components of the wave function, we then choose to adjust the NN ansatz in Eq. (7.2) to be

$$\Psi_{\mathbb{W}}(\mathbf{s}) = \exp\left(-\sum_{n=1}^{2N} a_n s_n^z\right) \prod_{m=1}^M \cosh\left(\sum_{n=1}^{2N} W_{mn} s_n^z + b_m\right) \exp\left(-\sum_{n=1}^{2N} c_n (s_n^z + s_{(n+N)\bmod(2N)}^z)^2\right), \quad (8.12)$$

where $\mathbf{a}, \mathbf{c} \in \mathbb{C}^{2N}$ and $\mathbf{b} \in \mathbb{C}^M$, while $\mathbf{W} \in \mathbb{C}^{2N \times M}$ is the weight matrix. The modification entails the inclusion of the last term governed by \mathbf{c} , which couples every spin with its respective ancilla. The ansatz in Eq. (8.12) is thus referred to as the modified neural network (mNN) ansatz. The (not normalized) infinite-temperature state $|\Psi(0)\rangle$ can now formally be obtained by $c_i \rightarrow +\infty$, $i = 1, \dots, 2N$, which exponentially suppresses all components where real site and ancilla site do not have opposite signs and gives all components where the signs are opposite equal weight in the wave function, if the other parameters are kept finite. In practice, however, this limit cannot be performed exactly because numerical gradient methods—required to change the parameters of the wave function during the imaginary-time evolution—will then not be able to find the proper descent path. We therefore set $c_i = c = \text{const} > 0$ and real. Furthermore, we choose the elements of \mathbf{a} and \mathbf{W} from a random normal distribution centred around zero with a very small width $w \sim 10^{-2}$ and keep $\mathbf{b} = 0$. In Refs.[194, 195], purification was employed without introducing the last term in Eq.(8.12). Instead, the parameters \mathbf{W} were adjusted to achieve the infinite-temperature state. It's worth noting that in our method, the suppression of unwanted contributions to the wave function is exponential, whereas using the \mathbf{W} parameters to realize the infinite-temperature state leads to only quadratic suppression if some noise is introduced to these parameters. Our approach allows for the addition of noise to the \mathbf{W} parameters while maintaining exponentially suppressed unwanted contributions.

8.1.2 Macroscopic Thermodynamic Variables

Besides expectation values of local quantum operators, it is also of interest in thermodynamics to compute macroscopic thermodynamic variables. They are in the VMC method, for example, attainable by indirect computations from the energy $E(T)$ [199]. To compute the free energy and the entropy of the system, the specific heat capacity is obtained first in the following way

$$C(T) = \partial_T E(T) = \frac{\Delta H^2}{T^2}, \quad (8.13)$$

where $\Delta H^2 = \langle H^2 \rangle - \langle H \rangle^2$. There is no additional computation needed in the algorithm since the quantity $\langle H^2 \rangle$ is computed for the β modification anyway (see Sec. 4.4). This can then be used to obtain the entropy and the free energy as well by

$$S(T) = \int_0^T \frac{C(T')}{T'} dT' \quad F(T) = E(T) - TS(T) . \quad (8.14)$$

For this computation to be accurate, it is crucial that the numerical results for the specific heat capacity and/or the energy are accurate. This can be circumvented by computing the free energy directly in the algorithm. This sounds counter-intuitive in the beginning since these quantities are in the ensemble formulation always related to the number of microstates. However, the normalization constant of the single, unnormalized state $|\Psi(\beta)\rangle$ contains the information necessary to calculate these thermodynamic quantities, as is derived in Refs. [198, 196], since the partition function is given by

$$Z(\beta) = \langle \Psi(\beta) | \Psi(\beta) \rangle , \quad (8.15)$$

and the norm of the state is itself adjusted to the correct value by setting $2^N = \langle \Psi(0) | \Psi(0) \rangle$. Thus, the computation of any other thermodynamic quantity, as for example the free energy

$$f(\beta) = \frac{-1}{N\beta} \log \langle \Psi(\beta) | \Psi(\beta) \rangle , \quad (8.16)$$

is feasible. Unfortunately, there is no direct β modification available. Thus, the imaginary time τ is for the free energy equal to the inverse temperature β . It is stressed that the β modification for correlation functions, on the other hand, is available and computed using the ratio γ_1 in Eq. (4.38).

8.1.3 Numerical Results

To investigate the performance of the proposed algorithm, the spin-1/2 Heisenberg chain is investigated, with the Hamiltonian introduced in Eq. (2.9), as a test case. We use open boundary conditions and the canonical transformation from Eq. (2.15). To consider the peculiarities of this method, the analysis is started with small system sizes in which the full Hilbert space can be taken into account, and compared to exact diagonalization results. The effect of the Monte Carlo sampling is subsequently investigated independently before the algorithm is applied to larger system sizes. We note that the approach was already considered for deep Boltzmann machines in Refs. [194, 195]. However, the method to obtain the initial, infinite-temperature state is different from our approach.

Numerical Results for Small Systems

First, the results for the isotropic Heisenberg model with $N = 7$ sites using the purification algorithm are shown in Fig. 8.1, where all basis states are kept. Note that due to the doubling of the number of sites when adding the ancillas, this already corresponds to a Hilbert space dimension of 2^{14} . The results show that for $M = 28$ hidden units in the mNN ansatz in Eq. (8.12), the exact results for

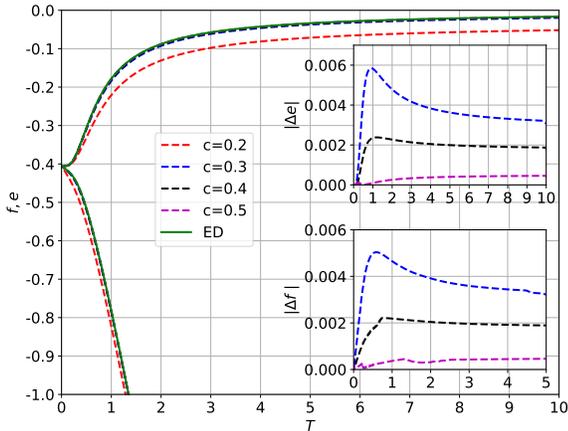


Figure 8.1: Inner and free energy for an open Heisenberg chain with $N = 7$ sites obtained using the purification algorithm without the β modification, a desired tolerance of $\epsilon = 10^{-9}$ and $R = 2 \cdot 10^3$ iterative steps [80]. In the mNN ansatz, $M = 28$ hidden units are used. The insets show the absolute errors for $c = 0.3, 0.4, 0.5$.

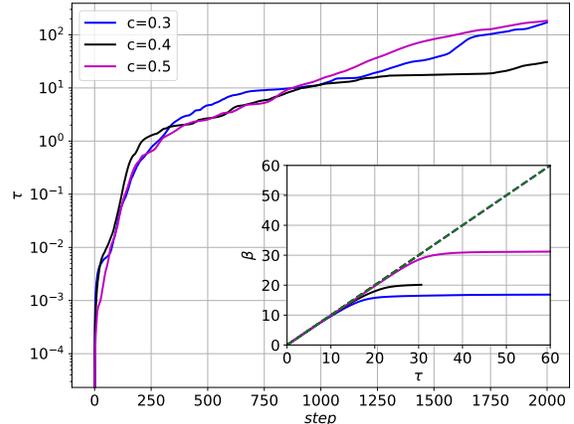


Figure 8.2: Imaginary time τ in the Heun algorithm versus number of iterative steps for an open Heisenberg chain with $N = 7$ sites [80]. The parameters are the same as in Fig. 8.1. The β modification using γ_1 in Eq. (4.38) is depicted in the inset.

the energy $e(T)$ and the free energy $f(T)$ are reproduced with absolute errors $\sim 10^{-3} - 10^{-4}$. If the parameter c in the mNN ansatz is chosen too small, the energy of the infinite-temperature state is strongly deviating from zero—the infinite-temperature value of the energy for the chosen form of the Hamiltonian in Eq. (2.9)—and thus a bad approximation of the infinite-temperature state. On the other hand, if c is too large and potentially infinite, the gradients reduce to zero. The course of the imaginary time in the Heun algorithm (see Sec. 4.3) is different depending on the respective initial state and the topology of the area that is traversed in the time evolution (see Fig. 8.2). The effect of the β modification using γ_1 in Eq. (4.38) depicted in the lower right inset is negligible for temperatures larger than the finite size gap and is smaller than the inherent error of the approach and is thus not visible in Fig. 8.1.

For larger systems, it will no longer be possible to use the full Hilbert space. Instead, a Monte Carlo sampling using the Metropolis algorithm, as explained in Fig. 4.1, is employed. To better understand potential additional issues, which may arise due to the sampling, the results for the same system size, $N = 7$, are shown in Fig. 8.3. Here, the results for the energy are as accurate as those for the full Hilbert space simulations if sufficiently many Monte Carlo samples are used and provided that the initial state is carefully chosen. The latter turns out to be more intricate here. First, a certain amount of noise in the \mathbf{a} and \mathbf{W} parameters is required for the imaginary time evolution to work properly. Both sets of parameters are drawn from independent normal distributions with a width $w = 0.02$. Second, the initial parameters $c_i = c = \text{const} > 0$ have to be chosen just large enough to obtain $\langle \Psi(0) | H | \Psi(0) \rangle \sim 10^{-3} - 10^{-4}$. If the initial c parameter is chosen too large, then the algorithm remains stuck close to $|\Psi(0)\rangle$ for a large number of steps and also shows instabilities. This is due to the nature of the infinite-temperature state. Increasing the parameter c comes with

an increase in the rejection probability in the Monte Carlo sampling and thus hinders the algorithm to explore all configurations violating the ergodicity feature. It is stressed that to go from one configuration that is fully correlated ($\mathbf{s} = -\mathbf{s}'$) to another one, two spins need to be flipped, namely one real and one ancilla spin. Thus, the Monte Carlo sampling, proposing the next configuration by only one spin flip, needs to traverse a configuration that is not fully correlated and thus has a small probability. This issue is investigated more thoroughly in Ch. 10, in which the purification method is modified in Sec. 10.2. It is stressed that the results for the energy $E_{\Psi_{\mathbb{W}}(\tau)}$ and $(H^2)_{\Psi_{\mathbb{W}}(\tau)}$ need to be fitted to obtain reasonable results for the β modification. However, the computation of γ_1 at low temperatures is challenging, because both the numerator and denominator are small quantities, which are calculated with an accuracy determined by the Monte Carlo sampling error.

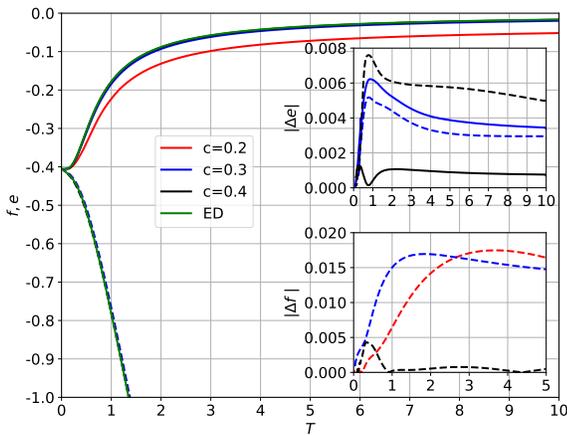


Figure 8.3: Inner and free energy for an open Heisenberg chain with $N = 7$ sites obtained using the purification algorithm and a Monte Carlo sampling of the Hilbert space with $M = 28$, $4 \cdot 10^4$ samples, $\epsilon = 10^{-7}$ and $R = 2 \cdot 10^3$ [80]. The results with β modification are shown as solid lines, the ones without as dashed lines. The error in the energy for the case $c = 0.2$ is not shown in the upper inset.

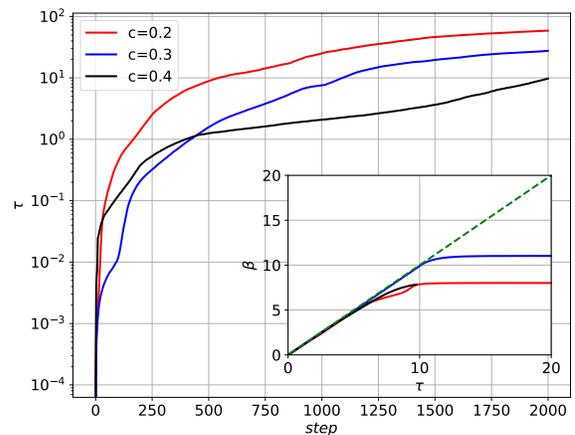


Figure 8.4: Imaginary time τ in the Heun algorithm versus number of iterative steps for an open Heisenberg chain with $N = 7$ sites [80]. The parameters are the same as in Fig. 8.3. The β modification using γ_1 in Eq. (4.38) is depicted in the inset.

Furthermore, the approximation of the infinite-temperature state is bounded by the sampling error. Thus, it is not feasible in this case to use values $c > 0.4$. Making c too small, on the other hand, leads to an initial error given by the non-zero initial energy which remains present down to low temperatures. In practice, we proceed as follows: With w small and fixed, the results for a series of calculations with increasing initial values for c and different values for the desired tolerance ϵ are obtained. The results are in general comparable for a certain range of initial values for c . Furthermore, it is noted that the error is dominated by the sampling error which is why an increase in the number of hidden units M does not have a measurable effect. This can be deduced as well from the β modification having no substantial effect up until low temperatures. Making M too large, generally seems to lead to additional instabilities. For the considered model, $M \in [N, 4N]$ seems to be optimal. While we find that this approach to find the right initial parameters and the

right number of hidden units does work, it is noted that it requires to run multiple simulations with different initial parameters c and ϵ . We also note, that without being able to compare with other methods it might not always be easy to decide what the optimal initial parameters are. Furthermore, we note that the free energy is more sensitive than the inner energy to the accuracy with which the infinite-temperature state is approximated (see Fig. 8.3).

Numerical Results for System Sizes beyond Exact Diagonalizations

In this section, we want to briefly show that the purification algorithm can be scaled up and used to access the thermodynamics of quantum lattice models for system sizes outside the reach of exact diagonalization. We want to stress again that the purpose of this part is a proof of principle demonstration of using neural networks to investigate thermodynamic properties and the discussion of certain generic obstacles. There are a number of directions in which the algorithms presented here could possibly be improved, including but not limited to the use of more complicated network structures such as CNNs or modifications to the imaginary-time evolution by using, for example, implicit iterative methods.

As an example, we present in Fig. 8.5 results obtained using the purification algorithm for a Heisenberg chain with 30 sites and compare the results to QMC results for the energy and the free energy, where the latter, in the QMC simulations, is obtained by the respective integration of the former (see Eq. (8.14)). We used the stochastic series expansion code available at [190] which is described in Ref. [191]. As for the smaller system sizes considered in the previous section, a post-selection of data is still required and can be done in a systematic manner. We find again that the parameter range $c = 0.3...0.4$ gives initial energies $|e| \sim 10^{-3}$ and allows for an accurate imaginary time evolution. The algorithm is executed for different tolerances. For $c = 0.34$ and $M = 60$, the results for the energy appear to be converged sufficiently for the β modified results to stay within the mentioned accuracy for all temperatures. Increasing the number of hidden units results in a small improvement for low temperatures, but instabilities that occur earlier in imaginary time τ . However, as the number of hidden units increases, the algorithm is able to proceed faster in inverse temperature β to reach lower temperatures. We want to emphasize that the β modification depicted in Fig. 8.6 circumvents the need to increase the number of hidden units to an extent that might not be numerically feasible anymore. The results for the energy without β modification shown as dashed lines are about one order of magnitude less accurate than the modified data. The β modification is thus a useful tool to achieve a good accuracy without having to increase the number of hidden units to the point where numerical instabilities become more and more problematic. The lowest temperatures reached are limited by the progress of the Heun method and the β modification, which at some point result in a temperature which is not decreasing anymore. To overcome this, increasing the number of hidden units and the number of sampled states is ultimately inevitable. To obtain a better approximation of the ground state, the regularization has to be adjusted by adding a diagonal matrix with entries being around $10^{-4} - 10^{-6}$ (see Ref. [11] and [95]) to the matrix \mathbf{G} explained in Sec. 4.5. This is however not possible for the computation of finite temperature results since it distorts the evolution itself. We find that the free energy is computed to an accuracy of $\sim 10^{-3}$ down to low temperatures as well. Unfortunately, the results differ more at temperatures close to or below the finite size gap (see free energy in Fig. 8.5). At even lower temperatures, the simulation

has to be stopped in this case due to larger sampling errors. This issue is well-known in Monte Carlo simulations. It is commonly referred to as *critical slowing down* and linked to an increase in the rejection probability of the Monte Carlo algorithm violating the ergodicity condition [200, 88]. It results in a flawed approximation of the probability distribution assigning high probabilities to configurations that are barely present. In NN theory, this is known as a spurious mode [201, 113]. It is in detail studied for the two-dimensional J_1 - J_2 model in Sec. 10.5.

Very recently, the authors of Ref. [194] proposed a Hadamard transformation of the infinite-

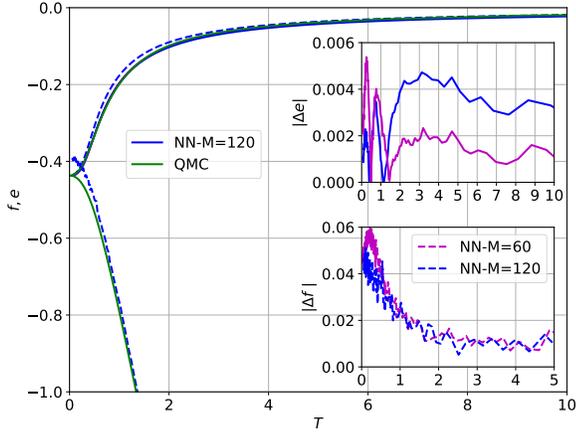


Figure 8.5: Inner and free energy compared to the QMC solution for an open Heisenberg chain with $N = 30$ sites obtained using the purification algorithm, $4 \cdot 10^4$ samples and $R = 2 \cdot 10^3$ iteration steps [80]. We choose $c = 0.34$ and $\epsilon = 10^{-6}$. Results without the β modification are represented by dashed lines. The results for $M = 60$ are omitted in the main plot for visual clarity.

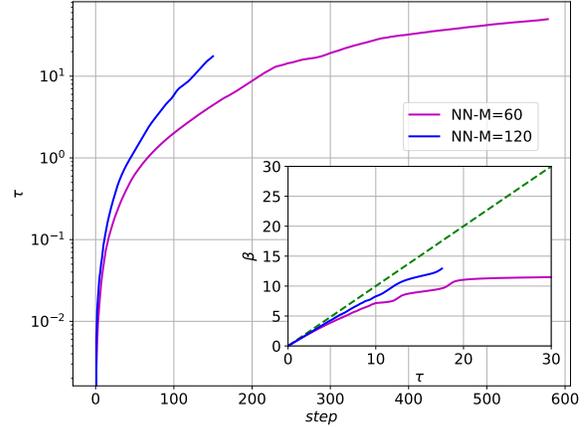


Figure 8.6: Imaginary time τ in the Heun algorithm versus number of iterative steps for an open Heisenberg chain with $N = 30$ sites [80]. The parameters are the same as in Fig. 8.5. The β modification using γ_1 in Eq. (4.38) is depicted in the inset.

temperature state to circumvent some of the issues related to small gradients at high temperatures. While all purifications are unitarily equivalent, it is possible that some provide a better starting point for numerical computations than others. It is stressed that there is no β modification applicable for the free energy, since it is sensitive to the norm of the variational state in contrast to the energy. The β modification is thus ineffective for the computation of other thermodynamic variables that have to be obtained from the free energy. Furthermore, the energy results cannot be computed accurately at temperatures low enough to allow for the computation of the free energy.

Numerical Results for Correlation Functions

To check how well the proper thermodynamics is approximated beyond the energy, we compute also the correlation functions $C_{\parallel}^d = \langle S_k^z S_{k+d}^z \rangle$ and $C_{\perp}^d = -\frac{1}{2} \langle S_k^+ S_{k+d}^- + S_k^- S_{k+d}^+ \rangle$. In Fig. 8.7, the results for $k = 10$ and $d = 2$ are compared to the QMC results and agree to an accuracy of order 10^{-3} .

The deviations observed in Fig. 8.7 are dominated by the Monte Carlo sampling error. We note that the SU(2) symmetry is broken, in general, but that it is restored at low temperatures. The implementation of the SU(2)-symmetry into the NN ansatz in Eq. (7.2) was proposed in Ref. [202]. However, this comes with an additional numerical cost for a basis transformation, which is why this approach has not been able to establish itself.

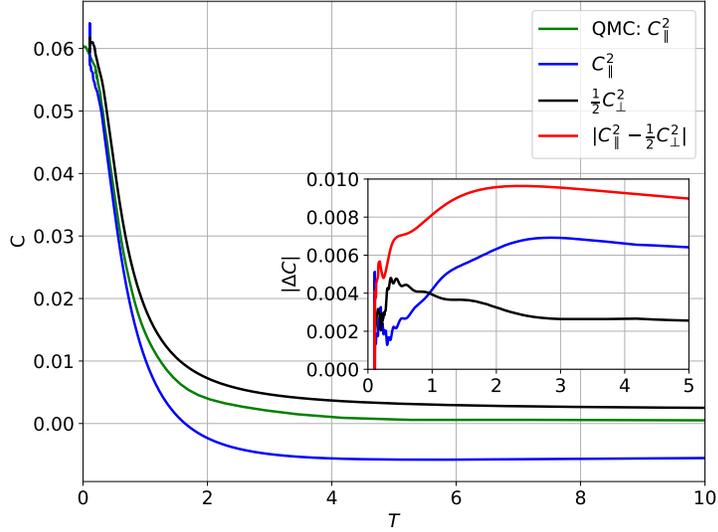


Figure 8.7: Correlation functions $C_{\parallel,\perp}^2$ for an open Heisenberg chain with $N = 30$ sites obtained using the purification algorithm are compared to the QMC solution. The parameters are the same as in Fig. 8.5 and $M = 60$ [80]. The inset shows $|C_{\parallel}^2 - \frac{1}{2}C_{\perp}^2|$ and the deviations of the two correlation functions from the QMC result.

8.2 Sampling Method

To compute the ensemble average of a quantum mechanical operator A (see Eq. (8.1)) at infinite temperature, one needs to compute the expectation value with respect to every vector of an orthonormal set $\{|k\rangle\}_{k=1}^K$

$$\langle A \rangle_{\beta=0} = \frac{\text{tr}[A]}{2^N} = \frac{1}{2^N} \sum_{k=1}^K \langle k|A|k\rangle. \quad (8.17)$$

The idea is now that instead of working with a full orthonormal set we can also sample the trace using random states. These are constructed by choosing K random vector components c_k from a symmetric probability distribution ([203]) to form the infinite-temperature states

$$|\Psi(\beta = 0)\rangle = \sum_{k=1}^K c_k |k\rangle. \quad (8.18)$$

It is essential that the vector components are uncorrelated. The ensemble average is then approximated by using S realizations of those states.

$$\frac{1}{S} \sum_{s=1}^S \langle \Psi_s(\beta = 0) | A | \Psi_s(\beta = 0) \rangle \approx \frac{1}{S} \sum_{s=1}^S \sum_{j,k=1}^K c_{j,s}^* c_{k,s} \langle j|A|k\rangle = E(|c_{j,s}|^2) \text{tr}(A) \quad (8.19)$$

For normalized states, the expectation value of any $|c_{j,s}|^2$ is $E(|c_{j,s}|^2) = \frac{1}{K}$ and thus the correct value of the ensemble average at infinite temperature is obtained. The error goes in general like $\frac{1}{\sqrt{SK}}$. An upper bound is computed in Ref. [203] in terms of the operator A . To include this idea into a variational algorithm using the NN ansatz, the parameters of the NN have to be chosen properly. States that fulfill the above criteria and are efficiently represented by the NN are, for example, the so called minimally entangled typical thermal states (METTS) [197]. The term *typical* will later be defined and used in Sec. 8.3 in the sense of quantum typicality. To avoid confusion, it is dropped in this work for the notation of these states.

8.2.1 Minimally Entangled Thermal States (METTS)

The minimally entangled thermal states (METTS) were first discussed as a tool to perform the imaginary-time evolution in a standard DMRG algorithm [197]. In Ref. [90], the approach was then generalized for the NN ansatz in Eq. (7.2). A random METS is defined as

$$|\Psi_0\rangle = \bigotimes_{j=1}^N \left(\frac{\xi_{j\uparrow} |\uparrow\rangle_j + \xi_{j\downarrow} |\downarrow\rangle_j}{\sqrt{|\xi_{j\uparrow}|^2 + |\xi_{j\downarrow}|^2}} \right), \quad (8.20)$$

where the vector $\boldsymbol{\xi}$ is chosen according to a standard complex normal distribution $\mathbb{CN}(\mu = 0, C = 1)$ with μ the mean and C the covariance matrix of the distribution

$$P_0(\boldsymbol{\xi}) = \left(\frac{1}{\pi}\right)^{2N} \exp\left(-\sum_{j=1}^N \sum_{\sigma=\uparrow,\downarrow} |\xi_{j\sigma}|^2\right). \quad (8.21)$$

This is equivalent to choosing the real and imaginary part according to a real normal distribution $N(\mu = 0, \sigma = 1/2)$ with σ being the variance. A METS state can be obtained in the NN ansatz in Eq. (7.2) up to an irrelevant overall constant by setting

$$a_j = -\frac{1}{2} \ln\left(\frac{\xi_{j\uparrow}}{\xi_{j\downarrow}}\right), \quad (8.22)$$

$\mathbf{W} = 0$ and $\mathbf{b} = 0$, see Ref. [90]. These states will be referred to as METS-NN from here on. Numerically, one needs to add some noise to the NN parameters \mathbf{W} and \mathbf{b} as in the purification method. The real and imaginary parts of both of them are chosen from a normal distribution centered around zero with $\sigma_{NN} = 1/N$, maintaining a constant variance of the wave function components across variable system sizes.

8.2.2 Numerical Results

The set of METS-NN states at infinite temperature is evolved by the adaptive Heun method as explained in Sec. 4.3. The imaginary time of the algorithm τ is afterwards modified as explained in Sec. 4.4.

Numerical Results for Inner and Free Energy

The results for the inner and the free energy of a Heisenberg chain of length $N = 30$ with $S = 30$ states are depicted in Fig. 8.8. The accuracy for the energy is here also of order 10^{-3} , which is expected from the sampling error. We also note that the temperatures reached using the purification method are lower than the ones reached here. This is depicted in Fig. 8.9 and due to the already mentioned convergence issues at low temperatures (see Sec. 8.1.3). Compared to the purification method, this problem is aggravated here because one needs to simulate ~ 30 states to low temperatures without any such instabilities occurring. Additionally, the state in which these instabilities occur first defines the minimum temperature that can be reached. The accuracy of the free energy is as in the purification method also of order 10^{-2} . The algorithm also breaks down at lower temperatures for the same reason mentioned in Sec. 8.1.3. In Sec. 10.5, it is investigated that this is related to a failure of the Monte Carlo sampling and even calls for a conceptual change of the way the configurations are sampled. To enable the computation of the free energy by integrating the energy results using Eq. (8.14), as was done for the QMC results, the evolution must reach lower temperatures.

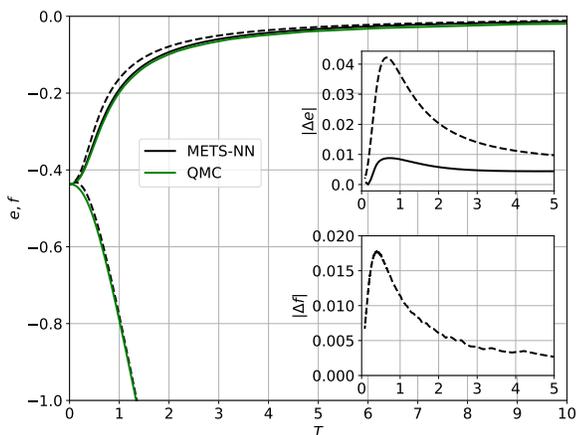


Figure 8.8: Inner and free energy of the METS-NN ansatz compared to the QMC solution for an open Heisenberg chain of length $N = 30$ obtained using the sampling algorithm with $4 \cdot 10^4$ Metropolis samples, $R = 2 \cdot 10^3$ and $\epsilon = 10^{-6}$ [80]. The average is taken over $S = 30$ initial states. The dashed lines show the data without β modification.

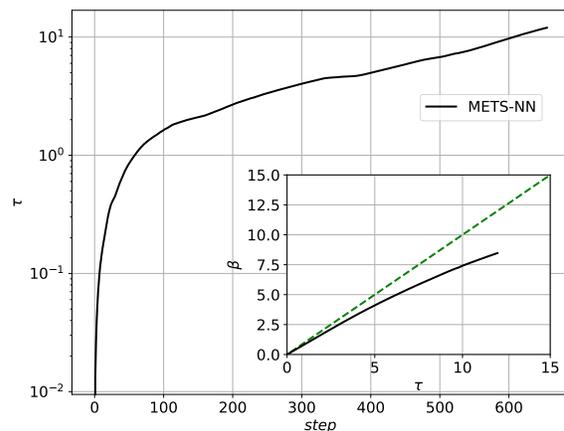


Figure 8.9: Imaginary time τ in the Heun algorithm versus number of iterative steps for an open Heisenberg chain of length $N = 30$ and parameters defined in Fig. 8.8 [80]. The β modification using γ_1 in Eq. (4.38) is depicted in the inset.

Numerical Results for Correlation Functions

For the sampling method, it is checked as well how accurate thermodynamics is approximated beyond the energy by computing spin-spin correlators like C_{\parallel}^d and C_{\perp}^d for $k = 10$ and $d = 2$ (see Sec. 8.1.3). These are not obtained as an average over local operators such as the energy where an average over $N - 1$ bond energies is taken. Thus, the METS-NN algorithm results in larger errors for single spin-spin correlators. In Fig. 8.10, the results for C_{\parallel}^2 and C_{\perp}^2 are compared to the QMC results. It is simply no longer sufficient to use ~ 30 states only. Instead, to obtain a result that is consistent within the Monte Carlo sampling error, ~ 1000 states are required. Note that a preselection of states would break the relation in Eq. (8.19). We emphasize that for any specific single state C_{\parallel}^d and C_{\perp}^d vary, with neither of them being more accurate in general. The results consequently also violate $SU(2)$ invariance.

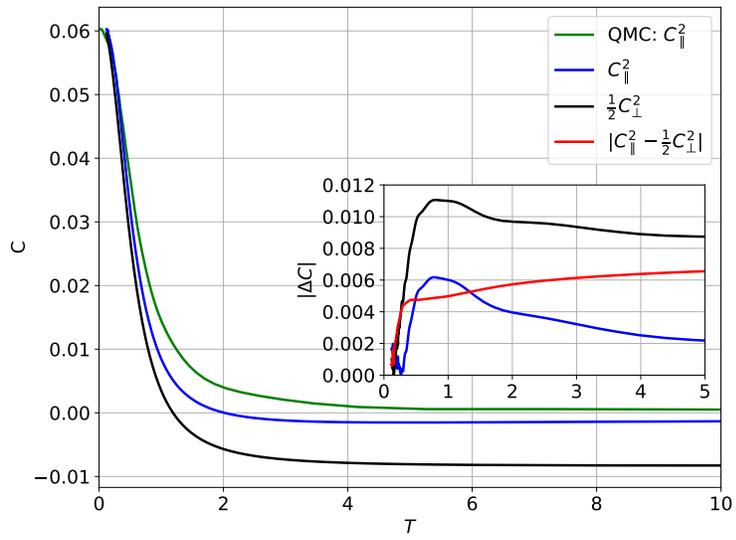


Figure 8.10: Correlation functions $C_{\parallel,\perp}^2$ compared to the QMC solution for a chain of length $N = 30$ obtained using the sampling algorithm [80]. The parameters are the same as in Fig. 8.8 and $M = 60$. The inset shows $|C_{\parallel}^2 - \frac{1}{2}C_{\perp}^2|$ and the deviations of the two correlation functions from the QMC result.

8.3 Thermal Pure Quantum State (TPQS)

The definition of thermal pure quantum states (TPQSs) or quantum typical states comes without the equal a priori probability postulate of statistical mechanics and replaces the ensemble average by individual states. It was first formulated for the canonical situation of a small system S coupled to an environment E [204] and later formulated in a way that is directly accessible to variational algorithms in which only one pure state is considered [196]. While it seems surprising that general quantum physical models at finite temperature, which are fully described by a density matrix, can for large system sizes be described (approximated) by only one pure state, this feature can be deduced from a general feature of normed spaces.

8.3.1 Measure Concentration on Normed Spaces

Normed spaces are vector spaces upon which a norm is defined. In the following, the example of interest will be the $(d - 1)$ -dimensional sphere S^{d-1} . The two-dimensional sphere was encountered in Ch. 3 when discussing the Hilbert space of a spin-1/2 particle, known as the Bloch sphere, which was illustrated in Fig. 3.1. The $(d - 1)$ -dimensional sphere S^{d-1} is defined as all those vectors $\mathbf{x} \in \mathbb{R}^d$ that fulfill the following normalization condition

$$\|\mathbf{x}\| = \sqrt{\sum_k^d x_j^2} = 1. \quad (8.23)$$

An interesting feature of normed spaces, and thus also the $(d - 1)$ -dimensional sphere, is the concentration of measure theorem.

Theorem 4. *Measure concentration on the sphere [4]*

Let $A \subseteq S^{d-1}$, $P[A] \geq \frac{1}{2}$ the probability measure of A and also the probability that a randomly chosen vector of S^{d-1} falls into A , $A_t = \{\mathbf{x} \in S^{d-1} : \text{dist}(A, \mathbf{x}) \leq t\}$ the t -neighbourhood of A , then

$$1 - P[A_t] \leq \exp\left(-\frac{dt^2}{2}\right). \quad (8.24)$$

This can be formulated in a more illustrative way by looking at how the vectors are spread around the equators $x_j = 0$. This is specified by looking at how wide the stripe is in which 90% of randomly chosen vectors are concentrated [4].

$$P[\{\mathbf{x} \in S^{d-1} : -w \leq x_j \leq w\}] = 0.9 \quad \forall j \quad (8.25)$$

The crucial point to note is that the width decreases for large dimensions as $w \propto \frac{1}{\sqrt{d}}$, which is illustrated in Fig. 8.11. Accordingly, if one chooses two random states of the high-dimensional sphere, they are with a high probability practically orthogonal [205]. While this may seem counterintuitive, it is an intrinsic feature of normed spaces that can be motivated by the fact that the expectation

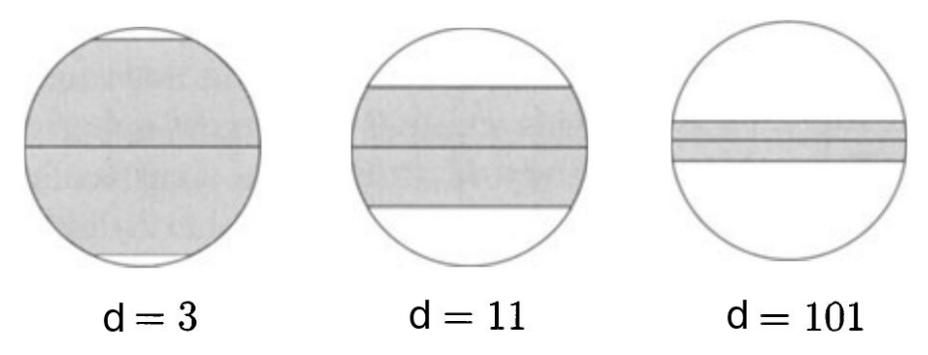


Figure 8.11: Illustration of the width within which 90% of randomly chosen states are located. In the illustration, the additional dimensions are imagined to be added without being visualized. Reprinted with permission from Ref. [4].

value of a single component is actually given by $x_j = \frac{1}{\sqrt{d}}$. Furthermore, a random vector is created by choosing the components from a normal distribution and subsequently normalizing them [206]. Expanding on these considerations, the second building block of this approach is represented (see Refs. [204, 207, 206]). It describes the concentration of real-valued functions on the S^{d-1} -sphere.

Theorem 5. *Levy's Lemma* [204, 207]

Let $f : S^{d-1} \rightarrow \mathbb{R}$ have a finite Lipschitz constant

$$L = \sup_{\mathbf{x}_2, \mathbf{x}_1} \frac{|f(\mathbf{x}_2) - f(\mathbf{x}_1)|}{\|\mathbf{x}_2 - \mathbf{x}_1\|} . \quad (8.26)$$

Then for a random normally distributed $\mathbf{x} \in S^{d-1}$

$$P[|f(\mathbf{x}) - \bar{f}| \geq \epsilon] \leq 2 \exp\left(-\frac{Cd\epsilon^2}{L^2}\right) , \quad (8.27)$$

where $C = \frac{1}{(9\pi)^3}$ and \bar{f} the mean of $f(\mathbf{x})$.

8.3.2 Infinite-Temperature Thermal Pure Quantum State (TPQS)

The obvious connection to quantum physics stems from the discussion of the projective (or normed) Hilbert space in Sec. 3.1. All quantum physical pure states can be written as

$$|\Psi\rangle = \sum_k^d \underbrace{(a_k + ib_k)}_{c_k} |k\rangle , \quad (8.28)$$

where $x_{2k}(\Psi) = a_k$, $x_{2k+1}(\Psi) = b_k$ and $d = 2^N$ for spin models. The vector $\mathbf{x} \in \mathbb{R}^{2d}$ is normalized according to Eq. (8.23) and thus parametrized on the $(2d - 1)$ -dimensional sphere S^{2d-1} . However,

a quantum mechanical system at finite temperature is, in general, in a mixed state. The vectors $\{|l\rangle\}_{l=1}^d$ are assumed to be the eigenvectors of the density matrix ρ

$$\rho = \sum_l^d p_l |l\rangle\langle l| \quad (8.29)$$

and p_l are the respective probabilities with

$$\sum_l^d p_l = 1 \quad \text{and} \quad p_l \geq 0 . \quad (8.30)$$

However, by averaging over the pure states in Eq. (8.28) chosen according to some probability distribution $p(\mathbf{k})$ with $\mathbf{k} = (k_1, \dots, k_d)$, the density matrix in Eq. (8.29) is recovered as

$$\overline{|\Psi\rangle\langle\Psi|} = \sum_{j,k} \overline{c_j^* c_k} |j\rangle\langle k| \stackrel{!}{=} \sum_k p_k |k\rangle\langle k| = \rho , \quad (8.31)$$

whenever the second moments are fixed by [205]

$$\overline{c_j^* c_k} = \delta_{jk} p_k . \quad (8.32)$$

The probability distribution is in general complicated. Fortunately, in the microcanonical ensemble, the probability assigned to every state, that fulfills a certain restriction, like the energy, is constant and given by the number of states $|\mathcal{S}|$ ($0 \leq |\mathcal{S}| \leq d$) of this subset \mathcal{S} .

$$p_l = \begin{cases} \frac{1}{|\mathcal{S}|} & \text{if } |l\rangle \in \mathcal{S} \\ 0 & \text{if } |l\rangle \notin \mathcal{S} \end{cases} \quad (8.33)$$

For the spin models of interest in this work, we consider the whole Hilbert space of dimension $d = 2^N$, for which $\text{Tr}[H] = 0$ the energy at infinite temperature. This is the case in which Levy's Lemma in Th. 5 can be shown to directly apply. The vector space over the real field is of dimension 2^{N+1} . It reduces by equipping it with the norm to a $(2^{N+1} - 1)$ -dimensional sphere $S^{2^{N+1}-1}$ and a random state therein is created easily by choosing its components a_k and b_k in Eq. (8.28) from a normal distribution and normalizing them subsequently. A real-valued function $f : S^{2^{N+1}-1} \rightarrow \mathbb{R}$ can be defined in this quantum physical notation as the expectation value of an operator A using Eq. (8.28).

$$f(\mathbf{x}_\Psi) = \langle\Psi|A|\Psi\rangle = \text{Tr}[|\Psi\rangle\langle\Psi|A] \quad (8.34)$$

The Lipschitz constant for any operator with finite operator norm $\|A\|_{EV}$ is given by $L \leq 2\|A\|_{EV}$ for any $|\Psi\rangle$, where $\|A\|_{EV}$ denotes the modulus of the maximum eigenvalue of the operator A . Furthermore, the ensemble average is rewritten as

$$\overline{\langle\Psi|A|\Psi\rangle} = \overline{\text{Tr}[|\Psi\rangle\langle\Psi|A]} = \text{Tr}[\overline{|\Psi\rangle\langle\Psi|}A] = \text{Tr}[\rho A] . \quad (8.35)$$

Thus, Levy's lemma for quantum expectation values can be formulated as [204, 207]

$$P[|\langle \Psi|A|\Psi \rangle - \overline{\langle \Psi|A|\Psi \rangle}| \geq \epsilon] \leq 2 \exp\left(-\frac{C2^{N-1}\epsilon^2}{\|A\|_{EV}^2}\right), \quad (8.36)$$

where Eq. (8.31) was used to obtain the ensemble average [205]. The crucial finding is that the difference between the expectation value of a randomly chosen pure state from the actual value in the microcanonical ensemble decreases exponentially in system size.

8.3.3 Extension to Finite Temperatures

It is of course desirable to extend the infinite-temperature thermal pure quantum state (TPQS) to the canonical setting at finite temperature. However, the geometrical situation is in principle not as simple since the probability is more advanced. Luckily, the availability of the infinite-temperature state ($\beta = 0$) can be used in combination with the imaginary-time evolution established in Sec. 4.2.1 to compute finite-temperature results. The principal idea is thus to use the infinite-temperature TPQS, defined in Eq. (8.28), as the initial state of the evolution

$$|\Psi(\beta = 0)\rangle = \sum_k^{2^N} c_k |k\rangle, \quad (8.37)$$

where the real and imaginary part of c_k are chosen randomly from a normal distribution centered around $\mu = 0$ and a standard deviation of $\sigma = 1$, and time evolve it in imaginary time τ as

$$|\Psi(\tau)\rangle = \exp\left(\frac{-\tau H}{2}\right) |\Psi(\beta = 0)\rangle. \quad (8.38)$$

The temperature $T = \frac{1}{\beta}$ is again computed via the β modification in Sec. 4.4. While the provided explanations are valid, it is not inherently clear to what extent the approximation from Levy's Lemma in Eq. (8.36) transfers to the finite temperature states obtained in the evolution. Fortunately, this was rigorously investigated in Ref. [196] and is sketched for completeness here.

8.3.4 Canonical Thermal Pure Quantum State (CTPQS)

The closeness of the finite-temperature TPQS, defined in Eq. (8.38), to the correct result from the canonical ensemble is directly computed in Refs. [196, 208] and this way the notion of CTPQSs is justified. The results are cited in the following and, this way, a concrete definition of CTPQSs for spin-1/2 models is provided:

Let $\{|k\rangle\}_{k=1}^{2^N}$ be an orthonormal basis of the Hilbert space \mathcal{H} and $|\Psi\rangle$ a pure state as before in Eq. (8.37). Let

$$\langle A \rangle^{CTPQS} = \langle \Psi|A|\Psi \rangle \quad (8.39)$$

be the expectation value of such a pure state with respect to a quantum physical operator A and

$$\langle A \rangle^{ens} = \frac{\text{Tr}[Ae^{-\beta H}]}{Z(\beta, H)} \quad (8.40)$$

its ensemble average. If $\forall \epsilon > 0$ a function $\eta_N(\epsilon)$ exists with $\eta_N(\epsilon) \xrightarrow{N \rightarrow \infty} 0$ such that the probability

$$P\left(\left|\langle A \rangle^{CTPQS} - \langle A \rangle^{ens}\right| \geq \epsilon\right) \leq \eta_N(\epsilon) , \quad (8.41)$$

then $|\Psi\rangle$ is called a CTPQS [196]. The right hand side is bounded from above for a Hilbert space \mathcal{H} of dimension 2^N by

$$\eta_N(\epsilon) \leq \frac{L^{2m} e^{-\mathcal{O}(N)}}{\epsilon^2} , \quad (8.42)$$

where the operator A is assumed to be a low-degree polynomial (degree $\leq m$) of local operators [196, 209]. This means $\eta_N(\epsilon)$ is decreasing exponentially in system size N . A similar result is obtained in Ref. [79] and even extended to dynamical correlation functions [210]. Therefore, it is for large system sizes sufficient to consider only one or at least just a few pure states.

8.3.5 Variational Canonical Thermal Pure Quantum State (VCTPQS)

The question remaining is how the aforementioned CTPQS can be accurately approximated in a variational setting. While the imaginary-time evolution is of no concern, the construction of the initial infinite-temperature TPQS is not as evident as it may look on first sight. In Ref. [211], it was investigated thoroughly which minimum requirements the randomly chosen wave function components of the state $|\Psi(\beta = 0)\rangle$ need to fulfill for the state to be an infinite-temperature TPQS resulting in the following two constraints:

- The components c_k are uncorrelated.

$$\overline{c_j^* c_k} = \overline{|c_j|^2} \delta_{j,k} \quad (8.43)$$

- The mixed state density matrix ρ has low purity

$$\text{tr } \rho^2 = \sum_k \rho_k^2 \ll 1 , \quad (8.44)$$

where

$$\rho_k = |c_k|^2 / \sqrt{\sum |c_k|^2} . \quad (8.45)$$

This constraint guarantees that averages are not dominated by just a few large ρ_k .

Some probability distributions like the normal and the uniform distribution, which fulfill these two

conditions, have been discussed in Refs. [196, 209]. Results for normally distributed wave-function components were obtained within the finite-temperature Lanczos method in Refs. [212, 213, 214] which include a detailed analysis of the accuracy of this method. In this work, we are, however, interested in using a variational ansatz like the NN ansatz to compute expectation values of operators at finite temperatures. Unfortunately, a FFNN ansatz in general and the NN ansatz in Eq. (7.2) in particular are problematic starting points. The non-linear activation function leads, in general, to a small number of amplitudes $|\Psi_{\mathbf{w}}(\mathbf{s})|$ dominating the wave function, thus violating the second condition stated above for a typical state. Choosing small initial network parameters leads on the other hand to a violation of the first condition. Additionally, it was proven in Ref. [215] and discussed in Refs. [216, 217] that infinite-temperature CTPQS have volume-law entanglement ruling out the METS-NN as possible candidates.

8.3.6 PP-NN Ansatz

Keeping in mind that the two conditions in Eq. (8.43) and Eq (8.44) are desired to be fulfilled, the PP-NN ansatz in Eq. (6.33) seems tempting. If one chooses the initial parameters of the NN, to be zero, or at least for numerical convenience so small that the variance of the NN part of the wave function is small compared to the one of the PP part, the infinite-temperature state is solely described by the PP part in Eq. (6.21). The correlation of the wave functions of two different spin configurations in the PP part is computed to be $\overline{\Phi_{PP}(\boldsymbol{\sigma}_j)\Phi_{PP}(\boldsymbol{\sigma}_k)} = 0$, since two configurations differ in every term of the Pfaffian by at least one random variable $X_{jk}^{\sigma_j\sigma_k}$. As long as the number of parameters is smaller than the Hilbert space dimension, non-zero higher order correlations will exist. However, we anticipate that such correlations have no influence on the expectation values of local observables. I.e., we expect that the difference in expectation values obtained by an actual CTPQS, given, for example, by normally distributed c_k and those obtained by the variational PP-NN ansatz will be small if a sufficient number of parameters is kept. To allow to distinguish the influence of the PP from the NN part, the PP-local ansatz is defined as the one for which the correlation parameters F_{ij} in Eq. (6.16) are only non-zero if $|i - j| \leq r$ with $r = 1$. In general, the effect of the PP part can be adjusted by varying the parameter r .

8.3.7 The Infinite-Temperature VCTPQS

First, we check if the states chosen according to the PP-NN ansatz, see Eq. (6.33), are good approximations of an infinite-temperature VCTPQS. As comparison, the normally distributed states chosen according to Eq. (8.37) and the METS-NN states chosen according to the NN ansatz, see Eqs. (8.20)-(8.22), are used. Second, the entanglement entropy is also computed as a measure for how elaborate it is to approximate those states by a variational ansatz (see also Sec. 5.10).

Independence of Components

It is not a priori clear to which extent the conditions for a infinite-temperature VCTPQS, discussed in Sec. 8.3.5, are fulfilled by the PP-NN ansatz. Additionally, apart from the entanglement entropy, it is not clear why the METS-NN states do not describe an infinite-temperature VCTPQS. We investigate first to what extent Eq. (8.43) is fulfilled compared to a CTPQS with normally distributed random numbers for the wave function components - from here on referred to as the Gaussian case. It is emphasized that in the PP-NN case the real and imaginary parts of all the NN parameters are not chosen as described in Sec. 8.2.1 but from a normal distribution centered around zero with $\sigma_{NN} = 1/N$ while those of the PP parameters are always chosen from a normal distribution centered around zero with $\sigma_{PP} = (N/2)! \cdot \frac{2^{N/2}}{(N!)^{2/N}}$ to keep the variance of both wave function components constant for growing system size. This way, the NN parameters in the PP-NN ansatz do not contribute to the properties of the infinite-temperature state but are only relevant for the numerics of the imaginary-time evolution. Thus, those parameters can be neglected in the following discussion. To check the validity of Eq. (8.43), the square root of the mean of the absolute value of the pairwise Pearson correlation coefficients

$$\rho_{jk} = \frac{\text{Cov}[\Psi_{\mathbb{W}}(\mathbf{s}_j), \Psi_{\mathbb{W}}(\mathbf{s}_k)]}{\sigma_{\Psi_{\mathbb{W}}(\mathbf{s}_j)}\sigma_{\Psi_{\mathbb{W}}(\mathbf{s}_k)}} \quad (8.46)$$

are computed, where $\text{Cov}[\cdot]$ denotes the covariance, \mathbf{s}_j the j -th s^z -configuration and σ the standard deviation. In Fig. 8.12, the Pearson correlation coefficients are plotted for varying n denoting the number of sets of parameters used. The full Hilbert space is considered for a spin chain of length $N = 12$. The diagonal elements are of course excluded. The results for the Gaussian case show as expected a $n^{-1/2}$ behavior as is checked by a respective fit. While this also holds for the PP ansatz up to an overall constant, the PP-local and the METS-NN ansatz differ significantly from this behavior.

Distribution of Expectation Values

To check what this means for the distribution of expectation values, the standard deviation of $\langle S_0^z \rangle$ is computed for $n = 10^3$ sets of parameters \mathbb{W} and varying length N in Fig. 8.13. The expectation values are obtained for every set of parameters by using 10^4 Metropolis samples. Since the Gaussian case reaches values affected by the sampling error, the full Hilbert space is used. While the Gaussian case shows exponential decay to the infinite-temperature result, the data for the PP ansatz can be fitted very well by a function that is proportional to $e^{-\lambda\sqrt{N}}$, where λ is a fitting parameter. However, the METS-NN and the PP-local ansatz show no decay with system size at all since they are product states and thus only depend on the parameters connected to the 1st spin. This is as well as the result for the Gaussian case analytically verified by computing the probability distribution of $\langle S_0^z \rangle$ which is in all those three cases given by

$$2\langle S_0^z \rangle = \frac{A}{A+B} - \frac{B}{A+B}, \quad (8.47)$$

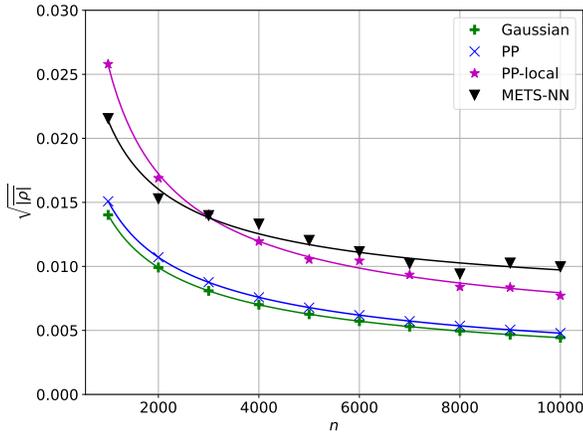


Figure 8.12: Square root of the mean of the absolute of the pairwise Pearson correlation coefficients of the wave function components for n different sets of parameters and a spin chain of length $N = 12$ [80].

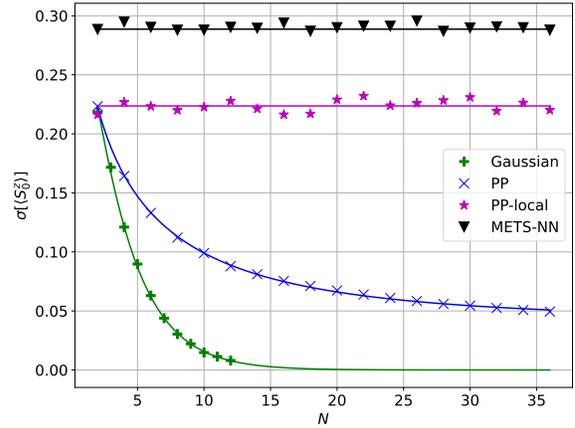


Figure 8.13: Standard deviation of the operator $\langle S_0^z \rangle$ computed using 10^4 Metropolis samples and $n = 10^3$ different sets of parameters chosen as discussed above [80]. For the Gaussian case, the full Hilbert space is considered.

where

$$A = \sum_k^{K/2} |z_k^\uparrow|^2 \quad B = \sum_k^{K/2} |z_k^\downarrow|^2 \quad (8.48)$$

and the real and imaginary part of z_k^\uparrow and z_k^\downarrow are chosen randomly from a normal distribution. The actual variance of this normal distribution is irrelevant in this case because it cancels out and is thus chosen to be one. Therefore, A and B follow a χ_K^2 distribution. Each term in Eq.(8.47) is computed to follow a probability distribution $p(y)$ with $y \in [0, 1]$. These, the respective value for K and the resulting variances are noted in Tab. 8.1 and plotted as solid lines in Fig. 8.13. The

Ansatz	K	$p(y)$	$\sigma[\langle S_0^z \rangle]$
Gaussian	2^N	$\frac{\Gamma(2^N)}{\Gamma^2(2^{N-1})} y^{2^{N-1}-1} (1-y)^{2^{N-1}-1}$	$\sqrt{\frac{2^{N-1}+1}{2^{N+1}+2} - \frac{1}{4}}$
PP-local	4	$6y(1-y)$	$\frac{1}{\sqrt{20}}$
METS-NN	2	1	$\frac{1}{\sqrt{12}}$

Table 8.1: Probability distributions and variances for the different ansatz states [80].

promising relations for the Gaussian case in Eqs. (8.39) and (8.41) thus do not translate to the discussed variational wave functions with a finite number of parameters. While the exponential decay is weakened for the PP ansatz, there is no decay at all with the METS-NN and the PP-local ansatz. For operators like the Hamiltonian in Eq. (2.16), there is still a decay of the standard deviation in all cases for larger spin chains observed. This is due to the benefit of averaging over local spin-spin operators along the chain. Naturally, the results for the mean of the expectation value converge in

all cases to the infinite-temperature value with increasing number of parameter sets (see Ref. [203] and the discussion in Sec. 8.2.1).

Entanglement Entropy

A significant difference between the ansatz wave functions becomes apparent when examining the entanglement entropy in Eq. (5.32), where the system is divided equally in the middle into the systems A and B . While the PP-NN ansatz shows a volume-law scaling due to the PP part of the wave function, the METS-NN ansatz shows no entanglement because the wave function is a product state. The different scaling with the length of the chain is illustrated for the full Hilbert space in Fig. 8.14. This is very relevant since states with volume-law entanglement can in general not be efficiently represented. The RBM is only capable of doing so for some specific states [142]. Thus, the METS-NN ansatz defines an initial state that is much easier to approximate. To allow for a similar state in the PP-NN ansatz, the PP-NN-local ansatz is used. This ansatz shows an entanglement entropy that is only nonzero if there is a PP pair that crosses the cut in the middle of the spin chain. This is only the case for chain lengths for which $N/2$ is odd. By increasing the support of the PP ansatz, one can obtain states whose properties lie in between the blue and magenta curves in Figs. 8.12, 8.13 and 8.14.

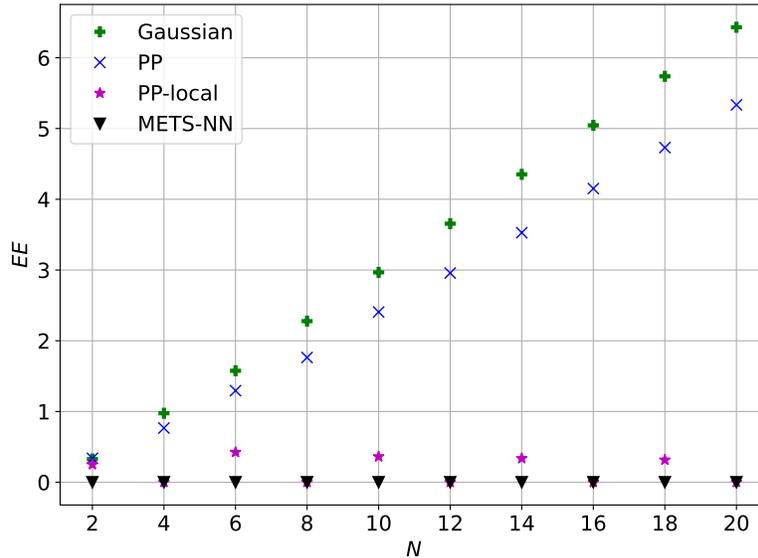


Figure 8.14: Entanglement entropy of the PP, the PP-local and the METS-NN initial states compared to the Gaussian case for different chain lengths of an open Heisenberg chain [80].

8.3.8 Numerical Results

Due to the results obtained in the previous section, the states obtained via the PP-NN ansatz are considered VCTPQSs. To compare the PP-NN and the PP-NN-local ansatz to the previous methods, we average in each case over 30 initial states and time evolve these according to Eq. (8.38). The averaging is still useful due to the fact that the error is dominated by the sampling error (see Fig. 8.13). The finite-temperature results are again computed for a Heisenberg chain of length $N = 30$. The results are shown in Fig. 8.15 and compared to the same QMC results as before. The number of samples and the number of iteration steps R is adjusted due to the larger computational cost for the updates of the PP parameters. For the computation of the Pfaffian after a spin flip, the effective update rule based on Cayley's identity, which was explained in Sec 6.3.4, is used [10].

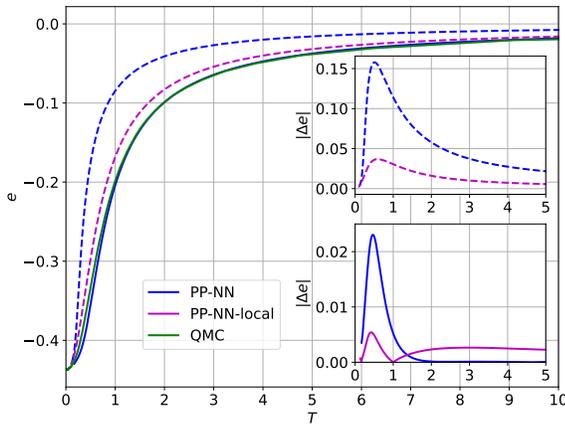


Figure 8.15: Inner energy of the PP-NN and the PP-NN-local ansatz compared to the QMC solution for an open Heisenberg chain of length $N = 30$ obtained using the typicality algorithm with 10^4 Metropolis samples, $R = 10^3$ and $\epsilon = 10^{-6}$ [80]. The average is taken over 30 initial states. The dashed lines show the data without β modification. The PP-NN-local results with β modification are omitted in the main plot because they are not distinguishable from the QMC results on this scale.

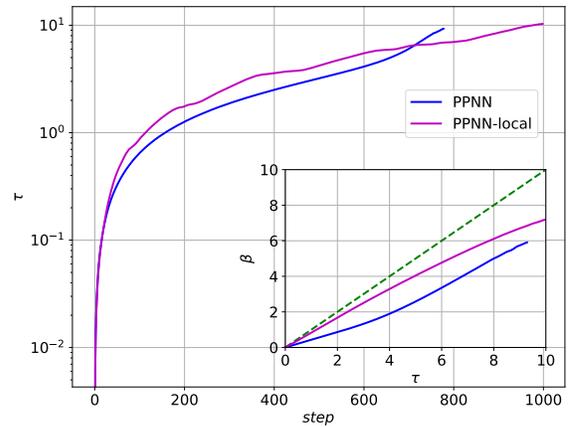


Figure 8.16: Imaginary time τ in the Heun algorithm versus number of iterative steps for the PP-NN and the PP-NN-local ansatz for an open Heisenberg chain of length $N = 30$ using the typicality algorithm and the parameters defined in Fig. 8.15 [80]. The β modification is shown in the inset.

The energy values for high temperatures are for the PP-NN ansatz more accurate, as is expected from the discussion in Sec. 8.3.7. However, during the course of the evolution, the PP-NN ansatz states tend to deviate further from the QMC results than in the METS-NN ansatz in the sampling algorithm in Sec. 8.2.2. The maximum deviation is roughly at the point where the heat capacity is maximal. The most elaborate ansatz with the largest number of parameters thus produces the most inaccurate results. This is a consequence of the large entanglement of the initial states $|\Psi(0)\rangle$ in the PP-NN ansatz. Therefore, it is more difficult to approximate the typical states, although the number of parameters in the PP-NN ansatz is much larger compared to the METS-NN ansatz. It is

stressed that this is not due to the smaller number of Metropolis samples, as is evident from the comparison to the PP-NN-local ansatz. Summarizing, there is not much of a numerical benefit of using VCTPQs in the PP-NN ansatz. The theoretical benefit is eaten up by the larger entanglement entropy, leading to a trajectory in the variational manifold that includes states that are harder to represent. It might be possible to find an ansatz that can represent the highly entangled states on the trajectory in special cases, for example for a certain model. While it seems unlikely that this can be done in general, deep NNs (DNNs) could possibly lead to a better representation of such states and thus alleviate this issue to some degree [155].

The modification of the imaginary time shown in Fig. 8.16 illustrates the issue with the VCTPQS in a different way. Especially for high temperatures, the imaginary time evolution in the PP-NN ansatz shows a very large β modification, which means that the proper evolution and the evolution projected onto the variational manifold deviate strongly. Although this feature turns somewhat around for lower temperatures, the PP-NN ansatz states appear to be not a good starting point due to their large initial entanglement. On the other hand, the error of the time evolution after the β modification is applied to the PP-NN-local states is of the order 10^{-3} which is thus again consistent with the Monte Carlo sampling error. It is stressed that within the PP-NN ansatz an accurate approximation of the ground state of the one-dimensional Heisenberg model is possible. Even the computation of the ground state phase diagram of the two-dimensional J_1 - J_2 model, which is depicted in Fig. 2.3, is achieved in Ref. [5]. The quantum number projection is additionally used, as in any ground state optimization method, to allow for an accurate computation of the results.

Chapter 9

Dynamical Correlation Functions for the One-Dimensional Heisenberg Model at Infinite Temperature

Besides the imaginary-time evolution, the question of how accurate the real-time evolution of a certain quantum state at finite temperature can be simulated is of natural interest closely related to experiments with ultracold atoms. Given the possible numerical methods available, it is even of outstanding interest since the QMC is due to the sign problem not applicable at all, while the time-dependent DMRG shows an exponential increase in the bond dimension with the entanglement entropy [150, 192]. Hence, it is briefly investigated in this chapter.

The expressive power of NNs, on the other hand, has been discussed extensively in this work. It started from principal, general theorems of real-valued NNs in Ch. 5 and resulted in the actual ability of efficiently representing quantum states with volume-law entanglement in Sec. 5.10 and Ch. 7. We remind the reader that special volume-law entangled states have been proven in Ref. [142] to be represented in the RBM or NN ansatz in Eq. (7.2) using $M \sim N$ hidden units (see also Ref. [9]). In their vicinity, one should thus be able to approximate similar states accurately. However, to be able to describe the dynamics accurately, this needs to be true for a variety of states, not only for some specific states and those in their vicinity. The algorithm for the results presented in this chapter was developed by Jesko Sirker. It is noted that the results have been published already in Ref. [80] (Wagner, Klümper, Sirker).

9.1 Spin-Spin Autocorrelation Function and Entanglement Entropy

The aim in the following is to conduct a numerical investigation to determine the number of hidden units necessary to achieve accurate results for the spin-spin autocorrelation function

$$\langle S_n^z(0)S_n^z(t) \rangle = \frac{1}{Z} \text{Tr} \left\{ S_n^z e^{itH} S_n^z e^{-itH} \right\} \quad (9.1)$$

at site n of an open Heisenberg chain at infinite temperature using the purification algorithm explained in Sec. 8.1.1. In addition, the entanglement entropy

$$S = -\text{Tr} \rho_A \ln \rho_A \quad (9.2)$$

is computed, where ρ_A is the reduced density matrix for a subsystem A consisting of half of the chain of real spins and ancilla spins.

9.2 Numerical Results

The full Hilbert space is considered in the following computations for small system sizes. The results for $N = 4$, depicted in Fig. 9.1, show that at least $M = 16$ hidden nodes are required to accurately approximate the exact result.

Looking at the entanglement entropy in the lower panel, one can conclude that while $M = 8$ is sufficient to represent certain states with volume-law entanglement, it does not capture the full entanglement structure of the time evolved state for times $Jt \gtrsim 2$.

The same computation for $N = 6$ in Fig. 9.2 proves that the number of hidden nodes required increases with system size. At least $M = 96$ hidden units are necessary to obtain accurate results for $Jt \gtrsim 5$. It is stressed, that already for $M = 12, 24, 48$ the entanglement entropy reaches values at long times of the order of the maximum entanglement in the time evolved state. The variational ansatz with $M = 12, 24, 48$ is thus capable of approximating states with large entanglement, but apparently fails to do so for all states on the trajectory. Therefore, it is not sufficient to describe the complete dynamics. Finally, in Fig. 9.3, results for $N = 8$ are presented, with the computation of the entanglement entropy omitted due to the numerical complexity involved. It is stressed that the combined system consists already of 16 sites. The number of hidden nodes required rises in this case to more than $M = 256$.

In summary, the number of hidden units M required to accurately simulate the infinite-temperature autocorrelation in Eq. (9.1) increases much faster than linear. Assuming that $M = 512$ hidden units yield reasonably converged results for $N = 8$, the increase follows, in the best case scenario, a power law $M \sim N^\alpha$ with $\alpha \gtrsim 4$. However, it is more consistent with an exponential scaling. While the number of states kept in a time-dependent DMRG calculation is only limited by the available memory and the compute time to perform a single-value decomposition or a diagonalization allowing to reach truncated dimensions of order $\chi \sim 10^5 - 10^6$, time-dependent neural networks for

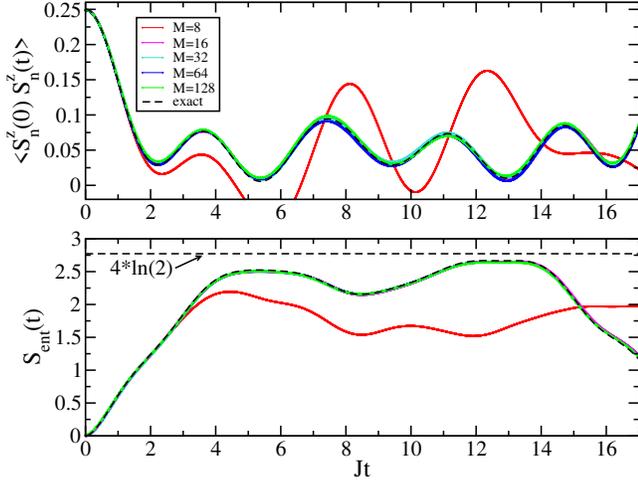


Figure 9.1: Top: Autocorrelation function $\langle S_n^z(0) S_n^z(t) \rangle$ for $n = N/2$ at infinite temperatures for an open Heisenberg chain with $N = 4$ sites. Bottom: Corresponding entanglement entropy for cutting the chain of real and ancilla spins into two equal halves. The RBM results for different M are compared to the exact result [80].

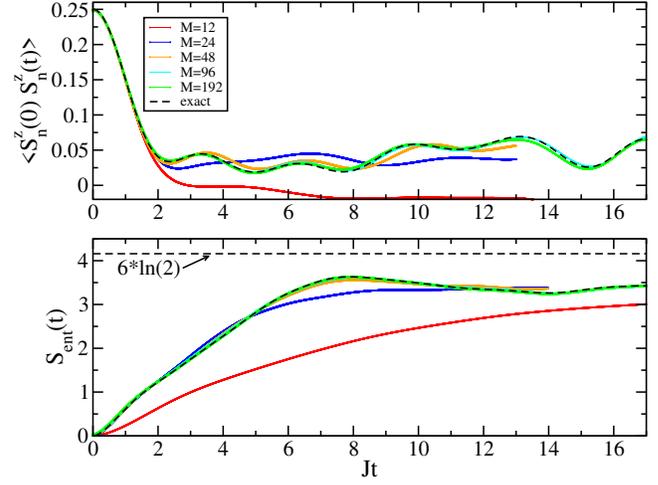


Figure 9.2: Same as Fig. 9.1 for a chain with $N = 6$ sites. The results for $M = 24, 48$ are only shown up to $Jt \sim 13$ [80].

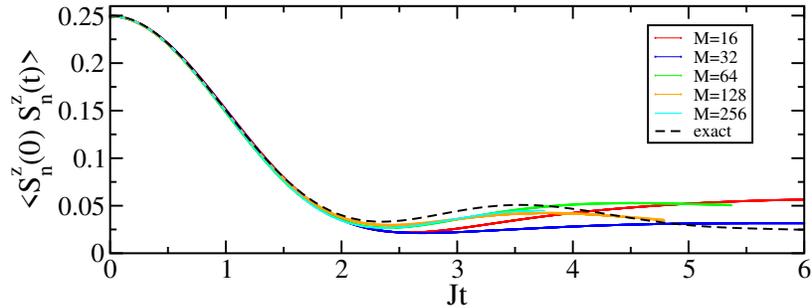


Figure 9.3: Autocorrelation function $\langle S_n^z(0) S_n^z(t) \rangle$ for $n = N/2$ at infinite temperatures for an open Heisenberg chain with $N = 8$ sites [80]. The corresponding entanglement entropies have not been calculated.

$N \sim 10^2$ with $M \sim 10^8$ hidden units might be difficult to optimize given that the Hilbert space needs to be sampled. It is also noted that in the DMRG the dimension χ can be adaptively increased with simulation time.

While a potential power-law increase in the hidden units with system size would make this approach in theory more powerful than time-dependent DMRG, it might be very difficult in practice to benefit from this scaling.

Furthermore, the NN results in Figs. 9.1, 9.2, 9.3 show, on the other hand, that the calculation becomes inaccurate already at short times if an insufficient number of hidden units is used. Additionally, there does not seem to be an easy way to adaptively increase the number of hidden units during a calculation. Finally, it is noted that the strong, likely exponential increase of the number of required hidden units is consistent with a recent publication where similar behavior was observed for quench dynamics [218]. It is emphasized that the β modification in Sec. 4.4 is also applicable for the real-time evolution. It can thus, in principle, improve the results obtained in this chapter. However, the unfavourable scaling is expected to prevail.

Chapter 10

Finite-Temperature Results for the Two-Dimensional J_1 - J_2 Model on the Square Lattice

A thermal Ising phase transition in the J_1 - J_2 model (see Sec. 2.3) in two dimensions, was first predicted analytically in Ref. [57]. However, numerical results are rare. Its classical counterpart, the J_1 - J_2 Ising model has been investigated thoroughly [219]. Numerical results are in this case also obtained by Monte Carlo methods and verify the existence of an Ising transition at finite temperatures [220]. On the other hand, the ground state phase diagram of the J_1 - J_2 model is well studied with some major disagreements remaining in the midsection around $J_2 \approx 0.5$. Since the finite-temperature phase transition is expected to occur at low temperatures, the ground state phase diagram is discussed first. Secondly, the computation of finite-temperature results via the purification and the sampling method is investigated for the two-dimensional Heisenberg model ($J_2 = 0$) in comparison to the results for the one-dimensional Heisenberg model in Ch. 8. Finally, the finite-temperature results for the frustrated J_1 - J_2 model ($J_1 = 1$, $J_2 > 0$) are compared to results from exact diagonalization (ED) before the manifestation of the phase transition at finite-temperatures is investigated. While J_2 is varied, the parameter $J_1 = 1$ is chosen to stay constant throughout this chapter.

10.1 Ground State Phase Diagram

The nature of the ground state of the J_1 - J_2 model for $J_1 = 1$ and variable J_2 has been studied extensively [5]. For small J_2 , a Néel-type phase exists, while for large J_2 , a stripe-type antiferromagnetic phase, also known as the collinear phase, is present. This is depicted in Fig. 2.3. The phase around $J_2 \approx 0.5$ has been investigated by various numerical techniques leading to different results ranging from a quantum spin liquid that is ungapped (variational method) [221] or gapped (DMRG) [222] to a valence bond solid of collinear [223] or plaquette type [224]. The term gap refers to the excitation spectrum. The most recent study in Ref. [5] makes use of the PP-NN

ansatz, which was discussed in Sec. 6.5. The authors therein managed to compute the result in Fig. 2.3. In Ref. [225], a highly sophisticated CNN was constructed, yet it did not yield conclusive results regarding the stability of the spin liquid phase. Summarizing, the nature of the phase between $J_2 \approx 0.5$ and $J_2 \approx 0.6$ is still a matter of debate and has not been consistently established.

10.2 Modifications of the Purification and the Sampling Algorithm

The primary generalization of the NN wave function in Eq. (7.2) to two-dimensional models is straightforward, as the input and hidden nodes of the NN are fully connected. Thus, for the NN ansatz wave function, the arrangement of the input s^z in either a chain or a square lattice is immaterial. Only the computation of correlation functions, as for example the energy, needs to be adjusted. Hence, the sampling algorithm is adapted with minor modifications.

The ancilla spins in the purification method, on the other hand, are introduced by an additional ancilla square lattice of the same size as the lattice of real spins. The infinite-temperature state in the purification method in Eq. (8.12) is especially unfavourable for Monte Carlo sampling methods due to its binary distribution assigning probabilities of either 0 or 1 to the s^z -configurations [195, 88]. This issue is amplified for the J_1 - J_2 model in two dimensions inhibiting a functioning algorithm. The value of the parameter c can not be increased to a value that allows for the correct infinite-temperature state to be approximated accurately, while at the same time allowing for an accurate evolution of the state in imaginary time τ . Luckily, due to the ambiguity of the purification with respect to a rotation of the ancilla spins, governed by the unitary matrix U in Eq. (8.8), the binary distribution is easily altered to a more favourable one for an efficient Monte Carlo sampling. One well-known transformation in this field is known as the Hadamard transform

$$H_{Had} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad (10.1)$$

which is in this case only applied to the ancilla spins by defining the following operator

$$T_{Had} = \left(\bigotimes_{i=1}^N \mathbf{1} \right) \otimes \left(\bigotimes_{i=1}^N H_{Had} \right). \quad (10.2)$$

It is noted that this corresponds to a change of basis of the ancilla spins from the s^z -basis to the s^x -basis. The idea of using the Hadamard transform was first proposed in Ref. [194]. The probabilities of the s^z -configurations are after the application of T_{Had} onto the correct infinite-temperature state all equal to 1, while the wave function components differ by their sign only. In the Monte Carlo sampling, the rejection probability is then substantially decreased, since all states are now equally probable and the failure of ergodicity, shining through for too large initial values of c in Sec. 8.1.3, is cured. The direct application of the operator in Eq. (10.2) onto the infinite-temperature state in Eq. (8.12), which was used for the one-dimensional Heisenberg model, comes with an additional cost for the computation of the matrix elements of physical operators. The application of this operator

onto the correct infinite-temperature state in Eq. (8.7), on the other hand, is simply expressed by the Hadamard factor

$$F_{Had}(\mathbf{s}) = \frac{1}{2^{\frac{N}{4}}} (-1)^{\mathbf{s}_R \cdot \mathbf{s}_A} . \quad (10.3)$$

The NN ansatz is added to allow for an accurate evolution of the variational ansatz wave function

$$\Psi_{\mathbb{W}}(\mathbf{s}) = \exp \left(- \sum_{n=1}^{2N} a_n s_n^z \right) \prod_{m=1}^M \cosh \left(\sum_{n=1}^{2N} W_{mn} s_n^z + b_m \right) F_{Had}(\mathbf{s}) , \quad (10.4)$$

where \mathbf{s}_R and \mathbf{s}_A are the vectors of the s^z -configurations of the real and the ancilla spins, respectively. The correct infinite-temperature state is obtained by choosing $\mathbf{a} = \mathbf{b} = \mathbf{W} = 0$. For the numerical computations, small noise needs to be added to the parameters governed by the variance of the normal distribution $w \sim 0.1$ the parameters are chosen from. The results are similar for a wide range of $w \in [10^{-3}, 0.1]$. It is stressed that the ansatz in Eq. (10.4) is not the Hadamard transform of the modified NN ansatz in Eq. (8.12), but the Hadamard transform of the exact infinite-temperature state adjusted by a small noise component described by the NN ansatz with all variational parameters chosen to be small. The correct infinite-temperature state can in both cases be approximated to an arbitrary extent by choosing the variational parameters accordingly. In the imaginary-time evolution, the parameters in the NN part of the variational ansatz in Eq. (10.4) change differently from the ones in the NN part in Eq. (8.12).

10.3 Finite-Temperature Results

Following the discussion of the finite-temperature results for the one-dimensional Heisenberg model in Ch. 8, there remain two methods available for the computation of finite-temperature results in two dimensions: the purification and the sampling method. With the slight modification of the infinite-temperature state described in Sec. 10.2, the purification and the sampling algorithm are directly applicable to two-dimensional models. To check the validity of the adjusted purification algorithm and the difference to the one-dimensional case in general, the finite-temperature results are first computed for the two-dimensional Heisenberg model analogous to the computations in the one-dimensional case in Ch. 8. After that, the algorithms are applied to the J_1 - J_2 model with $J_2 = 0.85$ to allow for a comparison to the results obtained with the CTMRG method in Fig. 10.7. New Monte Carlo samples are now proposed by either flipping a spin or exchanging the position of two randomly chosen spins. The reason for this will become clear in Sec. 10.5.

10.3.1 Finite-Temperature Results for the Two-Dimensional Heisenberg Model

To verify the validity of the algorithms at hand, the Heisenberg model (J_1 - J_2 model with $J_2 = 0$) is considered. In Fig. 10.1, the inner and free energy results are depicted for a periodic Heisenberg model on the square lattice with $N = 4 \times 4$ sites using the purification and the sampling algorithm.

They are compared to ED results computed with the *QuSpin* package explained in Ref. [226]. The energy results are in both methods comparable to an accuracy of order $\sim 10^{-3}$ up to a small deviation at low temperatures. The modified β in Fig. 10.2 tends to converge earlier compared to the results for the one-dimensional Heisenberg model in Fig. 8.6 and 8.9. This is in accordance with similar results for the sampling method for the Heisenberg model in Ref. [90]. The lowest

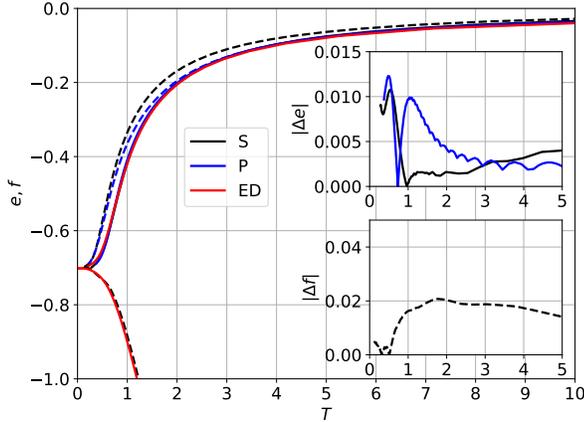


Figure 10.1: Inner and free energy for the periodic Heisenberg model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm P ($\epsilon = 10^{-6}$) and the sampling algorithm S ($\epsilon = 10^{-5}$) with $4 \cdot 10^4$ samples and $\alpha = \frac{M}{N} = 2$. The ED results are obtained using the *QuSpin* package [226]. For the free energy, the dashed lines denote the results without the β modification. The difference between the results for the energy and the free energy is depicted in the inset.

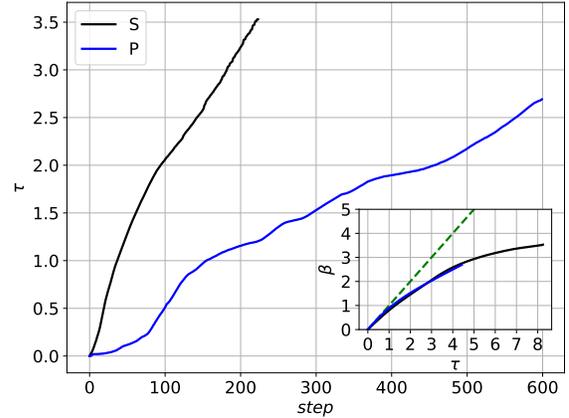


Figure 10.2: Imaginary time τ in the Heun algorithm versus number of iterative steps for the periodic Heisenberg model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm P and the sampling algorithm S . The parameters are the same as in Fig. 10.1. The β modification is depicted in the inset.

temperature reached is similar for the considered methods at hand. In both cases, the algorithm breaks down at those temperatures (see Fig. 10.8). Unfortunately, the minimal temperature reached increases with system size and thus prevents meaningful results at low temperatures. This is in accordance with more detailed studies of the critical slowing down, for example in Ref. [227]. It is stressed that this issue can neither significantly be overcome by an increase in the number of Monte Carlo samples nor an increase in the number of hidden units in the variational ansatz, since the failure of the algorithm at low temperatures can be attributed to a failure of the Monte Carlo sampling itself. This is in detail studied and explained in Sec. 10.5. It is less pronounced in the sampling method due to the smaller Hilbert space. Furthermore, the results for the free energy in the sampling method are of similar accuracy as in the one-dimensional case (see Fig. 8.8). It seems unclear what causes the strong deviation of the free energy in the purification method, which is thus not shown in Fig. 10.1. However, the consistency of the finite-temperature results for the energy and the free energy in the sampling algorithm let us conclude that the sampling method is also applicable in two dimensions. This only holds with respect to the energy and probably other correlation functions in the purification algorithm.

10.3.2 Finite-Temperature Results for the Two-Dimensional J_1 - J_2 Model

The consideration of not only nearest but also next-nearest neighbors in the J_1 - J_2 model comes with a significant increase in the computation time for the energy and $H_{\Psi_{\text{opt}}}^2$, leading to this part dominating for smaller system sizes. Other than that, the purification and the sampling algorithm used for the two-dimensional Heisenberg model are directly applied to the J_1 - J_2 model with $J_2 = 0.85$. The results for the energy and the free energy are depicted in Fig. 10.3. Both algorithms show a greater

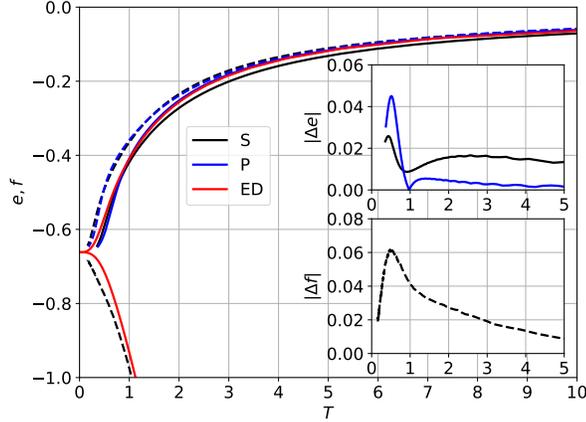


Figure 10.3: Inner and free energy for the periodic J_1 - J_2 model on the square lattice with $J_2 = 0.85$ and $N = 4 \times 4$ sites obtained using the purification algorithm P ($\epsilon = 10^{-6}$) and the sampling algorithm S ($\epsilon = 10^{-5.4}$), $4 \cdot 10^4$ samples and $\alpha = 2$. The ED results are obtained using the *QuSpin* package [226]. The dashed lines denote the results without the respective β modification. The difference between the results for the energy and the free energy is depicted in the inset.

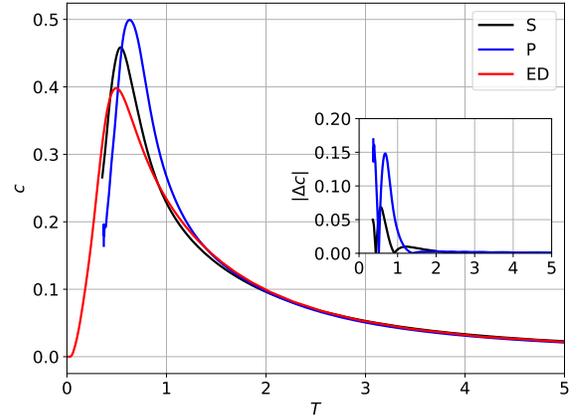


Figure 10.4: Specific heat capacity for the periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites and $J_2 = 0.85$ obtained using the purification algorithm P and the sampling algorithm S . The parameters are the same as in Fig. 10.3.

deviation from the ED results compared to the results for the two-dimensional Heisenberg model. In the sampling algorithm, the energy is less accurate compared to the results in Fig. 10.1. This is already the case for larger temperatures. However, it is stressed that the infinite-temperature energy is still approximated with an accuracy that is dominated by the Monte Carlo sampling error. The imaginary-time evolution in the sampling algorithm at large temperatures is apparently inaccurate and thus the reason for the deviation. In the purification algorithm, the energy is approximated accurately at larger temperatures, but deviates more strongly at low temperatures resulting in a large peak in the inset showing the difference to the ED results. The accuracy of the results should be improved by increasing the number of hidden units, the number of Monte Carlo samples and the desired accuracy of the imaginary-time evolution ϵ (see Sec. 4.3). Unfortunately, significant improvement couldn't be attained by adjusting any of these outer parameters individually. It seems that one either needs to adjust all of these parameters collectively and substantially or even reduce

the complexity of the approximation by using, for example, the translational symmetry. Additionally, the specific heat capacity is considered and depicted in Fig. 10.4. While the sampling algorithm basically only fails to accurately approximate the magnitude of the peak of the specific heat capacity, the peak of the purification algorithm is shifted. Thus, the sampling algorithm is the more reliable method to investigate the presumed phase transition in the following. However, the reduced accuracy of the finite-temperature results compared to the ones in the two-dimensional Heisenberg model need to be kept in mind when considering system sizes for which no other results for comparison are available. It is noted that this difference between the two-dimensional Heisenberg model and the J_1 - J_2 model, especially for the results for the specific heat capacity, was found also in Ref. [195] for a reduced Hilbert space ($\sum_i S_i^z = 0$).

10.4 Phase Transition at Finite Temperatures

Unfortunately, the early convergence of the modified temperature in the two-dimensional case (see Fig. 10.2) renders the detection of the predicted finite-temperature phase transition of the J_1 - J_2 model on the square lattice difficult, especially, since quite recently, a tensor network ansatz has determined the critical temperature to be at $\frac{T_C}{J_1} = 0.093$ for $J_2 = 0.85$ [6]. Additionally, the algorithm even breaks down at temperatures, where the modified β curve is not yet finally converged. This is in detail discussed in Sec. 10.5.

Still, the behavior of the J_1 - J_2 model with $J_2 = 0.85$ is studied at low but finite temperatures. It is aimed to workout whether possible indicators for a phase transition are observable in the numerical results even for finite system sizes. Based on the finite-temperature results in Sec. 10.2, the sampling algorithm is selected as the preferred method. Following the discussion in Sec. 2.6 there is a variety of critical exponents to study. Fortunately, the constituents of the specific heat capacity $\langle \Delta H \rangle$ are due to the β modification computed in every imaginary-time step anyway. An accurate computation of the susceptibility (see Eq. (2.18)) demands a computation of the finite-temperature results for zero and small magnetic field $h_i = \pm h$. The following correlation function

$$o_{vh} = \frac{1}{N} \left(\sum_{\langle i,j \rangle, v} \mathbf{s}_i \mathbf{s}_j - \sum_{\langle i,j \rangle, h} \mathbf{s}_i \mathbf{s}_j \right) \quad (10.5)$$

allows for the detection of the stripe phase for $J_2 \gtrsim 0.61$ being maximal when ordering occurs in the vertical direction and minimal when ordering occurs in the horizontal direction. The sums in Eq. (10.5) are taken in horizontal (h) and vertical (v) direction, respectively. The Néel phase for $J_2 \lesssim 0.49$ is detected by computing all the next-neighbour spin interactions in the following way

$$o_{Néel} = \frac{1}{N} \left(\sum_{\langle i,j \rangle} (-1)^{i+j} \frac{\mathbf{s}_i \mathbf{s}_j}{N} \right). \quad (10.6)$$

This way, the stripe and the Néel phase are detected correctly in the respective J_2 sectors depicted in Fig. 2.3. To investigate the expected finite-temperature phase transition, the correlation function

$|o_{vh}|$ and the specific heat capacity are computed for $J_2 = 0.85$ and shown in Figs. 10.5 and 10.6. They are compared to the already mentioned results obtained with the corner transfer matrix

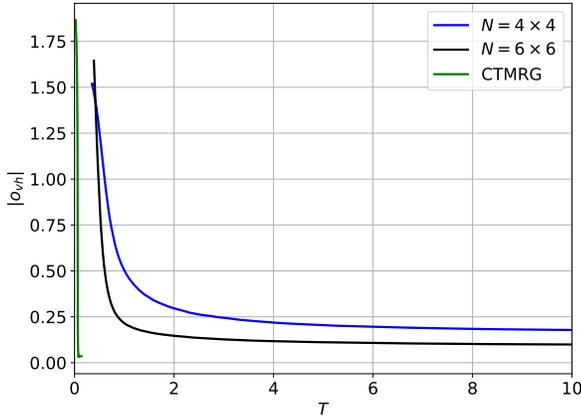


Figure 10.5: Absolute value of the correlation function o_{vh} for a periodic J_1 - J_2 model on the square lattice for various system sizes using the sampling algorithm with $J_2 = 0.85$, $\alpha = 2$, $4 \cdot 10^4$ samples and $\epsilon = 10^{-5.4}$. The CTMRG results are extracted for the corner dimension $\chi = 256$ from Fig. 10.7.

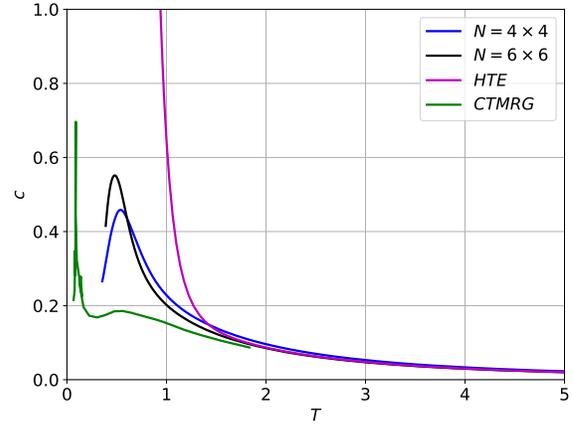


Figure 10.6: Specific heat capacity for a periodic J_1 - J_2 model on the square lattice with $J_2 = 0.85$ for various system sizes using the sampling algorithm. The parameters are the same as in Fig. 10.5. The coefficients for the high-temperature expansion (HTE) up to 8-th order are obtained from Ref. [228].

renormalization group (CTMRG) method in Fig. 10.7. These results are obtained directly in the thermodynamic limit. The formation of the phase transition for the system sizes considered is hardly detectable. On one hand, the significant rise of the correlation function $|o_{vh}|$ exhibits a more abrupt and rapid trend with expanding system size, potentially culminating in a similar abrupt surge detected by the CTMRG method. On the other hand, the results for the specific heat capacity already drastically deviate for lower temperatures, which is probably due to the finite size of the system. However, the results for the high-temperature expansion (HTE) up to 8-th order from Ref. [228] are consistent with the results obtained by the sampling algorithm for $N = 6 \times 6$ in the expected temperature range, while they deviate slightly from the CTMRG results. Since the authors of Ref. [6] themselves compare their results to the HTE, it is most probable that the graphical extraction comes, due to the small number of measurements for $T > 1$ and the rather large squares used in Fig. 10.7, with an error that can explain the slight difference for $T < 2$ until the HTE diverges. The critical temperature $T_C = 0.093$, according to the results obtained by the CTMRG method, is not reachable by the sampling algorithm due to the critical slowing down that seems to get even worse with larger system sizes. All in all, the results obtained with the sampling method do not show any sign of a phase transition because the system sizes investigated are too small and the temperatures reached are not low enough. To allow for more conclusive results, the imaginary-time evolution thus needs to reach lower temperatures. Additionally, an analysis of the correlation length is desirable to judge which system sizes are required to allow for results at low temperatures, which are not spoiled by finite-size effects. The results in Ref. [6] indicate that the correlation length is of the magnitude N already for temperatures $T \sim 2$. It is noted that an increase in the number

of samples and the number of hidden units (variational parameters) barely allows to reach lower temperatures. The reason for this is the critical slowing down of the Monte Carlo sampling, which is studied in detail in Sec. 10.5.

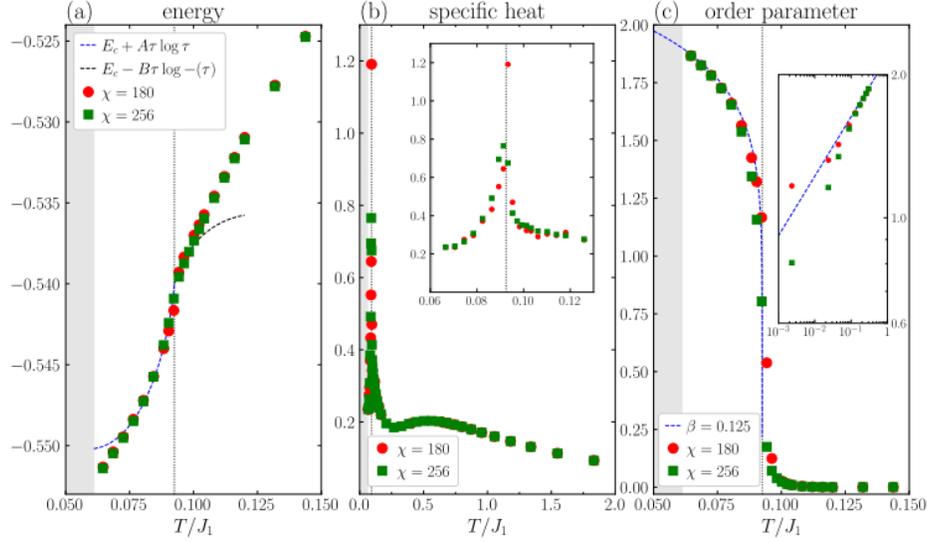


Figure 10.7: Results for the energy, the specific heat capacity and the correlation function o_{vh} computed by the corner transfer matrix renormalization group (CTMRG) algorithm with corner dimension χ for $J_2 = 0.85$. Reprinted with permission from Ref. [6].

10.5 Critical Slowing Down

The breakdown of the Monte Carlo sampling is depicted for the energy in Fig. 10.8. The reason can be attributed to what is known as the critical slowing down at low temperatures [229]. The Monte Carlo sampling error is determined by the number of statistically independent configurations that are used to estimate the expectation value of a physical operator. The number of configurations, which need to be sampled in a Markov chain until another independent configuration appears, is computed by the autocorrelation time [230]. The detection of the finite-temperature phase transition is undermined by long autocorrelation times in critical regions. This issue is quantified in the following. First, the estimator for the autocorrelation function of an operator O of interest

$$\Gamma_O(t) = \frac{1}{N_{\text{Met}} - t} \sum_{k=1}^{N_{\text{Met}} - t} f_O^k f_O^{k+t} \quad (10.7)$$

is computed, where $f_O^k = O_k - \langle O \rangle$ and O_k the expectation value of the operator O with respect to the configuration k in the sampling list. The autocorrelation function is obtained for $N_{\text{Met}} \rightarrow \infty$ [231]. The autocorrelation function thus serves to quantify the statistical error in the Monte Carlo sampling or, to put it different, the autocorrelation function helps to quantify how many Monte

Carlo samples are needed for an accurate computation of the quantity of interest. The expected exponential decay of the estimator of the autocorrelation function

$$\frac{\Gamma_O(t)}{\Gamma_O(0)} \propto e^{-\frac{t}{\tau}} \quad (10.8)$$

allows to obtain the exponential autocorrelation time τ by a corresponding fit [231]. The integrated autocorrelation time is summed up to be

$$\tau_{int} = \frac{1}{2} + \sum_{t=1}^W \frac{\Gamma_O(t)}{\Gamma_O(0)}, \quad (10.9)$$

where the factor $\frac{1}{2}$ is a matter of convention. This is in detail discussed in Ref. [232]. The choice of W is either computed therein in a more advanced way or by looking closely at the plot and choosing a point where the curve is sufficiently converged to zero. This is extensively discussed in Refs. [233, 234]. However, an accurate computation of the integrated autocorrelation time is here not of interest, but just a qualitative description of the issue at low temperatures. For this purpose, the parameter is chosen to be finite $W = 2000$. The critical slowing down of the Monte Carlo algorithm can be detected by an increase in the autocorrelation time. Finally, the integrated autocorrelation time can be used to compute an error estimate for the operator O

$$\sigma_O = \sqrt{2\tau_{int} \frac{Var[f_O^k]}{N_{Met}}}, \quad (10.10)$$

where $Var[f_O^k]$ the variance of f_O^k [231]. To clarify the critical slowing down at low temperatures, the numerical results for a periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites and $J_2 = 0, 0.85$ using the purification algorithm are discussed in detail. It is stressed that the issue is also present for the Heisenberg model, even in one dimension. However, it occurs already at higher temperatures for two-dimensional models. The breakdown of the algorithm is observed for the expectation values of physical operators, for example the energy, which is depicted in Fig. 10.8. This is directly linked to an increase of the rejection probability of the Monte Carlo sampling shown in Fig. 10.9. The integrated autocorrelation time is computed for the Hamiltonian and illustrated in Fig. 10.10. It is stressed that $1/\tau$ denotes the imaginary time in the purification algorithm, which can be used to compute the temperature via the β modification. In any other case in this section, τ refers to the (integrated) autocorrelation time. The integrated autocorrelation time is directly linked to the rejection probability. The breakdown of the algorithm is observed if the rejection probability and thus the autocorrelation time exceeds a critical value. In this event, the Monte Carlo sampling becomes trapped in certain states because they have a significantly higher probability $|\Psi_{\mathbb{w}}(\mathbf{s})|^2$ compared to those proposed in the subsequent sampling step. This prevents the Monte Carlo sampling algorithm to explore the space of possible configurations to a sufficient extent. It is only slightly refined by allowing the local updates to consist of either a spin flip or an exchange of positions of two randomly chosen spins. This issue has been recently studied extensively for the Potts-model in Ref. [227], where an exponential relation between the rejection probability and the integrated autocorrelation time $\tau_{int} = ae^{-bp_{rej}}$ was detected for large rejection probabilities. It has also been described for frustrated

models [144]. The larger rejection probability at lower temperatures for $J_2 = 0.85$ compared to $J_2 = 0$ in Fig. 10.9 and the earlier breakdown of the algorithm in Figs. 10.8 and 10.10, are consistent with the presumed finite-temperature phase transition. However, this can only serve as an indication for a critical behavior at low but finite temperatures, given the inconclusive results in Sec. 10.4. The results are computed in the purification algorithm, however, it is stressed that the critical slowing down also shows up in the sampling algorithm at low temperatures. Furthermore, it is stressed that the shortcoming of the NN ansatz is in NN theory denoted as a lack of the generalization property of this specific FFNN [144]. Still, in the setting at hand, it becomes clear that the failure of the algorithm can not be assigned to the variational ansatz itself, but the nature of the Hilbert space, the chosen s^z -basis and the local spin changes proposed. For example, the implementation of some kind of importance sampling, would obviously decrease the rejection probability without touching the NN ansatz itself [3]. For this to work out, one could use the probability distribution in the previous time step, for example. Other than that, it is quite elaborate to overcome this issue especially when

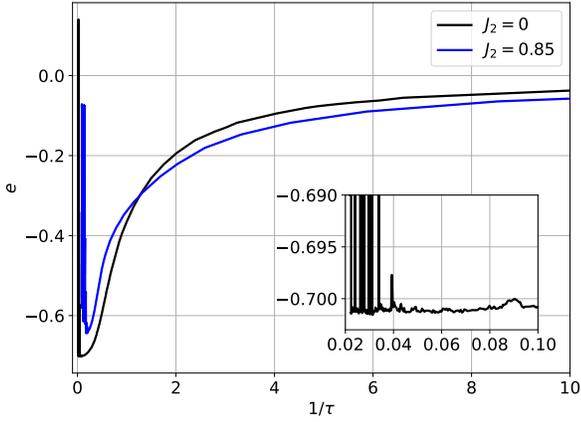


Figure 10.8: Energy for a periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm with $\alpha = 2, 4 \cdot 10^4$ samples and $\epsilon = 10^{-5}$. The inset depicts the breakdown of the algorithm for $J_2 = 0$.

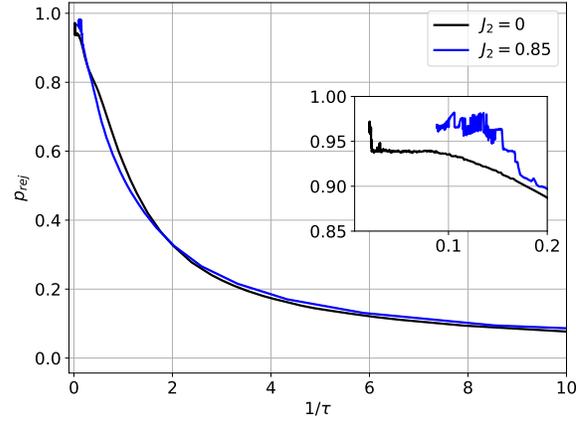


Figure 10.9: Rejection probability for a periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm. The parameters are the same as in Fig. 10.8. The inset depicts the breakdown of the algorithm.

using the VMC method. The task is to propose new configurations in the Monte Carlo sampling in a way that the rejection probability does not grow larger than $\sim 80\%$. It is worth noting that there exists a golden rule stating that the rejection probability should ideally be around 50% for the Monte Carlo sampling to be efficient [229]. One way to overcome this issue is to make use of cluster algorithms like the Swendsen-Wang algorithm [235] or the Wolff-algorithm [236]. However, these are hardly efficiently applicable for models that go beyond simple exchange interaction models [229]. The efficient application of such a cluster algorithm is worth a separate study itself, especially, since they are rarely used in a variational setting. Another way to overcome the issue of critical slowing down is to change the variational NN ansatz to a so-called autoregressive NN ansatz [237]. These are inspired by classical models, where the probability is decomposed into a product of conditional

probabilities

$$p_{\mathbb{P}}(\mathbf{v}) = \prod_{i=1}^N p_{\mathbb{P},i}(v_i | \mathbf{v}_{<i}) \quad (10.11)$$

with \mathbb{P} the variational parameters and \mathbf{v} the visible or input nodes [238]. This allows for every factor in Eq. (10.11) to be parametrized by a separate FFNN [239]. The crucial point is that, due to the product structure, the sampling of clusters of the model at hand is enabled, for example by proposing a transition of the form $\mathbf{s} \rightarrow \mathbf{s}'$ with $\mathbf{s}_{\leq N-k} = \mathbf{s}'_{\leq N-k}$ held fixed for any $k = 1, \dots, N$ [240]. However, this way, there are only N specific and fixed clusters allowed. With the exception of Ref. [237], they have been mostly deployed for classical models or the approximation of classical probability distributions in general. In the quantum case, this would mean that the absolute value of the variational wave function ansatz needs to be of the product structure in Eq. (10.11). Therefore, it seems to be beneficial to divide the wave function in a part describing its absolute value and another part describing its phase. A complex function of this kind is known as a real-valued split function and was mentioned in Sec. 5.5. The sigmoidals ansatz (SA), discussed in Sec. 5.6.1, could for example be used by replacing the NN describing the absolute value by an autoregressive NN. Since they define a very different and complex architecture, the detailed investigation of the complex-valued autoregressive NNs goes beyond the scope of this work.

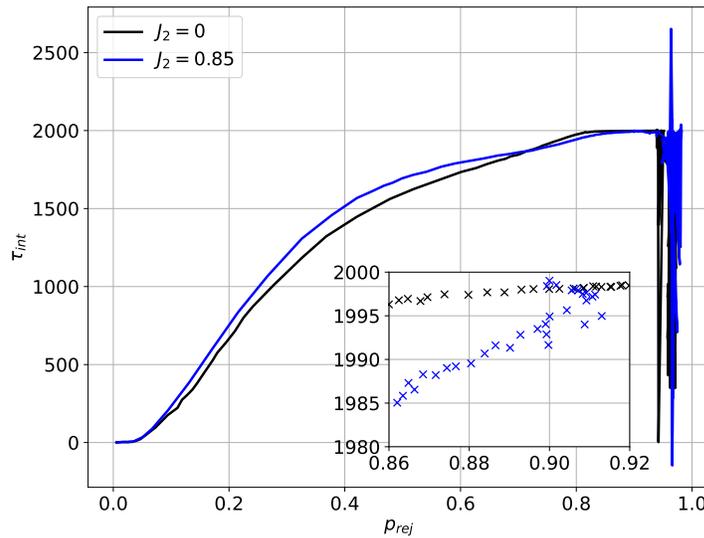


Figure 10.10: Integrated autocorrelation time τ_{int} for a periodic J_1 - J_2 model on the square lattice with $N = 4 \times 4$ sites obtained using the purification algorithm. The parameters are the same as in Fig. 10.8.

It is stressed that this issue is not prevalent in the ground state optimization since quantum number projections, for example for the total spin, are used in this case [5]. For the elaborate CNN used in Ref. [225], a certain sign structure needs to be imposed additionally for the initial state in order for the optimization to reach the ground state. All in all, it can be concluded that without restricting the wave function to be in a certain subspace of the Hilbert space the ground state of the

two-dimensional J_1 - J_2 is difficult to approximate variationally. This is in accordance with similar results for eigenstates of this model [241].

Chapter 11

Conclusion

This thesis has contributed to classify to what extent the idea of using FFNNs in a VMC algorithm is efficiently applicable to quantum spin models in order to obtain an accurate evolution in real and imaginary time. The imaginary-time evolution for the one-dimensional Heisenberg model was studied in the purification, the sampling and the typicality algorithm to obtain finite-temperature correlation functions. In the purification and the sampling algorithm, results for local correlation functions in a large temperature range have been obtained and compared to ED and QMC results. The accuracy is dominated by the sampling error. Therefore, the NN ansatz used in this work can sufficiently approximate the wave functions along the trajectory of the imaginary-time evolution. The number of variational parameters necessary is quadratic in the length of the spin chain. Additionally, even the free energy and thus other thermodynamic variables are feasible. However, the free energy is in the purification algorithm more reliant on the accuracy with which the infinite-temperature state is approximated and on an accurate Monte Carlo sampling. This leads to an error for low temperatures that dominates over the sampling error. The results obtained in this work allow not only to be compared to results derived by other methods, but also to be applied to models in which other numerical results are lacking. Combined with the β modification, the reliability of these results can be judged and the error of the method can be quantified. This way, a lasting critique of variational algorithms in physics in specific and NN algorithms in general is managed.

Furthermore, the criteria that a variational ansatz needs to fulfill in order to be a variational canonical thermal pure quantum state (VCTPQS), is investigated in detail. Building upon these, the combined PP-NN ansatz is recognized and shown to be an intriguing candidate. Unfortunately, it is, due to its high entanglement entropy, also cumbersome to evolve accurately. Since quantum typical states are in general highly entangled, exploiting quantum typicality in a variational setting for computing accurate results at low temperatures seems unlikely. However, it is stressed that the VCTPQs provide more accurate results than the sampling algorithm at higher temperatures. Furthermore, the confusing naming of minimally-entangled typical thermal states (METTS) is discussed. It is shown that those states do not show any sign of quantum typicality, as defined in this project, and thus a clarification in notation is proposed by dropping the term typical.

For the dynamical correlation functions of the infinite-temperature state of the purification algorithm, the scaling of the number of hidden units, that is necessary to obtain accurate results, was shown

to scale with the system size at least like $M \sim N^\alpha$ with $\alpha \gtrsim 4$, but is more likely to show an exponential scaling. The RBM thus fails to efficiently approximate a finite number of states with a large entanglement entropy, although it was proven to do so for some specific states. The reason for this failure is thus the same as the one for the inaccurate imaginary-time evolution of VCTPQSs. This finding is in accordance with similar results in DMRG and the scaling of the bond dimension with the entanglement entropy. In this regard, the VMC method using FFNNs does not represent a significant breakthrough, which aligns with the correspondence observed between MPSs and RBM ansatz wave functions.

In the last part of this study, the purification and the sampling algorithm are applied to the two-dimensional Heisenberg model and the two-dimensional J_1 - J_2 model with $J_1 = 1$ and $J_2 > 0$. Especially for the latter model, numerical results are rare. For the two-dimensional Heisenberg model with $N = 4 \times 4$ sites, the results obtained for the energy with the purification and the sampling algorithm agree with the ED results to an accuracy of $\sim 10^{-3}$. For the two-dimensional J_1 - J_2 model with $J_2 \neq 0$, the construction of the infinite-temperature state in the purification algorithm needs to be altered using the Hadamard transform. The finite-temperature results are for this model in the purification and the sampling algorithm less accurate. An increase in the number of hidden units, the number of Monte Carlo samples and the desired accuracy of the imaginary-time evolution is inevitable. Unfortunately, the critical slowing down of the Monte Carlo sampling is more pronounced for the two-dimensional J_1 - J_2 model and inhibits accurate results at the presumed critical temperature, which is suggested by CTMRG results. However, the sampling algorithm is used to investigate if there are signs of a finite-temperature phase transition detectable at temperatures right above the critical temperature for finite system sizes $N = 4 \times 4$ and $N = 6 \times 6$. While there is a sudden increase in the spin-spin correlator observable, the results for the specific heat capacity can not confirm an emerging peak and thus differ significantly from the CTMRG results in the thermodynamic limit. This can mainly be attributed to finite-size effects. All in all, one needs to admit that the system sizes considered and the temperatures reached with the sampling algorithm are not sufficient to conclusively support the notion of the phase transition, let alone provide a conclusive approximation of the critical temperature.

Chapter 12

Outlook

The finite-temperature properties of the one-dimensional Heisenberg model are extensively studied in this project. While correlation functions are accurately obtained in the purification and the sampling algorithm, the accuracy of the free energy could be further improved by choosing the infinite-temperature state more accurately. Furthermore, it would be interesting to further investigate how the β modification can be adjusted to include an improvement of the free energy results. The developed algorithm for the purification as well as the sampling method can thus serve as a basis for further investigations and computations of correlation functions of physical interest, for example for even more complex spin models with larger interaction ranges. It is stressed, that the NN ansatz used throughout this work has not been tailored to any specific properties of the model, which could be done to improve the accuracy of the results. Furthermore, the results for the models considered in this work could also help to better understand which class of highly-entangled states on the trajectory of the time evolution can be efficiently represented by the NN ansatz used throughout this work and similar FFNN ansatz wave functions, and what kind of states can not. While a similar relation is available in DMRG, it is lacking for FFNN ansatz wave functions. It is stressed that this is the reason for the inaccuracy of both the imaginary-time evolution of the VCTPQS and the real-time evolution of the infinite-temperature state in the purification algorithm, whenever the number of variational parameters is not adjusted drastically with system size.

For the investigation of the finite-temperature results in the two-dimensional J_1 - J_2 model, it is desirable to investigate how the hyperparameters need to be adjusted to obtain results that are as accurate as in the one-dimensional case. Additionally, it would be interesting to see how the application of the translational symmetry could potentially ease the necessity to increase these outer parameters.

Fundamentally, it would also be of great interest to compare the results obtained in this work to the ones that are attainable via the method using exact representations, which was mentioned in Sec. 7.3. This could allow to tailor the universal ansatz wave functions to some extent to some specific features of the J_1 - J_2 model, for example, and this way improve the accuracy and the efficiency of the algorithm.

To allow for accurate results for the order parameter and the specific heat capacity in the two-dimensional J_1 - J_2 model at small temperatures, the finite-size effects need to be studied in detail

to identify what system sizes need to be considered to obtain results that are in accordance with the CTMRG results in the thermodynamic limit. An accurate investigation of the presumed finite-temperature phase transition is only achieved by altering the Monte Carlo sampling fundamentally. In principle, there are three paths to go along. One is to try to translate the Swendsen-Wang or Wolff algorithm directly to the VMC formalism and tailor it to the J_1 - J_2 model. Another one is to make use of the autoregressive NNs discussed in this work. This way, a certain form of clusters is directly implied, whereas in the Swendsen-Wang or Wolff algorithm the clusters form in every step specifically depending on the Hamiltonian at hand. The third one is to make use of the probability distribution of the previous step in the setting of importance sampling. This specific probability distribution could help to propose states that have a finite probability. This way, the rejection probability should not grow to values that result in a critical slowing down.

Bibliography

- [1] B.Y. Pan, Y. Xu, J.M. Ni, S.Y. Zhou, X.C. Hong, X. Qiu, and S.Y. Li. Unambiguous experimental verification of linear-in-temperature spinon thermal conductivity in an antiferromagnetic Heisenberg chain. *Phys. Rev. Lett.*, 129:167201, Oct 2022.
- [2] M.A. Nielsen and I.L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, USA, 10th edition, 2010.
- [3] Federico Becca and Sandro Sorella. *Quantum Monte Carlo Approaches for Correlated Systems*. Cambridge University Press, 2017.
- [4] J. Matoušek. *Lectures on Discrete Geometry*. Springer New York, 2002.
- [5] Yusuke Nomura and Masatoshi Imada. Dirac-type nodal spin liquid revealed by refined quantum many-body solver using neural-network wave function, correlation ratio, and level spectroscopy. *Phys. Rev. X*, 11:031034, Aug 2021.
- [6] Olivier Gauthé and Frédéric Mila. Thermal Ising transition in the spin-1/2 J_1 - J_2 Heisenberg model. *Phys. Rev. Lett.*, 128:227202, Jun 2022.
- [7] Gaopei Pan and Zi Yang Meng. The sign problem in quantum Monte Carlo simulations. In Tapash Chakraborty, editor, *Encyclopedia of Condensed Matter Physics (Second Edition)*, pages 879–893. Academic Press, Oxford, second edition edition, 2024.
- [8] U. Schollwöck. The density-matrix renormalization group. *Rev. Mod. Phys.*, 77:259–315, Apr 2005.
- [9] I. Glasser, N. Pancotti, M. August I.D. Rodriguez and J.I. Cirac. Neural network quantum states, string-bond states, and chiral topological states. *Phys. Rev. X* 8, 011006, 2018.
- [10] D. Tahara and M. Imada. Variational Monte Carlo method combined with quantum-number projection and multi-variable optimization. *J. Phys. Soc. Jpn.*, 77, Nov 2008.
- [11] G. Carleo and M. Troyer. Solving the quantum many-body problem with artificial neural networks. *Science*, Vol. 355, Issue 6325:602–606, 2017.
- [12] A. Altland and B.D. Simons. *Condensed Matter Field Theory*. Cambridge University Press, Cambridge, 2010.

- [13] W. Nolting. *Quantentheorie des Magnetismus 1*. Vieweg+Teubner Verlag, Wiesbaden, 1968.
- [14] W. Nolting. *Quantentheorie des Magnetismus 2*. Vieweg+Teubner Verlag, Wiesbaden, 1968.
- [15] Libor Šmejkal, Jairo Sinova, and Tomas Jungwirth. Emerging research landscape of altermagnetism. *Phys. Rev. X*, 12:040501, Dec 2022.
- [16] W. Heisenberg. Zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 49:619–636, 1928.
- [17] G. Czycholl. *Theoretische Festkörperphysik*. Springer-Verlag Berlin Heidelberg, 3 edition, 2008.
- [18] M. A. Ruderman and C. Kittel. Indirect exchange coupling of nuclear magnetic moments by conduction electrons. *Phys. Rev.*, 96:99–102, Oct 1954.
- [19] Tadao Kasuya. A theory of metallic ferro- and antiferromagnetism on Zener’s model. *Progress of Theoretical Physics*, 16(1):45–57, 07 1956.
- [20] Kei Yosida. Magnetic properties of Cu-Mn alloys. *Phys. Rev.*, 106:893–898, Jun 1957.
- [21] J Schlappa, U Kumar, K J Zhou, S Singh, M Mourigal, V N Strocov, A Revcolevschi, L Patthey, H M Rønnow, S Johnston, and T Schmitt. Probing multi-spinon excitations outside of the two-spinon continuum in the antiferromagnetic spin chain cuprate Sr2CuO3. *Nature Communications*, 9(1):5394, December 2018.
- [22] Oliver Breunig, Markus Garst, Andreas Klümper, Jens Rohrkamp, Mark M. Turnbull, and Thomas Lorenz. Quantum criticality in the spin-1/2 Heisenberg chain system copper pyrazine dinitrate. *Science Advances*, 3(12):eaao3773, 2017.
- [23] H. Kühne, M. Günther, S. Grossjohann, W. Brenig, F. J. Litterst, A. P. Reyes, P. L. Kuhns, M. M. Turnbull, C. P. Landee, and H.-H. Klauss. Low frequency spin dynamics in the quantum magnet copper pyrazine dinitrate. *physica status solidi (b)*, 247(3):671–675, 2010.
- [24] Fabio Franchini. *An Introduction to Integrable Techniques for One-Dimensional Quantum Systems*. Springer Cham, 2017.
- [25] H. Bethe. Zur Theorie der Metalle. *Zeitschrift für Physik*, 71:205–226, 1931.
- [26] Y. Wang, W.L. Yang, J. Cao, and K. Shi. *Off-Diagonal Bethe Ansatz for Exactly Solvable Models*. Springer Berlin, Heidelberg, 2015.
- [27] C. N. Yang. Some exact results for the many-body problem in one dimension with repulsive delta-function interaction. *Phys. Rev. Lett.*, 19:1312–1315, Dec 1967.
- [28] Rodney J. Baxter. *Exactly Solved Models in Statistical Mechanics*. Courier Corporation, Mineola, New York, 2015.
- [29] Jean-Sébastien Caux and Jorn Mossel. Remarks on the notion of quantum integrability. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(02):P02023, Feb 2011.

- [30] L A Takhtadzhan and Lyudvig D Faddeev. The quantum method of the inverse problem and the Heisenberg XYZ model. *Russian Mathematical Surveys*, 34(5):11, Oct 1979.
- [31] L. A. Takhtajan. Introduction to algebraic Bethe ansatz. In B. S. Shastry, S. S. Jha, and V. Singh, editors, *Exactly Solvable Problems in Condensed Matter and Relativistic Field Theory*, pages 175–219, Berlin, Heidelberg, 1985. Springer Berlin Heidelberg.
- [32] E K Sklyanin. Boundary conditions for integrable quantum systems. *Journal of Physics A: Mathematical and General*, 21(10):2375, May 1988.
- [33] Andreas Klümper. Thermodynamics of the anisotropic spin-1/2 Heisenberg chain and related quantum chains. *Zeitschrift für Physik B Condensed Matter*, 91:507–519, May 1988.
- [34] Henrik Johannesson Abolfazl Bayat, Sougato Bose, editor. *Entanglement in Spin Chains*. Springer Cham, 2022.
- [35] O. V. Marchukov, A. G. Volosniev, M. Valiente, D. Petrosyan, and N. T. Zinner. Quantum spin transistor with a Heisenberg spin chain. *Nature Communications*, 7:13070, 2016.
- [36] Marta P. Estarellas, Irene D’Amico, and Timothy P. Spiller. Topologically protected localised states in spin chains. *Scientific Reports*, 7:42904, 2017.
- [37] John S. Van Dyke, Yadav P. Kandel, Haifeng Qiao, John M. Nichol, Sophia E. Economou, and Edwin Barnes. Protecting quantum information in quantum dot spin chains by driving exchange interactions periodically. *Phys. Rev. B*, 103:245303, 2021.
- [38] Frédéric Mila Claudine Lacroix, Philippe Mendels. *Introduction to Frustrated Magnetism*. Springer Berlin, Heidelberg, 2011.
- [39] Francesco Ferrari and Federico Becca. Dynamical structure factor of the J_1 - J_2 Heisenberg model on the triangular lattice: Magnons, spinons, and gauge fields. *Phys. Rev. X*, 9:031026, Aug 2019.
- [40] Leon Balents. Spin liquids in frustrated magnets. *Nature*, 464:199–208, 2010.
- [41] Roderich Moessner and Arthur P. Ramirez. Geometrical frustration. *Physics Today*, 59(2):24–29, 02 2006.
- [42] P. Chandra and B. Doucot. Possible spin-liquid state at large s for the frustrated square Heisenberg lattice. *Phys. Rev. B*, 38:9335–9338, Nov 1988.
- [43] R. Melzi, S. Aldrovandi, F. Tedoldi, P. Carretta, P. Millet, and F. Mila. Magnetic and thermodynamic properties of $\text{Li}_2\text{VOSiO}_4$: A two-dimensional $s = 1/2$ frustrated antiferromagnet on a square lattice. *Phys. Rev. B*, 64:024409, Jun 2001.
- [44] F. Essler, H. Frahm, F. Göhmann et al. *The One-Dimensional Hubbard Model*. Cambridge University Press, Cambridge, 2005.

- [45] W. Marshall and Rudolf Ernst Peierls. Antiferromagnetism. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 232(1188):48–68, 1955.
- [46] Elliott Lieb, Theodore Schultz, and Daniel Mattis. Two soluble models of an antiferromagnetic chain. *Annals of Physics*, 16(3):407–466, 1961.
- [47] Andreas Voigt, Johannes Richter, and Nedko B. Ivanov. Marshall-Peierls sign rule for excited states of the frustrated J_1 - J_2 Heisenberg antiferromagnet. *Physica A: Statistical Mechanics and its Applications*, 245(3):269–275, 1997.
- [48] E. Lieb and D. Mattis. Ordering energy levels of interacting spin systems. *J. Math. Phys.*, 3(4):749–751, 1962.
- [49] Elliott Lieb and Daniel Mattis. *Ordering Energy Levels of Interacting Spin Systems*, pages 43–45. Springer Berlin Heidelberg, Berlin, Heidelberg, 2002.
- [50] Assa Auerbach. *Ground States of the Heisenberg Model*, pages 51–60. Springer New York, New York, NY, 1994.
- [51] Sergey Bravyi. Monte Carlo simulation of stoquastic Hamiltonians. *Quantum Info. Comput.*, 15(13–14):1122–1140, Oct 2015.
- [52] W. Nolting. *Theoretical Physics 8*. Springer Cham, Cham, Switzerland, 2017.
- [53] Hidetoshi Nishimori and Gerardo Ortiz. *Elements of Phase Transitions and Critical Phenomena*. Oxford University Press, 12 2010.
- [54] Kenneth G. Wilson. Renormalization group and critical phenomena. i. Renormalization group and the Kadanoff scaling picture. *Phys. Rev. B*, 4:3174–3183, Nov 1971.
- [55] P. Francesco, P. Mathieu and D. Sénéchal. *Conformal Field Theory*. Springer New York, NY, 1997.
- [56] Aron J. Beekman, Louk Rademaker, and Jasper van Wezel. An introduction to spontaneous symmetry breaking. *SciPost Phys. Lect. Notes*, page 11, 2019.
- [57] P. Chandra, P. Coleman, and A. I. Larkin. Ising transition in frustrated Heisenberg models. *Phys. Rev. Lett.*, 64:88–91, Jan 1990.
- [58] Ingemar Bengtsson and Karol Zyczkowski. *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge University Press, 2006.
- [59] Lucas Hackl, Tommaso Guaita, Tao Shi, Jutho Haegeman, Eugene Demler, and J. Ignacio Cirac. Geometry of variational methods: dynamics of closed quantum systems. *SciPost Phys.*, 9:48, 2020.
- [60] A. Ashtekar and T.A. Schilling. *Geometrical Formulation of Quantum Mechanics*, pages 23–65. Springer New York, New York, NY, 1999.

- [61] Chae-Yeun Park and Michael J. Kastoryano. Geometry of learning neural quantum states. *Phys. Rev. Res.*, 2:023232, May 2020.
- [62] Guido Fubini. Sulla teoria delle funzioni automorfe e delle loro trasformazioni. *Annali di Matematica Pura ed Applicata (1898-1922)*, 14:33–67, 1908.
- [63] E. Study. Kürzeste Wege im komplexen Gebiet. *Math. Ann.*, 60:321–378, 1905.
- [64] Jing Liu, Haidong Yuan, Xiao-Ming Lu, and Xiaoguang Wang. Quantum Fisher information matrix and multiparameter estimation. *Journal of Physics A: Mathematical and Theoretical*, 53(2):023001, Dec 2019.
- [65] Xiao-Ming Lu Jing Liu, Haidong Yuan and Xiaoguang Wang. The quantum geometric tensor from generating functions. *J. Phys. A: Math. Theor.*, 53(02):023001, 2020.
- [66] Shan-Ming Ruan. Purification complexity without purifications. *Journal of High Energy Physics*, 2021:92, 2021.
- [67] J. Jost. *Riemannian Geometry and Geometric Analysis*. Springer Berlin, Heidelberg, 2011.
- [68] Y. Le Floch. *A Brief Introduction to Berezin–Toeplitz Operators on Compact Kähler Manifolds*. Springer Cham, 2018.
- [69] Sam McArdle, Tyson Jones, Suguru Endo, Ying Li, Simon C. Benjamin, and Xiao Yuan. Variational ansatz-based quantum simulation of imaginary time evolution. *npj Quantum Information*, 5:177202, 2019.
- [70] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum natural gradient. *Quantum*, 4:269, May 2020.
- [71] Alex Harvey, editor. *On Einstein’s Path*. Springer New York, NY, 1999.
- [72] J.P. Provost and G Vallee. Riemannian structure on manifolds of quantum states. *Commun.Math. Phys*, pages 289–301, Sep 1980.
- [73] Michael V. Berry. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 392:45 – 57, 1984.
- [74] Eliahu Cohen, Hugo Larocque, Frédéric Bouchard, Farshad Nejdatsattari, Yuval Gefen, and Ebrahim Karimi. Geometric phase from Aharonov–Bohm to Pancharatnam–Berry and beyond. *Nature Reviews Physics*, 1:437–449, 2019.
- [75] Yuan Yao, Pierre Cussenot, Richard A. Wolf, and Filippo Miatto. Complex natural gradient optimization for optical quantum circuit design. *Phys. Rev. A*, 105:052402, May 2022.
- [76] S.-I. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2):251–276, 1998.

- [77] J. Dong D. Xu and H. Zhang. Deterministic convergence of Wirtinger-gradient methods for complex-valued neural networks. *Neural Processing Lett*, 45(2):445–456, 2017.
- [78] J. Anandan and Y. Aharonov. Geometry of quantum evolution. *Phys. Rev. Lett.*, 65:1697–1700, Oct 1990.
- [79] J. Gemmer, M. Michel and G. Mahler. *Quantum Thermodynamics*. Springer Berlin, Heidelberg, 2009.
- [80] Dennis Wagner, Andreas Klümper, and Jesko Sirker. Thermodynamics based on neural networks. *Phys. Rev. B*, 109:155128, Apr 2024.
- [81] W. Nolting. *Theoretical Physics 7*. Springer Cham, Cham, Switzerland, 2018.
- [82] Y. Yamaji Y. Nomura, A. S. Darmawan and M. Imada. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. *PHYSICAL REVIEW B* 96, 205152, 2017.
- [83] David D. L. Minh and Do Le (Paul) Minh. Understanding the Hastings algorithm. *Communications in Statistics - Simulation and Computation*, 44(2):332–349, 2015.
- [84] Hidemaro Suwa and Synge Todo. Markov chain Monte Carlo method without detailed balance. *Phys. Rev. Lett.*, 105:120603, Sep 2010.
- [85] Hal Tasaki. *Physics and Mathematics of Quantum Many-Body Systems*. Springer Cham, 2020.
- [86] Sandro Sorella. Green function Monte Carlo with stochastic reconfiguration. *Phys. Rev. Lett.*, 80:4558–4561, May 1998.
- [87] James Martens. New insights and perspectives on the natural gradient method. *J. Mach. Learn. Res.*, 21(1), Jan 2020.
- [88] Alessandro Sinibaldi, Clemens Giuliani, Giuseppe Carleo, and Filippo Vicentini. Unbiasing time-dependent variational Monte Carlo by projected quantum evolution. *Quantum*, 7:1131, October 2023.
- [89] Markus Schmitt and Markus Heyl. Quantum many-body dynamics in two dimensions with artificial neural networks. *Phys. Rev. Lett.*, 125:100503, Sep 2020.
- [90] Douglas Hendry, Hongwei Chen, and Adrian Feiguin. Neural network representation for minimally entangled typical thermal states. *Phys. Rev. B*, 106:165111, Oct 2022.
- [91] M. Fuhry and L. Reichel. A new Tikhonov regularization method. *Numer Algor*, 59:433–445, March 2012.
- [92] F.S.V. Bazán. Simple and efficient determination of the Tikhonov regularization parameter chosen by the generalized discrepancy principle for discrete ill-posed problems. *J Sci Comput*, 63:163–184, April 2015.
- [93] Thomas N. E. Greville Adi Ben-Israel. *Generalized Inverses*. Springer New York, NY, 2010.

- [94] Nieves Castro-González, Froilán M. Dopico, and Juan M. Molera. Multiplicative perturbation theory of the Moore–Penrose inverse and the least squares problem. *Linear Algebra and its Applications*, 503:1–25, 2016.
- [95] Yusuke Nomura, Andrew S. Darmawan, Youhei Yamaji, and Masatoshi Imada. Restricted Boltzmann machine learning for solving strongly correlated quantum systems. *Phys. Rev. B*, 96:205152, Nov 2017.
- [96] U. Tautenhahn. Regularization of linear ill-posed problems with noisy right hand side and noisy operator. *J of Inverse and Ill-Posed Problems*, 16:507–523, 2008.
- [97] S. Lu, S.V. Pereverzev, Y. Shao et al. On the generalized discrepancy principle for Tikhonov regularization in Hilbert scales. *The Journal of Integral Equations and Applications*, 22:483–517, 2010.
- [98] M. Schmitt and M. Reh. jVMC: Versatile and performant variational Monte Carlo leveraging automated differentiation and GPU acceleration. *SciPost Phys. Codebases*, page 2, 2022.
- [99] D. Hilbert. Mathematical problems. *Bull. Amer. Math. Soc.*, 8:437–479, 1902.
- [100] D. Hilbert. Über die Gleichung neunten Grades. *D. Math. Ann.*, 97: 243, 1927.
- [101] Ehrenfried Walter von Tschirnhaus. Methodus auferendi omnes terminos intermedios ex data aequatione. *Acta Eruditorum*, 2(0):204–207, 1683.
- [102] Ehrenfried Walter von Tschirnhaus and R. F. Green. A method for removing all intermediate terms from a given equation. *SIGSAM Bull.*, 37(1):1–3, Mar 2003.
- [103] A G Vitushkin. On Hilbert’s thirteenth problem and related questions. *Russian Mathematical Surveys*, 59(1):11, Feb 2004.
- [104] D. A. Sprecher. A survey of solved and unsolved problems on superposition of functions. *Journal of Approximation Theory*, 6:123–134, 19672.
- [105] Alexander B. Givental, Boris A. Khesin, Jerrold E. Marsden, Alexander N. Varchenko, Victor A. Vassiliev, Oleg Ya. Viro, and Vladimir M. Zakalyukin, editors. *On functions of three variables*, pages 5–8. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [106] V.I. Arnold. On functions of three variables. *Amer. Math. Soc. Transl.*, 28:51–54, 1963.
- [107] V.I. Arnold. On functions of three variables. *Dokl. Akad. Nauk SSSR*, 114:679–681, 1957.
- [108] V.I. Arnold. On the representation of functions of several variables as a superposition of functions of a smaller number of variables. *Mat. Prosveshchenie*, 3:41–61, 1958.
- [109] Alexander B. Givental, Boris A. Khesin, Jerrold E. Marsden, Alexander N. Varchenko, Victor A. Vassiliev, Oleg Ya. Viro, and Vladimir M. Zakalyukin, editors. *On the representation of functions of several variables as a superposition of functions of a smaller number of variables*, pages 25–46. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

- [110] Johannes Schmidt-Hieber. The Kolmogorov–Arnold representation theorem revisited. *Neural Networks*, 137:119–126, 2021.
- [111] Federico Girosi and Tomaso Poggio. Representation Properties of Networks: Kolmogorov’s Theorem Is Irrelevant. *Neural Computation*, 1(4):465–469, 12 1989.
- [112] A.N. Gorban. Approximation of continuous functions of several variables by an arbitrary nonlinear continuous function of one variable, linear functions, and their superpositions. *Applied Mathematics Letters*, 11(3):45–49, 1998.
- [113] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [114] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [115] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signal Systems 2*, 303, 1989.
- [116] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.
- [117] Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015.
- [118] Robert Hecht-Nielsen. *Neurocomputing : The Technology of Non-Algorithmic Information Processing*. Addison-Wesley Pub. Co., 1990.
- [119] Robert H. Nielsen. Kolmogorov’s mapping neural network existence theorem. In *Proceedings of the IEEE First International Conference on Neural Networks* (San Diego, CA), volume III, pages 11–13. Piscataway, NJ: IEEE, 1987.
- [120] David Sprecher. On the structure of representations of continuous functions of several variables as finite sums of continuous functions of one variable. *Proceedings of The American Mathematical Society - PROC AMER MATH SOC*, 17, 02 1966.
- [121] Phillip A. Ostrand. Dimension of metric spaces and Hilbert’s problem 13. *Bulletin of the American Mathematical Society*, 71(4):619 – 622, 1965.
- [122] Jürgen Braun. *An Application of Kolmogorov’s Superposition Theorem to Function Reconstruction in Higher Dimensions*. PhD thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, December 2009.
- [123] Věra Kůrková. Kolmogorov’s theorem and multilayer neural networks. *Neural Networks*, 5(3):501–506, 1992.
- [124] A. Hirose. *Complex-Valued Neural Networks*. Springer-Verlag Berlin Heidelberg, 2012.
- [125] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5:115–133, 1943.

- [126] Snehashish Chakraverty, Deepti Moyi Sahoo, and Nisha Rani Mahato. *McCulloch–Pitts Neural Network Model*, pages 167–173. [Springer Singapore](#), Singapore, 2019.
- [127] Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová. Machine learning and the physical sciences. *Rev. Mod. Phys.*, 91:045002, Dec 2019.
- [128] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1):53, 2021.
- [129] Petia Koprinkova-Hristova, Valeri Mladenov, and Nikola K. Kasabov. *Artificial Neural Networks*. [Springer Cham](#), 2014.
- [130] Shuang Cong and Yang Zhou. A review of convolutional neural network architectures and their optimizations. *Artificial Intelligence Review*, 56:1905–1969, 2023.
- [131] Simone Scardapane, Steven Van Vaerenbergh, Amir Hussain, and Aurelio Uncini. Complex-valued neural networks with nonparametric activation functions. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 4(2):140–150, 2020.
- [132] ChiYan Lee, Hideyuki Hasegawa, and Shangce Gao. Complex-valued neural networks: A comprehensive survey. *IEEE/CAA Journal of Automatica Sinica*, 9(JAS-2022-0298):1406, 2022.
- [133] Marin Bukov, Markus Schmitt, and Maxime Dupont. Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape. *SciPost Phys.*, 10:147, 2021.
- [134] Charu Aggarwal. *Restricted Boltzmann Machines*, pages 231–264. [Springer International Publishing](#), Cham, 2023.
- [135] G.E Hinton. Products of experts. In *9th International Conference on Artificial Neural Networks: ICANN '99*, pages 1–6, London, 1999. [IEE](#).
- [136] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.
- [137] James Martens, Arkadev Chattopadhyay, Toniann Pitassi, and Richard Zemel. On the representational efficiency of restricted Boltzmann machines. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, page 2877–2885, Red Hook, NY, USA, 2013. [Curran Associates Inc.](#)
- [138] Nicolas Le Roux and Yoshua Bengio. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Computation*, 20(6):1631–1649, 2008.

- [139] Guido Montufar and Nihat Ay. Refinements of universal approximation results for deep belief networks and restricted Boltzmann machines. *Neural Computation*, 23(5):1306–1319, 2011.
- [140] Kihyuk Sohn and Honglak Lee. Learning invariant representations with local transformations. In *Proceedings of the 29th International Conference on Machine Learning*, ICML’12, page 1339–1346, Madison, WI, USA, 2012. [Omnipress](#).
- [141] H. Lee, R. Grosse, R. Ranganath and A. Y. Ng. Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, 54:95–103, October 2011.
- [142] D. L. Deng, X. Li and S. Das Sarma. Quantum entanglement in neural network states. *Phys. Rev. X*, 7:021021, May 2017.
- [143] Hugh. Cartwright. *Artificial Neural Networks*. Methods in Molecular Biology, 2190. [Springer US](#), New York, NY, 3rd ed. 2021. edition, 2021.
- [144] T. Westerhout, N. Astrakhantsev, K.S. Tikhonov et al. Generalization properties of neural network approximations to frustrated magnet ground states. *Nat. Commun.*, 11:1593, 2019.
- [145] Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, and Oriol Vinyals. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM*, 64(3):107–115, Feb 2021.
- [146] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, 810:1–124, 2019.
- [147] Dong Eui Chang Anthony L. Caterini. *Deep Neural Networks in a Mathematical Framework*. [Springer Cham](#), 2018.
- [148] Boris Hanin. Universal function approximation by deep neural nets with bounded width and ReLU activations. *Mathematics*, 7(10), 2019.
- [149] Ronen Eldan and Ohad Shamir. The power of depth for feedforward neural networks. In Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir, editors, *29th Annual Conference on Learning Theory*, volume 49 of *Proceedings of Machine Learning Research*, pages 907–940, Columbia University, New York, New York, USA, 23–26 Jun 2016. PMLR.
- [150] David Poulin, Angie Qarry, Rolando Somma, and Frank Verstraete. Quantum simulation of time-dependent Hamiltonians and the convenient illusion of Hilbert space. *Phys. Rev. Lett.*, 106:170501, Apr 2011.
- [151] M B Hastings. An area law for one-dimensional quantum systems. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(08):P08024, Aug 2007.
- [152] M. B. Plenio, J. Eisert, J. Dreißig, and M. Cramer. Entropy, entanglement, and area: Analytical results for harmonic lattice systems. *Phys. Rev. Lett.*, 94:060503, Feb 2005.

- [153] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, 2014.
- [154] Román Orús. Tensor networks for complex quantum systems. *Nature Reviews Physics*, 1:538–550, 2019.
- [155] Xun Gao and Lu-Ming Duan. Efficient representation of quantum many-body states with deep neural networks. *Nature Communications*, 8:662, 2017.
- [156] Jing Chen, Song Cheng, Haidong Xie, Lei Wang, and Tao Xiang. Equivalence of restricted Boltzmann machines and tensor network states. *Phys. Rev. B*, 97:085104, Feb 2018.
- [157] E.M. Stoudenmire and Steven R. White. Studying two-dimensional systems with the density matrix renormalization group. *Annual Review of Condensed Matter Physics*, 3(1):111–128, 2012.
- [158] W. Heitler and F. London. Wechselwirkung neutraler Atome und homöopolare Bindung nach der Quantenmechanik. *Zeitschrift für Physik*, 44:455–472, 06 1927.
- [159] V. Fock. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik*, 61:126–148, 01 1930.
- [160] J.C. Slater. The theory of complex spectra. *Phys. Rev.*, 34:1293–1322, Nov 1929.
- [161] M.C. Gutzwiller. Effect of correlation on the ferromagnetism of transition metals. *Phys. Rev. Lett.*, 10:159–162, Mar 1963.
- [162] R. Jastrow. Many-body problem with strong forces. *Phys. Rev.*, 98:1479–1484, Jun 1955.
- [163] T. Misawa, S. Morita, K. Yoshimi et al. mVMC—Open-source software for many-variable variational Monte Carlo method. *Computer Physics Communications*, 235:447–462, 2019.
- [164] J.F. Annett. *Superconductivity, Superfluids and Condensates*. Oxford University Press, 2004.
- [165] J. Bardeen, L.N. Cooper and J.R. Schrieffer. Theory of superconductivity. *Phys. Rev.*, 108:1175–1204, Dec 1957.
- [166] C. Gros. Superconductivity in correlated wave functions. *Phys. Rev. B*, 38:931–934, Jul 1988.
- [167] P. W. Anderson. The resonating valence bond state in La_2CuO_4 and superconductivity. *Science*, 235(4793):1196–1198, 1987.
- [168] Michele Casula, Seiji Yunoki, Claudio Attaccalite, and Sandro Sorella. Resonating valence bond wave function: from lattice models to realistic systems. *Computer Physics Communications*, 169(1):386–393, 2005. Proceedings of the Europhysics Conference on Computational Physics 2004.

- [169] Linus Pauling. *The nature of the chemical bond and the structure of molecules and crystals : An introduction to modern structural chemistry*. George Fisher Baker non-resident lectureship in chemistry at Cornell University ; [v. 18]. Cornell University Press, Ithaca, N.Y., 3d ed. – edition, 1960.
- [170] P.W. Anderson. Resonating valence bonds: A new kind of insulator? *Materials Research Bulletin*, 8(2):153–160, 1973.
- [171] M. Bajdich, L. Mitas, G. Drobný et al. Pfaffian pairing wave functions in electronic-structure quantum Monte Carlo simulations. *Phys. Rev. Lett.*, 96:130201, Apr 2006.
- [172] M. Wimmer. Algorithm 923: Efficient numerical computation of the Pfaffian for dense and banded skew-symmetric matrices. *ACM Trans. Math. Softw.*, 38(4), Aug 2012.
- [173] A. Cayley. Sur les déterminants gauches. (suite du mémoire t. xxxii. p. 119). *Journal für die reine und angewandte Mathematik*, 1849(38):93–96, 1849.
- [174] Hans-Peter Eckle. 241Models of Strongly Interacting Quantum Matter. In *Models of Quantum Matter: A First Course on Integrability and the Bethe Ansatz*. Oxford University Press, 07 2019.
- [175] Roger G. Melko, Giuseppe Carleo, Juan Carrasquilla, and J. Ignacio Cirac. Restricted Boltzmann machines in quantum physics. *Nature Physics*, 15:887–892, 2019.
- [176] Di Luo and Bryan K. Clark. Backflow transformations via neural networks for quantum many-body wave functions. *Phys. Rev. Lett.*, 122:226401, Jun 2019.
- [177] Hiroki Saito. Solving the Bose–Hubbard model with machine learning. *Journal of the Physical Society of Japan*, 86(9):093001, 2017.
- [178] Dong-Ling Deng, Xiaopeng Li, and S. Das Sarma. Machine learning topological states. *Phys. Rev. B*, 96:195145, Nov 2017.
- [179] Juan Carrasquilla and Roger G. Melko. Machine learning phases of matter. *Nature Physics*, 13:431–434, 2017.
- [180] J.R. McClean, S. Boixo, V.N. Smelyanskiy et al. Barren plateaus in quantum neural network training landscapes. *Nat Commun*, 9:4812, 2018.
- [181] A. Szabó and C. Castelnovo. Neural network wave functions and the sign problem. *Phys. Rev. Research*, 2:033075, 2020.
- [182] Kenny Choo, Titus Neupert, and Giuseppe Carleo. Two-dimensional frustrated J_1 - J_2 model studied with neural network quantum states. *Phys. Rev. B*, 100:125124, Sep 2019.
- [183] Luciano Loris Viteritti, Francesco Ferrari, and Federico Becca. Accuracy of restricted Boltzmann machines for the one-dimensional J_1 - J_2 Heisenberg model. *SciPost Phys.*, 12:166, 2022.

- [184] Yusuke Nomura. Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry. *Journal of Physics: Condensed Matter*, 33(17):174003, Apr 2021.
- [185] Takahiro Mizusaki and Masatoshi Imada. Quantum-number projection in the path-integral renormalization group method. *Phys. Rev. B*, 69:125110, Mar 2004.
- [186] Giuseppe Carleo, Yusuke Nomura, and Masatoshi Imada. Constructing exact representations of quantum many-body systems with deep neural networks. *Nature Communications*, 9:5322, 2018.
- [187] Chih-Chun Chien, Sebastiano Peotta, and Massimiliano Di Ventra. Quantum transport in ultracold atoms. *Nature Physics*, 11(12):998–1004, Dec 2015.
- [188] Anton Mazurenko, Christie S. Chiu, Geoffrey Ji, Maxwell F. Parsons, Márton Kanász-Nagy, Richard Schmidt, Fabian Grusdt, Eugene Demler, Daniel Greif, and Markus Greiner. A cold-atom Fermi–Hubbard antiferromagnet. *Nature*, 545(7655):462–466, May 2017.
- [189] V I Yukalov. Theory of cold atoms: basics of quantum statistics. *Laser Physics*, 23(6):062001, Apr 2013.
- [190] Sandvik, A. W. Quantum Monte Carlo methods at work for novel phases of matter, 2012. <http://physics.bu.edu/~sandvik/trieste12/index.html>, Last accessed on 2024-02-13.
- [191] Anders W. Sandvik. Computational studies of quantum spin systems. *AIP Conference Proceedings*, 1297(1):135–338, 11 2010.
- [192] Frank Verstraete, Tomotoshi Nishino, Ulrich Schollwöck, Mari Carmen Bañuls, Garnet K. Chan, and Miles E. Stoudenmire. Density matrix renormalization group, 30 years on. *Nature Reviews Physics*, 5(5):273–276, May 2023.
- [193] Adrian E. Feiguin and Steven R. White. Finite-temperature density matrix renormalization using an enlarged Hilbert space. *Phys. Rev. B*, 72:220401, Dec 2005.
- [194] Jannes Nys, Zakari Denis, and Giuseppe Carleo. Real-time quantum dynamics of thermal states with neural thermofields. *arXiv:2309.07063*, 2023.
- [195] Yusuke Nomura, Nobuyuki Yoshioka, and Franco Nori. Purifying deep Boltzmann machines for thermal quantum states. *Phys. Rev. Lett.*, 127:060601, Aug 2021.
- [196] S. Sugiura. *Formulation of Statistical Mechanics Based on Thermal Pure Quantum States*. Springer Singapore, 2017.
- [197] R.S. White. Minimally entangled typical quantum states at finite temperature. *Phys. Rev. Lett.*, 102:190601, May 2009.
- [198] Masuo Suzuki. Thermo field dynamics in equilibrium and non-equilibrium interacting quantum systems. *Journal of the Physical Society of Japan*, 54(12):4483–4485, 1985.

- [199] D.P. Landau and K. Binder. *A Guide to Monte Carlo simulations in statistical physics*. Cambridge University Press, Cambridge, 2005.
- [200] J.-C. Walter and G.T. Barkema. An introduction to Monte Carlo methods. *Physica A: Statistical Mechanics and its Applications*, 418:78–87, 2015. Proceedings of the 13th International Summer School on Fundamental Problems in Statistical Physics.
- [201] Huan Zhang, Robert J. Webber, Michael Lindsey, Timothy C. Berkelbach, and Jonathan Weare. Understanding and eliminating spurious modes in variational Monte Carlo using collective variables. *Phys. Rev. Res.*, 5:023101, May 2023.
- [202] Tom Vieijra, Corneel Casert, Jannes Nys, Wesley De Neve, Jutho Haegeman, Jan Ryckebusch, and Frank Verstraete. Restricted Boltzmann machines for quantum states with non-abelian or anyonic symmetries. *Phys. Rev. Lett.*, 124:097201, Mar 2020.
- [203] A. Hams and H. De Raedt. Fast algorithm for finding the eigenvalue distribution of very large matrices. *Phys. Rev. E*, 62:4365–4377, Sep 2000.
- [204] S. Popescu, A. Short and A. Winter. Entanglement and the foundations of statistical mechanics. *Nature Phys*, 2:754–758, 2006.
- [205] P.Reimann. Typicality of pure states randomly sampled according to the Gaussian adjusted projected measure. *J Stat Phys*, 132:921–935, 2008.
- [206] V.D. Milman and G. Schechtman. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Springer Berlin, Heidelberg, 2009.
- [207] Sandu Popescu, Anthony J. Short, and Andreas Winter. The foundations of statistical mechanics from entanglement: Individual states vs. averages. *arXiv:quant-ph/0511225*, 2005.
- [208] S. Sugiura and A. Shimizu. Thermal pure quantum states at finite temperature. *Phys. Rev. Lett.*, 108:240401, 2012.
- [209] S. Sugiura and A. Shimizu. Canonical thermal pure quantum state. *Phys. Rev. Lett.*, 111:010401, 2013.
- [210] Robin Steinigeweg, Jochen Gemmer, and Wolfram Brenig. Spin-current autocorrelations from single pure-state propagation. *Phys. Rev. Lett.*, 112:120601, Mar 2014.
- [211] P.Reimann. Typicality for generalized microcanonical ensembles. *Phys. Rev. Lett.*, 99:160404, 2007.
- [212] Jürgen Schnack, Johannes Richter, Tjark Heitmann, Jonas Richter, and Robin Steinigeweg. Finite-size scaling of typicality-based estimates. *Zeitschrift für Naturforschung A*, 75(5):465–473, 2020.
- [213] Jürgen Schnack, Johannes Richter, and Robin Steinigeweg. Accuracy of the finite-temperature Lanczos method compared to simple typicality-based estimates. *Phys. Rev. Res.*, 2:013186, Feb 2020.

- [214] Henrik Schlüter, Florian Gayk, Heinz-Jürgen Schmidt, Andreas Honecker, and Jürgen Schnack. Accuracy of the typicality approach using Chebyshev polynomials. *Zeitschrift für Naturforschung A*, 76(9):823–834, 2021.
- [215] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291–1294, Aug 1993.
- [216] Eugenio Bianchi, Lucas Hackl, Mario Kieburg, Marcos Rigol, and Lev Vidmar. Volume-law entanglement entropy of typical pure quantum states. *PRX Quantum*, 3:030201, Jul 2022.
- [217] Yuya O. Nakagawa, Masataka Watanabe, Hiroyuki Fujita, and Sho Sugiura. Universality in volume-law entanglement of scrambled pure quantum states. *Nature Communications*, 9:1635, 2018.
- [218] Sheng-Hsuan Lin and Frank Pollmann. Scaling of neural-network quantum states for time evolution. *physica status solidi (b)*, 259(5):2100172, 2022.
- [219] J Oitmaa. The square-lattice Ising model with first and second neighbour interactions. *Journal of Physics A: Mathematical and General*, 14(5):1159, May 1981.
- [220] Cédric Weber, Luca Capriotti, Grégoire Misguich, Federico Becca, Maged Elhadj, and Frédéric Mila. Ising transition driven by frustration in a 2d classical model with continuous symmetry. *Phys. Rev. Lett.*, 91:177202, Oct 2003.
- [221] Wen-Jun Hu, Federico Becca, Alberto Parola, and Sandro Sorella. Direct evidence for a gapless Z_2 spin liquid by frustrating Néel antiferromagnetism. *Phys. Rev. B*, 88:060402, Aug 2013.
- [222] Hong-Chen Jiang, Hong Yao, and Leon Balents. Spin liquid ground state of the spin- $\frac{1}{2}$ square J_1 - J_2 Heisenberg model. *Phys. Rev. B*, 86:024424, Jul 2012.
- [223] R. Haghshenas and D. N. Sheng. $U(1)$ -symmetric infinite projected entangled-pair states study of the spin-1/2 square J_1 - J_2 Heisenberg model. *Phys. Rev. B*, 97:174408, May 2018.
- [224] Shou-Shu Gong, Wei Zhu, D. N. Sheng, Olexei I. Motrunich, and Matthew P. A. Fisher. Plaquette ordered phase and quantum phase diagram in the spin- $\frac{1}{2}$ J_1 - J_2 square Heisenberg model. *Phys. Rev. Lett.*, 113:027201, Jul 2014.
- [225] Kenny Choo, Titus Neupert, and Giuseppe Carleo. Two-dimensional frustrated J_1 - J_2 model studied with neural network quantum states. *Phys. Rev. B*, 100:125124, Sep 2019.
- [226] Phillip Weinberg and Marin Bukov. QuSpin: a Python package for dynamics and exact diagonalisation of quantum many body systems part I: Spin chains. *SciPost Phys.*, 2:003, 2017.
- [227] Hidemaro Suwa. Reducing rejection exponentially improves Markov chain Monte Carlo sampling. *Physica A: Statistical Mechanics and its Applications*, 633:129368, 2024.
- [228] Heinz-Jürgen Schmidt, Andre Lohmann, and Johannes Richter. Eighth-order high-temperature expansion for general Heisenberg Hamiltonians. *Phys. Rev. B*, 84:104443, Sep 2011.

- [229] J D Alzate-Cardona, D Sabogal-Suárez, R F L Evans, and E Restrepo-Parra. Optimal phase space sampling for Monte Carlo simulations of Heisenberg spin systems. *Journal of Physics: Condensed Matter*, 31(9):095802, Jan 2019.
- [230] Piotr Białas, Piotr Korcyl, and Tomasz Stebel. Analysis of autocorrelation times in neural markov chain monte carlo simulations. *Phys. Rev. E*, 107:015303, Jan 2023.
- [231] Fabian Joswig, Simon Kuberski, Justus T. Kuhlmann, and Jan Neuendorf. pyerrors: A python framework for error analysis of monte carlo data. *Computer Physics Communications*, 288:108750, 2023.
- [232] Neal Madras and Alan D. Sokal. The pivot algorithm: A highly efficient Monte Carlo method for the self-avoiding walk. *Journal of Statistical Physics*, 50(1):109–186, Jan 1988.
- [233] Piotr Białas, Piotr Korcyl, and Tomasz Stebel. Analysis of autocorrelation times in neural Markov chain Monte Carlo simulations. *Phys. Rev. E*, 107:015303, Jan 2023.
- [234] Meng Yao, Da Wang, and Qiang-Hua Wang. Reducing autocorrelation time in determinant quantum Monte Carlo using the Wang-Landau algorithm: Application to the Holstein model. *Phys. Rev. E*, 104:025305, Aug 2021.
- [235] Robert H. Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, 58:86–88, Jan 1987.
- [236] Ulli Wolff. Collective Monte Carlo updating for spin systems. *Phys. Rev. Lett.*, 62:361–364, Jan 1989.
- [237] Thomas D. Barrett, Aleksei Malyshev, and A. I. Lvovsky. Autoregressive neural-network wavefunctions for ab initio quantum chemistry. *Nature Machine Intelligence*, 4(4):351–358, Apr 2022.
- [238] Dian Wu, Lei Wang, and Pan Zhang. Solving statistical mechanics using variational autoregressive networks. *Phys. Rev. Lett.*, 122:080602, Feb 2019.
- [239] Benigno Uribe, Marc-Alexandre Côté, Karol Gregor, Iain Murray, and Hugo Larochelle. Neural autoregressive distribution estimation. *J. Mach. Learn. Res.*, 17(1):7184–7220, Jan 2016.
- [240] Dian Wu, Riccardo Rossi, and Giuseppe Carleo. Unbiased Monte Carlo cluster updates with autoregressive neural networks. *Phys. Rev. Res.*, 3:L042024, Nov 2021.
- [241] Tom Vieijra and Jannes Nys. Many-body quantum states with exact conservation of non-Abelian and lattice symmetries through variational Monte Carlo. *Phys. Rev. B*, 104:045123, Jul 2021.

Appendix A

Schrödinger Equation on the Variational Manifold of the Projective Hilbert Space

To clarify the validity of the calculation to get from Eq. (4.23) to Eq. (4.24), the steps are explained here in detail. Multiplying with $\langle V_{\mathbb{W},\gamma}(\tau)|$ from the left of Eq. (4.23) and using the definition of the projector $P_{\Psi_{\mathbb{W}}(\tau)}$ in Eq. (3.23) leads to the following equation:

$$\frac{\partial \mathbb{W}^\alpha}{\partial \tau} \langle V_{\mathbb{W},\gamma} | V_{\mathbb{W},\alpha} \rangle = - \frac{(G^{-1})^{\alpha\mu}}{\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle} \langle V_{\mathbb{W},\gamma} | V_{\mathbb{W},\alpha}(\tau) \rangle \langle V_{\mathbb{W},\mu} | H | \Psi_{\mathbb{W}} \rangle \quad (\text{A.1})$$

It is stressed that the Einstein sum convention is used. On both sides, the definition of the QGT in Eq. (3.16) is applied and the definition of the projector Q_Ψ in Eq. (3.6) is inserted:

$$G_{\gamma\alpha} \frac{\partial \mathbb{W}^\alpha}{\partial \tau} \langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle = -G_{\gamma\alpha} (G^{-1})^{\alpha\mu} \langle \partial_\mu \Psi_{\mathbb{W}} | (H - E) | \Psi_{\mathbb{W}} \rangle \quad (\text{A.2})$$

Dividing by $\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle$ allows to identify the forces, defined in Eq. (4.9), on the right hand side. Additionally, the QGT $G_{\gamma\alpha}$ appears on both sides of the equation and is thus cancelled.

$$\frac{\partial \mathbb{W}^\alpha}{\partial \tau} = -(G^{-1})^{\alpha\mu} F_\mu \quad (\text{A.3})$$

Appendix B

Schrödinger Equation on the Variational Manifold of the Hilbert Space

The computation in appendix A is analogously carried out for the variational manifold of the Hilbert space $\mathcal{V}(\mathcal{H}, \mathbb{C})$. The variational Schrödinger equation in Eq. (4.23) is adjusted to

$$\frac{d}{d\tau} |\Psi_{\mathbb{W}}(\tau)\rangle\rangle = \sum_{\alpha} \frac{\partial \mathbb{W}_{\alpha}}{\partial \tau} |\hat{V}_{\mathbb{W},\alpha}(\tau)\rangle\rangle = -\hat{P}_{\Psi_{\mathbb{W}}(\tau)} H |\Psi_{\mathbb{W}}(\tau)\rangle\rangle, \quad (\text{B.1})$$

where the local basis of the tangent space is now given as in Eq. (3.29) by

$$|\hat{V}_{\mu,\mathbb{W}}\rangle\rangle = |\partial_{\mu} \Psi_{\mathbb{W}}\rangle\rangle \quad (\text{B.2})$$

and thus differs from the one for the projective Hilbert space in Eq. (3.14) by the projector \mathbb{Q}_{Ψ} . This thus changes also the definition of the projector onto the tangent space in Eq. (3.31), which is denoted by $\hat{P}_{\Psi_{\mathbb{W}}(\tau)}$. Multiplying with $\langle\langle \hat{V}_{\mathbb{W},\gamma}(\tau) |$ from the left of Eq. (B.1) and using the definition of the projector $\hat{P}_{\Psi_{\mathbb{W}}(\tau)}$ leads to the following equation:

$$\frac{\partial \mathbb{W}^{\alpha}}{\partial \tau} \langle\langle \hat{V}_{\mathbb{W},\gamma} | \hat{V}_{\mathbb{W},\alpha} \rangle\rangle = -\frac{(\hat{G}^{-1})^{\alpha\mu}}{\langle\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle\rangle} \langle\langle \hat{V}_{\mathbb{W},\gamma} | \hat{V}_{\mathbb{W},\alpha}(\tau) \rangle\rangle \langle\langle \hat{V}_{\mathbb{W},\mu} | H | \Psi_{\mathbb{W}} \rangle\rangle \quad (\text{B.3})$$

It is stressed that the Einstein sum convention is used. On both sides, the definition of the QGT in Eq. (3.30) is applied:

$$\hat{G}_{\gamma\alpha} \frac{\partial \mathbb{W}^{\alpha}}{\partial \tau} \langle\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle\rangle = -\hat{G}_{\gamma\alpha} (\hat{G}^{-1})^{\alpha\mu} \langle\langle \partial_{\mu} \Psi_{\mathbb{W}} | H | \Psi_{\mathbb{W}} \rangle\rangle \quad (\text{B.4})$$

Dividing by $\langle\langle \Psi_{\mathbb{W}} | \Psi_{\mathbb{W}} \rangle\rangle$ allows to identify the forces, defined in Eq. (4.25), on the right hand side. Additionally, the QGT $\hat{G}_{\gamma\alpha}$ appears on both sides of the equation and is thus cancelled.

$$\frac{\partial \mathbb{W}^{\alpha}}{\partial \tau} = -(\hat{G}^{-1})^{\alpha\mu} \hat{F}_{\mu} \quad (\text{B.5})$$