

Arbeit zur Erlangung des akademischen Grades
Master of Science

Exploring Entanglement Hamiltonians with Variational Algorithms

Yanick Sebastian Kind
geboren in Farfield(Ohio)

2025

AG Fauseweh
Fakultät Physik
Technische Universität Dortmund

Erstgutachter: JProf. Dr. Benedikt Fauseweh
Zweitgutachter: Dr. Jörg Bünemann
Abgabedatum: 06.03.2025

Abstract

This work presents a quantum variational algorithm with which the Entanglement Hamiltonian of lattice systems can be learned. It was first introduced by Kokail *et al.* The Ansatz is motivated by the Bisognano-Wichmann theorem from quantum field theory, which will be extended to quantum critical systems with conformal symmetry. The fundament is an optimized implementation in the programming language `julia`. An important part of this work is to provide a complete understanding of this algorithm and to investigate its convergence properties. With the knowledge of convergence, accurate results for the thermodynamic limit are obtained. The limitations of the Bisognano-Wichmann theorem will be shown and corrections to this theorem will be discussed.

Kurzfassung

In dieser Arbeit wird ein quanten-variationeller Algorithmus vorgestellt, mit welchem der Entanglement Hamiltonian von Gittersystemen gelernt werden kann. Der Algorithmus stammt von Kokail *et al.* Das Bisognano-Wichmann Theorem aus der Quantenfeldtheorie liefert einen variationellen Ansatz, welcher auch auf quantenkritische Systeme mit konformer Symmetrie erweitert wird. Als Fundament dient eine optimierte Implementierung in der Programmiersprache `julia`. Ein wichtiger Teil dieser Arbeit ist es, vollständiges Verständnis von diesem Algorithmus zu bieten und aufzuzeigen, wie konvergierte Ergebnisse erhalten werden können. Mit den Erkenntnissen der Konvergenz können genau Ergebnisse für den thermodynamischen Limes erzielt werden. Es werden die Einschränkungen des Bisognano-Wichmann Theorems aufgezeigt und Korrekturen zu dem Theorem diskutiert.

Contents

1	Introduction	1
2	Theoretical foundation	3
2.1	Bipartite entanglement	3
2.2	Schmidt decomposition	3
2.3	Reduced density matrix and Entanglement Hamiltonian	4
2.4	Bisognano-Wichmann theorem	5
2.5	The original algorithm	7
2.6	Improvement of the cost function	9
2.7	Variational Ansätze	10
2.8	Computational basis	11
2.9	Quantum phase transition and scale invariance	12
2.10	Transverse field Ising model	13
2.11	XXZ model	14
3	Implementation and runtime optimization	15
3.1	General considerations	15
3.2	Numerical integration method	17
3.3	Gradient	20
4	Results: Convergence	25
4.1	Cost function and convergence of the midpoint rule for the BW-like Ansatz	25
4.2	Influence of the maximum observation time for the BW-like Ansatz	28
4.3	Convergence properties of the BW-violating Ansatz	35
4.4	Benchmarks	37
5	Results: Physics	41
5.1	Violation of the BW theorem	41
5.2	Long range corrections in the XXZ model	46
5.3	Comparison of the Entanglement spectra	48
5.4	TFIM and XXZ model across the phase diagram	51
5.5	Violation of the BW theorem in the XXZ model in the thermodynamic limit	53

6 Summary and outlook	59
Bibliography	61

1 Introduction

This work deals with the simulation of quantum systems using the tools of modern quantum information theory. The simulation of quantum many-body systems is one of the key challenges addressed with a quantum computer. Computational access to quantum many-body systems to gain knowledge about e.g. complex phase diagrams or non-equilibrium quantum systems, is useful to advance in these research areas. However, accessing large systems poses a significant challenge with classical computers. The reason for that hurdle is the tensor product structure of the Hilbert space of composite quantum systems, resulting in an exponential growth of the Hilbert space dimension. This work deals exclusively with spins-1/2 systems, i.e., with two-level systems (qubits). Its individual Hilbert space dimension is 2. Each added qubit doubles the composite system Hilbert space dimension, leading to a dimension of 2^N , where N is the number of qubits [12].

State-of-the-art quantum computers are limited by the number of qubits, lack of connectivity between the qubits as well as coherent and incoherent errors reducing the possible computation time and the reliability of the results [5]. These devices are part of the Noisy Intermediate-Scale Quantum (NISQ) era [42]. To make use of the current NISQ devices, despite the noise, the Variational Quantum Algorithms[5] (VQAs) are a promising strategy. Within the framework of such algorithms, classical computers are used for optimization of a cost function, while the quantum time evolution is done on a quantum device. VQAs use parameterized quantum circuits to be run on the quantum computer and then utilize a classical computer to optimize the parameters. This approach has the advantage that the circuit depths can be held shallow and errors can be mitigated [5].

Part of the VQAs is Hamiltonian learning, where, as its name suggests, Hamiltonians are learned with parameterized quantum circuits. Specifically, here, the Entanglement Hamiltonian (EH) is learned by leveraging quantum time evolution under a variational Ansatz for the EH. The underlying algorithm was first presented by Kokail *et al.* [26]. The quantity of interest is the eigenspectrum of the EH, the Entanglement Spectrum (ES), which yields rich information about entanglement in quantum many-body systems [29]. For example, Li and Haldane[29] proposed that the ES works as a generalization of the Entanglement Entropy, leading to the identification of topological order. To this day, entanglement in quantum many-body systems is an open topic in research. Its experimental measurement is often difficult. However, there are some procedures for the quantification and detection

of multipartite entanglement, such as the Quantum Fisher Information, shown by Hauke *et al.*[17].

In general, analytical results for the EH are very difficult to obtain. For quantum field theories (QFTs) and certain geometries, the Bisognano-Wichmann (BW) theorem[3, 4] provides an exact result. However, lattice systems are, in general, not described by QFTs[32] and the form of the EH on lattice systems and the applicability of the BW theorem to lattices still remains an open question, which is thoroughly addressed in this work.

2 Theoretical foundation

2.1 Bipartite entanglement

Given a composite Hilbert space $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, composed of two subsystems A and B with its respective Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ of dimensions $d_A = \dim(\mathcal{H}_A)$ and $d_B = \dim(\mathcal{H}_B)$, spanned by the orthonormal bases $\{|\mu_A^i\rangle\}$ and $\{|\mu_B^j\rangle\}$, a general pure state $|\Psi\rangle \in \mathcal{H}$ can be written as

$$|\Psi\rangle = \sum_{i=1}^{d_A} \sum_{j=1}^{d_B} M_{ij} |\mu_A^i\rangle \otimes |\mu_B^j\rangle, \quad (2.1)$$

where the rank $\chi \leq \min(d_A, d_B)$ of the complex matrix M is called the Schmidt rank [19]. Since the dimensions of the sub-Hilbert spaces can in general differ, the matrix M , sometimes called the entanglement matrix, is rectangular in general and obeys

$$\sum_{i=1}^{d_A} \sum_{j=1}^{d_B} |M_{ij}|^2 = 1,$$

to ensure normalization of the state $|\Psi\rangle$. In case of $\chi = 1$, the state $|\Psi\rangle$ takes the form of a simple product state

$$|\Psi\rangle = |\Psi_A\rangle \otimes |\Psi_B\rangle$$

and is said to be separable, entangled otherwise.

2.2 Schmidt decomposition

Recalling a general pure state describing a composite system (see Equation (2.1)), the entanglement matrix M can be brought into a diagonal form D via a singular value decomposition (SVD)[43]

$$M = UDV^\dagger.$$

The matrices U and V are of size $d_A \times \min(d_A, d_B)$ and $d_B \times \min(d_A, d_B)$, respectively, and obey $U^\dagger U = \mathbb{1}$ and $V V^\dagger = \mathbb{1}$ [43]. The non-negative entries (the singular values of M) of the diagonal matrix D with dimension $\min(d_A, d_B)$ are called Schmidt-coefficients[38] and can be expressed as $e^{-\xi_\alpha/2}$ (the explicit choice of this representation should become clear in Section 2.3). Using the SVD, Equation (2.1) reads

$$|\Psi\rangle = \sum_{i=1}^{d_A} \sum_{j=1}^{d_B} \sum_{\alpha=1}^{\min(d_A, d_B)} e^{-\xi_\alpha/2} U_{i\alpha} V_{j\alpha}^* |\mu_A^i\rangle \otimes |\mu_B^j\rangle.$$

Defining a new orthonormal basis set $\{|\Phi_A^\alpha\rangle = \sum_{i=1}^{d_A} U_{i\alpha} |\mu_A^i\rangle\}$ and $\{|\Phi_B^\alpha\rangle = \sum_{j=1}^{d_B} V_{j\alpha}^* |\mu_B^j\rangle\}$ yields[43]

$$|\Psi\rangle = \sum_{\alpha=1}^{\chi} e^{-\xi_\alpha/2} |\Phi_A^\alpha\rangle \otimes |\Phi_B^\alpha\rangle, \quad (2.2)$$

where $\{\xi_\alpha\}$ will be referred to as the Entanglement Spectrum (ES) (see Section 2.3). Since the rank is preserved under a SVD, the number of non-zero singular values coincides with the Schmidt rank χ [27], and thus, the sum in Equation (2.2) is restricted to χ . The lower and upper bound of summation will be dropped from now on as long as it is unambiguous.

2.3 Reduced density matrix and Entanglement Hamiltonian

Given a pure state $\hat{\rho} = |\Psi\rangle\langle\Psi|$, which describes the composite system $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, after a Schmidt decomposition as in Equation (2.2), the reduced density matrix (RDM) on a subsystem A after tracing out the degrees of freedom related to subsystem B is defined as

$$\hat{\rho}_A = \text{Tr}_B [\hat{\rho}] = \sum_{\alpha} e^{-\xi_\alpha} |\Phi_A^\alpha\rangle\langle\Phi_A^\alpha| = e^{-\hat{H}_A}, \quad (2.3)$$

parameterized by the EH $\hat{H}_A$ [26, 44]. The EH and its non-negative eigenvalues $\{\xi_\alpha\}$, the ES, completely characterize all correlations in partition A and reveal much more than the entanglement entropy [29] or the entanglement witness [45]. On the one hand, the ES can be used to detect quantum phase transitions as for example in spin models, where quantum phase transitions are signalled by a singular behaviour of the Schmidt gap (difference between the two largest eigenvalues of the reduced density matrix) [9, 28, 45]. On the other hand, the ES can be utilized to identify topological order [13, 29, 41, 45]. In general, it is hard to derive an analytical form of the EH especially for lattice theories. The BW theorem (Section 2.4) delivers one way to obtain the EH analytically for specific cases for QFTs.

2.4 Bisognano-Wichmann theorem

In a $d + 1$ -dimensional relativistic QFT with a local Hamiltonian-density $\hat{\mathcal{H}}(\mathbf{x})$, the EH of the ground state for the special case of a bipartition of an infinite system A ($A = \{\mathbf{x} \in \mathbb{R}^d | x_1 > 0\}$) is

$$\hat{H}_A = \int_A d^d x \beta(x_1) \hat{\mathcal{H}}(\mathbf{x}) + c' \quad (2.4)$$

with $\beta(x_1) = \frac{2\pi}{c} x_1$ [8, 14, 26], whereby the “speed of sound” c of the underlying QFT is set to unity from now on. The constant c' ensures the normalization $\text{Tr}[\hat{\rho}_A] = 1$. This is the seminal BW-theorem, which is exact for QFTs. In Equation (2.4) it becomes apparent that the EH is a deformation of the system Hamiltonian [26]. Additionally, the RDM $\hat{\rho}_A$ can be interpreted as a thermal state with a locally varying entanglement temperature, which is very high near the entanglement cut (boundary between both partitions) and decreases with $1/x_1$ away from it [26].

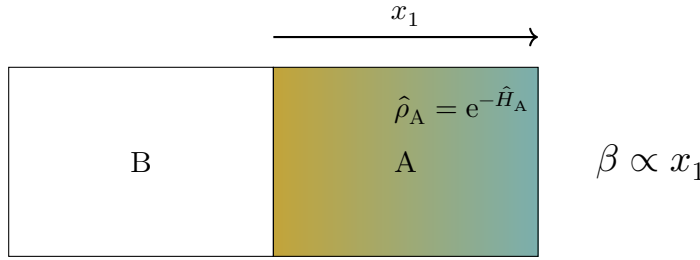


Figure 2.1: Interpretation of the reduced density matrix $\hat{\rho}_A$ as a thermal state with a locally varying temperature, the entanglement temperature. The inverse entanglement temperature takes the form of a linear ramp, and thus, the entanglement temperature decreases as $\propto 1/x_1$, indicated by the color gradient from orange to blue.

For lattice systems, it is straightforward to propose a discretized version of Equation (2.4) s.t.

$$\hat{H}_A \approx \sum_{i \in A} g_i \hat{h}_i + c', \quad (2.5)$$

where the substitution $\beta(x_1) \rightarrow g_i$ and $\hat{\mathcal{H}}(\mathbf{x}) \rightarrow \hat{h}_i$ with $\hat{h}_i$ as a quasi-local few-body operator for the i -th lattice site is utilized. A natural question is whether the BW theorem works for lattice systems since it is defined for relativistic QFTs at first. Although the presence of a lattice breaks the Lorentz invariance[8] (even when it

is recovered as a low-energy symmetry[14]), numerical calculations[8, 14, 32, 44] suggest that the discretized version of the BW theorem (2.5) is often a good first approximation for lattice systems.

2.4.1 Conformal extensions

For systems, which have conformal symmetry in addition to Lorentz invariance, the BW theorem (Equation (2.4)) can be extended to different geometries [14]. Since this work only deals with one dimensional systems, only conformal extensions for one spatial dimension will be listed.

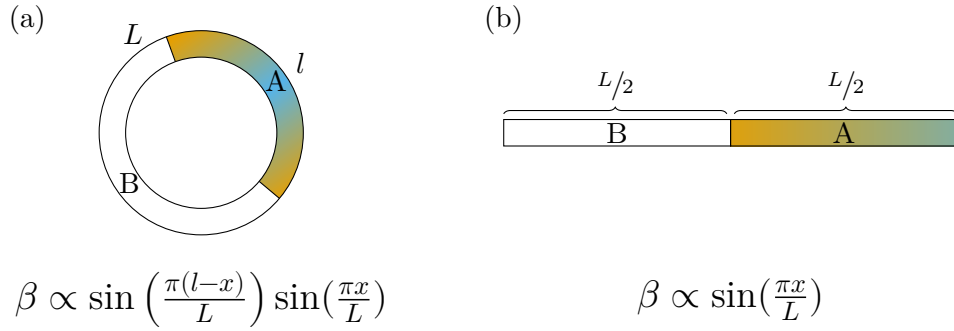


Figure 2.2: The BW theorem extended to one dimensional systems with conformal symmetry. Schematic representation of the entanglement temperature for (a) a subsystem of length l embedded in a system of length L with periodic boundary conditions and (b) a subsystem of length $L/2$ embedded at a boundary of an open system of length L . The colors indicate a high (orange) and low (blue) entanglement temperature.

In case of a finite subsystem of length l in a ring of circumference L , the EH is given by

$$\hat{H}_A^{\text{CFT1}} = 2L \int_0^l dx \frac{\sin\left(\frac{\pi(l-x)}{L}\right) \sin\left(\frac{\pi x}{L}\right)}{\sin\left(\frac{\pi l}{L}\right)} \hat{\mathcal{H}}(x) + c'. \quad (2.6)$$

Since the system obeys periodic boundary conditions (PBC), there are two entanglement cuts, where the inverse temperature rises approximately linearly for small distances from the entanglement cut in agreement with the BW theorem (Equation (2.4)). For a finite partition of length $L/2$ at the edge of a finite open system of length L , the EH reads

$$\hat{H}_A^{\text{CFT2}} = 2L \int_0^{L/2} dx \sin\left(\frac{\pi x}{L}\right) \hat{\mathcal{H}}(x) + c', \quad (2.7)$$

again with a linear rise of the inverse entanglement temperature near the entanglement cut. For a finite subsystem of length l in an infinite composite system, the EH is given by

$$\hat{H}_A^{\text{CFT3}} = 2\pi \int_0^l dx x \left(\frac{l-x}{l} \right) \hat{\mathcal{H}}(x) + c'. \quad (2.8)$$

2.5 The original algorithm

The following algorithm was first presented in Reference [26] and the information in the following is taken from that article.

The main goal of the algorithm is to learn the EH via a hybrid quantum-classical feedback loop (QCFL) utilizing the variational Ansatz $\hat{H}_A^{\text{Var}}(\mathbf{g}) = \sum_i g_i \hat{h}_i$, which acts as a generator for the time evolution operator

$$\hat{U}_A(\mathbf{g}, t) = e^{-i\hat{H}_A^{\text{Var}}(\mathbf{g})t},$$

acting on subsystem A for some time t s.t.

$$\hat{\rho}_A \rightarrow \hat{U}_A(\mathbf{g}, t) \hat{\rho}_A \hat{U}_A^\dagger(\mathbf{g}, t).$$

The parameters g_i act as variational parameters. The QCFL works as follows:

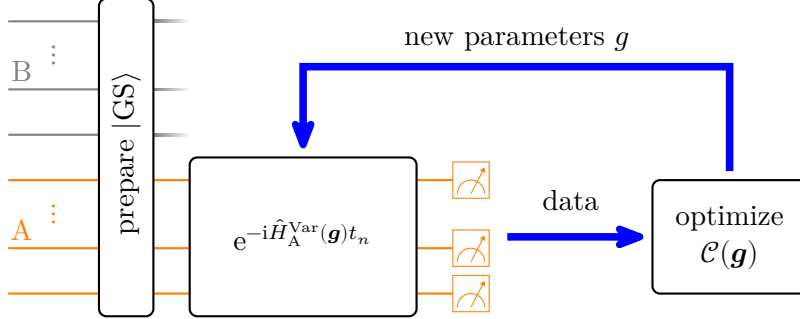


Figure 2.3: Quantum classical feedback loop (QCFL). The composite system is initialized with the ground state, $|\text{GS}\rangle$, of the system Hamiltonian. The subsystem A is then evolved under the variational Ansatz $\hat{H}_A^{\text{Var}}(\mathbf{g})$ and some observables $\langle \hat{\mathcal{O}}_j^A \rangle_{t_n}$ are measured at time instances $\{t_n\}$. The cost function is then evaluated with the measurements and the new parameters suggested by the optimizer are used to repeat the procedure.

1. Prepare an initial state $\hat{\rho}_A = \text{Tr}_B [|\text{GS}\rangle\langle\text{GS}|]$ with $|\text{GS}\rangle$ as the ground state of the composite system.

2 Theoretical foundation

2. Evolve the subsystem A under the variational Ansatz for some time $t_n > 0$, leaving the complementary subsystem untouched.
3. Evaluate the expectation values $\langle \hat{\mathcal{O}}_j^A \rangle_{t_n}$ after each time t_n .
4. Calculate a suitable cost function $\mathcal{C}(\mathbf{g})$.
5. Repeat step 2 to 4 for different variational parameters and minimize $\mathcal{C}(\mathbf{g})$.

The expectation value after the subsystem A has been evolved under the variational Ansatz reads

$$\langle \hat{\mathcal{O}}_j^A \rangle_{t_n} = \text{Tr}_A \left[\hat{\mathcal{O}}_j^A \hat{U}_A(\mathbf{g}, t_n) \hat{\rho}_A \hat{U}_A^\dagger(\mathbf{g}, t_n) \right],$$

where the operators $\hat{\mathcal{O}}_j^A$ are only defined on subsystem A and are restricted to be (quasi-)local. The optimal parameters $\mathbf{g}^{\text{opt}}$ are learned by minimizing the time variation of the observables s.t. $\langle \hat{\mathcal{O}}_j^A \rangle_{t_n} = \text{const.}$ A suitable cost function left to be minimized is given as

$$\mathcal{C}(\mathbf{g}) = \sum_{j=1}^{N_O} \sum_{n=1}^{N_T} \left(\langle \hat{\mathcal{O}}_j^A \rangle_{t_n} - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2$$

with N_O as the number of observables and N_T as the number how often the subsystem A is evolved and each observable is measured. For sufficiently many observation times t_n and observables $\hat{\mathcal{O}}_j^A$, a cost function value of zero implies

$$[\hat{H}_A^{\text{Var}}(\mathbf{g}^{\text{opt}}), \hat{H}_A] = 0, \quad (2.9)$$

where $\hat{H}_A$ is the exact EH and $\mathbf{g}^{\text{opt}}$ are the optimal variational parameters. Equivalently, a cost function value of zero implies $[\hat{H}_A^{\text{Var}}(\mathbf{g}^{\text{opt}}), \hat{\rho}_A] = 0$, too, since the exact RDM $\hat{\rho}_A = \exp(-\hat{H}_A)$ is given by a power series in $\hat{H}_A$. This results in a thermalized subsystem A and the observables are constant in time. The precise choice of observables is not crucial, since an operator is expected to evolve into a complex operator under the dynamics as long as $[\hat{H}_A^{\text{Var}}(\mathbf{g}), \hat{\mathcal{O}}_j^A] \neq 0$. Since the aforementioned commutator is still fulfilled if a solution $\mathbf{g}^{\text{opt}}$ is scaled by a factor γ , the scale factor remains undetermined by the algorithm as well as the normalization constant c' (see Equation (2.5)). To compare the ES of the variational solution and the exact ES, the universal ratios

$$\kappa_\alpha = \frac{\xi_\alpha - \xi_{\alpha_0}}{\xi_{\alpha_1} - \xi_{\alpha_0}} \quad (2.10)$$

are defined s.t. the undetermined scaling factor γ and the normalization constant c' are eliminated by division and subtraction, respectively.

2.6 Improvement of the cost function

The original algorithm, introduced in the previous section, provides a good way to determine the EH of lattices systems, which is hard to obtain otherwise. However, throughout the investigation of this algorithm, some difficulties, and thus, possibilities to improve the algorithm have been noticed. A reliable algorithm should provide converged results (in this case the optimal variational parameters $\mathbf{g}^{\text{opt}}$), which could not be achieved with the algorithm initially presented. A major challenge is determining how to select the observation times t_n , which can be easily solved by choosing another interpretation of the cost function. This section addresses this problem and presents the improved interpretation of the cost function of this work. Numerical examples and benchmarks in regard to convergence will be presented in the results (see Chapter 4).

Recalling the cost function from Section 2.5

$$\mathcal{C}(\mathbf{g}) = \sum_{j=1}^{N_O} \sum_{n=1}^{N_T} \left(\langle \hat{\mathcal{O}}_j^A \rangle_{t_n} - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2,$$

it is hard to compare numerical values of the cost function, since it is not normalized to the number of observables N_O and to the number of observation times N_T , which is easily fixed by diving by these aforementioned quantities. Since the algorithm is based on monitoring observables, it is, in general, not enough to choose a few arbitrary discrete time points. Otherwise, the variational parameters $\mathbf{g}^{\text{opt}}$ will not be converged. Assuming equidistant time points i.e. a step size Δt for the observation times, the cost function can be, together with the aforementioned normalization, rewritten as

$$\mathcal{C}(\mathbf{g}) = \frac{\Delta t}{T_{\max} N_O} \sum_{j=1}^{N_O} \sum_{n=1}^{N_T} \left(\langle \hat{\mathcal{O}}_j^A \rangle_{n\Delta t} - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2, \quad (2.11)$$

defining the maximum observation time $T_{\max} = N_T \Delta t$. To obtain a cost function, monitoring not at discrete time points but at all times, the discrete sum is replaced by an integral. That is, in the continuum limit $\Delta t \rightarrow 0$ with $T_{\max} = \text{const}$, then

$N_T \rightarrow \infty$, the cost function reads

$$\begin{aligned}
 \mathcal{C}(\mathbf{g}) &= \lim_{\Delta t \rightarrow 0} \frac{1}{T_{\max}} \sum_{n=1}^{N_T} \frac{1}{N_O} \sum_{j=1}^{N_O} \left(\langle \hat{\mathcal{O}}_j^A \rangle_{n\Delta t} - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2 \Delta t \\
 &= \frac{1}{T_{\max}} \int_0^{T_{\max}} \frac{1}{N_O} \sum_{j=1}^{N_O} \underbrace{\left(\langle \hat{\mathcal{O}}_j^A \rangle_t - \langle \hat{\mathcal{O}}_j^A \rangle_0 \right)^2}_{:=c(\mathbf{g},t)} dt \\
 &= \frac{1}{T_{\max}} \int_0^{T_{\max}} c(\mathbf{g}, t) dt, \tag{2.12}
 \end{aligned}$$

which boils down to the mean value of the integrand $c(\mathbf{g}, t)$ over an interval $[0, T_{\max}]$. Still, one open degree of freedom to properly choose remains, namely the maximum observation time $T_{\max}$. The influence of $T_{\max}$ will be thoroughly discussed in Section 4.2. Besides the choice of $T_{\max}$, the only remaining challenge is to evaluate the integral as accurately and as quickly as possible, which will be discussed in the implementation and runtime optimization Chapter 3.

2.7 Variational Ansätze

In the previous Section 2.5 the terminology “variational Ansatz”, denoted as $\hat{H}_A^{\text{Var}}(\mathbf{g})$, is already introduced. There are two variational Ansätze in this work. The first Ansatz is the BW-like Ansatz, denoted as $\hat{H}_A^{\text{BW}}$, which is used in Reference [26]. As the name suggests, it follows the BW theorem. The second Ansatz, $\hat{H}_A^{\text{BWV}}$, is used to show a violation of the BW theorem in lattice models. From now on, if $\hat{H}_A^{\text{Var}}(\mathbf{g})$ is written, both Ansätze, $\hat{H}_A^{\text{BW}}$ and $\hat{H}_A^{\text{BWV}}$, are addressed. This will be of use for general derivations such as the derivation of the gradient. In general, both Ansätze are given by a linear combination

$$\hat{H}_A^{\text{Var}}(\mathbf{g}) = \sum_i g_i \hat{h}_i, \tag{2.13}$$

where g_i is a variational parameter and $\hat{h}_i$ is a quasi-local few-body operator, which will be referred to as a block. The subsystem A will always be on the right border of the composite system.

2.7.1 BW-like Ansatz

The BW theorem predicts that the EH is a spatially deformed version of the system Hamiltonian on a subsystem. That is, each lattice site i is assigned a block $\hat{h}_i$

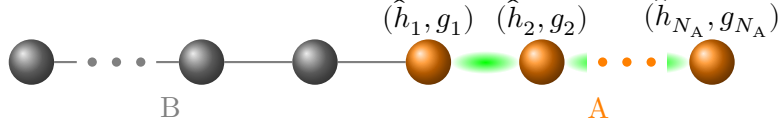


Figure 2.4: Schematic illustration of the variational Ansatz $\hat{H}_A^{\text{BW}} = \sum_i g_i \hat{h}_i$. Each lattice site in the subsystem A is assigned a few-body quasi-local operator $\hat{h}_i$ together with a variational parameter g_i . Only interactions within the subsystem A are taken into account, as suggested by the green highlighting.

together with one variational parameter g_i , as illustrated in Figure 2.4. That means that the index i in Equation (2.13) coincides with the i -th lattice site in the subsystem A. The blocks $\hat{h}_i$ are not local and act on more than one qubit. It is important to note that all interactions are restricted to be within subsystem A as well.

2.7.2 BW-violating Ansatz

The BW-violating Ansatz is not given by a spatially deformed Hamiltonian. Thus, each lattice site is assigned multiple blocks $\hat{h}_i$ and multiple variational parameters g_i and the index i in Equation (2.13) does not coincide with the lattice site i . From now on, the dependence of the variational Ansätze on the variational parameters $\mathbf{g}$ will be omitted. The explicit form of the blocks will become clear in the Sections 2.10 and 2.11, where the transverse field Ising model (TFIM) and the XXZ model will be discussed.

2.8 Computational basis

The basis used in this work is the commonly used basis in quantum simulation/-computing, namely the computational basis spanned by the basis states $|\uparrow\rangle$ and $|\downarrow\rangle$, corresponding to “spin up” and “spin down” respectively. These states are the eigenstates of the z -component of the spin-1/2 operator. In most quantum computing literature, these states are denoted as $|0\rangle$ and $|1\rangle$, but since this work deals exclusively with spin models, the first notation is used. In the computational basis, sometimes called z -basis in the context of spin models, the spin-1/2 operators take the simple form

$$\hat{S}_x = \frac{1}{2}X = \frac{1}{2} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{S}_y = \frac{1}{2}Y = \frac{1}{2} \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{S}_z = \frac{1}{2}Z = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

with $\{X, Y, Z\}$ as the three pauli matrices. All spin models will be expressed in terms pauli matrices, i.e. in the computational basis.

2.9 Quantum phase transition and scale invariance

A phase transition exhibits a qualitative change of equilibrium properties and separates two states of matter with different characteristics. It is a point in parameter space, at which the a system undergoes change from a disordered to an ordered phase (or vice versa). The qualitative change can be detected by an order parameter, which is finite in the ordered phase and zero in the disordered phase. As an example, the magnetization could act as an order parameter in magnetic systems. There are two kinds of phase transitions, the continuous (second order) and discontinuous (first order) phase transition. Both are given their name by a continuous variation and a discontinuity of the order parameter at the phase transition, respectively. Phase transitions can be driven by thermal fluctuations at finite temperatures T hence giving it the name thermal phase transition (TPT).

A quantum phase transition (QPT) is a phase transition at temperature $T = 0$. Therefore, it is not driven by thermal fluctuations but by quantum fluctuations. A QPT is induced by changes in control parameters such as pressure, magnetic

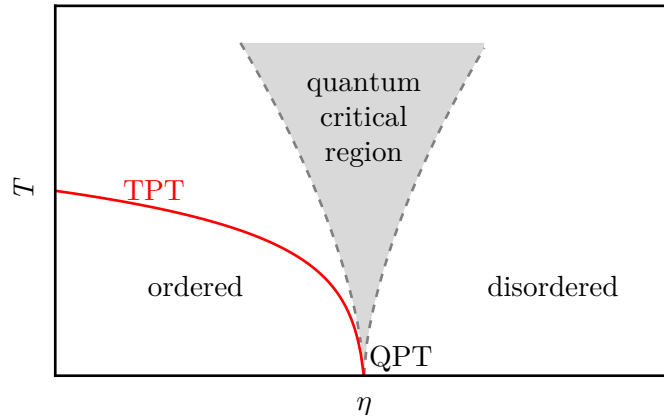


Figure 2.5: Schematic drawing of a phase diagram with a control parameter η . A thermal phase transition (TPT), which separates the ordered state from the disordered state, is drawn by the red line. At $T = 0$ the transition is a quantum phase transition (QPT). The characteristics of the QPT translate into the quantum critical region for finite temperature.

field strengths, etc. [20]. However, there is some caution needed with the term “driven by quantum fluctuations”, since a quantum mechanical system is described

by single coherent many-body wavefunction. Deviations from a reference state, e.g. an ordered magnet, is the most appropriate use of the term “fluctuations”. Despite being experimentally unavailable, a QPT is still physically relevant, because it influences systems properties even at finite T , like the scaling behaviour of the heat capacity. A continuous phase transition sets in at the critical point at which the correlation length diverges. At the critical point, the correlation length is the only relevant length scale governing the low-energy physics [48]. Now, the importance of the (quantum) criticality for this work is the accompanying scale invariance, which exists since there is no fixed length scale in critical systems. The importance of the scale invariance is that QFTs are scale invariant. Thus, for the accurate applicability of the BW theorem, which holds for QFTs, to lattice systems, the systems are mostly considered when they are critical.

2.10 Transverse field Ising model

The Hamiltonian of the TFIM with N sites, open boundary conditions (OBC), nearest neighbour coupling strength J and transverse field strength Γ reads[14]

$$\hat{H} = -J \sum_{i=1}^{N-1} Z_i Z_{i+1} - \Gamma \sum_{i=1}^N X_i. \quad (2.14)$$

The first term favors a ferromagnetic state for $J > 0$ and an antiferromagnetic state for $J < 0$ while the transverse field introduces fluctuations s.t. an orientation along the x -axis is favored by the transversal term. It possesses a $\mathbb{Z}_2$ symmetry, where the Hamiltonian is invariant under flipping all spins, i.e.

$$Z_i \rightarrow -Z_i.$$

In the limit $J \gg \Gamma$, the ground state is two-fold degenerate and the system is fully polarized with all spins pointing either up or down

$$|\text{GS}\rangle = \bigotimes_{i=1}^N |\uparrow\rangle \quad \text{or} \quad |\text{GS}\rangle = \bigotimes_{i=1}^N |\downarrow\rangle,$$

breaking the $\mathbb{Z}_2$ symmetry spontaneously, whereas all spins are completely aligned in the x -direction in the limit $\Gamma \gg J$

$$|\text{GS}\rangle = \bigotimes_{i=1}^N \underbrace{\frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)}_{|\rightarrow\rangle},$$

exhibiting a paramagnetic behaviour. The TFIM has a quantum critical point at $J/\Gamma = 1$, separating the ordered ferromagnetic and the disordered paramagnetic

phase [11]. From now on, $J = 1$ holds. In the case of the Ansatz $\hat{H}_A^{\text{BW}}$, one block of site i is given by

$$\hat{h}_i = -\frac{1}{2} \sum_{j \in \langle j, i \rangle \cap A} Z_j Z_i - \Gamma X_i,$$

where $\langle j, i \rangle \cap A$ denotes nearest neighbour coupling only if i and j are in the subsystem A. One block for the Ansatz $\hat{H}_A^{\text{BWV}}$ reads

$$\hat{h}_i \in \{-Z_j Z_{j+1} | 1 \leq j < N_A - 1\} \cup \{-\Gamma X_j | 1 \leq j < N_A\}.$$

That is, the complete BW-violating Ansatz for the TFIM is given by

$$\hat{H}_A^{\text{BWV}} = - \sum_{i=1}^{N_A-1} J_{i,i+1} Z_i Z_{i+1} - \sum_{i=1}^{N_A} \Gamma_i X_i. \quad (2.15)$$

with $J_{i,i+1}$ and Γ_i as variational Parameters.

2.11 XXZ model

The Hamiltonian of the XXZ model with N lattice sites and OBC is defined as

$$\hat{H} = \sum_{i=1}^{N-1} (X_i X_{i+1} + Y_i Y_{i+1} + \Delta Z_i Z_{i+1}),$$

where Δ is the anisotropy. For $\Delta = 1$, the isotropic case, the Heisenberg model is recovered. The XXZ model is ferromagnetic for $\Delta < -1$, quantum critical for $-1 < \Delta \leq 1$, exhibiting a Luttinger liquid phase, and antiferromagnetic for $\Delta > 1$ [14]. The phase transition at $\Delta = -1$ is of first order, s.t. the ferromagnetic state is exact for $\Delta < -1$, while the phase transition at $\Delta = 1$ is of second order [10]. Again in the ferromagnetic phase, the $\mathbb{Z}_2$ symmetry is spontaneously broken [14]. One block for the Ansatz $\hat{H}_A^{\text{BW}}$ reads

$$\hat{h}_i = \frac{1}{2} \sum_{j \in \langle j, i \rangle \cap A} (X_i X_j + Y_i Y_j + \Delta Z_i Z_j), \quad (2.16)$$

while the blocks for the BW-violating Ansatz are given by

$$\hat{h}_i \in \{X_j X_{j+1} + Y_j Y_{j+1} | 1 \leq j < N_A - 1\} \cup \{\Delta Z_j Z_{j+1} | 1 \leq j < N_A - 1\}.$$

Thus, the complete variational Ansatz $\hat{H}_A^{\text{BWV}}$ for the XXZ model reads

$$\hat{H}_A^{\text{BWV}} = \sum_{i=1}^{N_A-1} (J_{i,i+1}^{\text{XX}} (X_i X_{i+1} + Y_i Y_{i+1}) + J_{i,i+1}^Z \Delta Z_i Z_{i+1}). \quad (2.17)$$

3 Implementation and runtime optimization

For this package, the **julia** Programming Language[2] was chosen, because it offers a very readable syntax similar to python and if the code is properly written, the performance approaches that of programming languages like C. Its type system and multiple dispatch allowed to write very readable and generic but still short code, which can be extended to e.g. different variational Ansätze or lattice models. Another argument for **julia** is the extensible and efficient open-source Quantum Computing framework Yao.jl[31], whose purpose in this work is to construct Hamiltonians and density matrices conveniently. For optimization, the LBFGS-algorithm[40, 30], a gradient-based optimizer, from the package Optim.jl[33] is used. As a convergence criterium, the infinity norm of the gradient $\|\nabla_{\mathbf{g}} \mathcal{C}(\mathbf{g})\|_{\infty} = \max(|\partial \mathcal{C} / \partial g_1|, \dots, |\partial \mathcal{C} / \partial g_{N_p}|)$ is required to be less than 10^{-16} if not mentioned otherwise. Here, the number of parameters is denoted as N_p . This convergence criterium will be referenced to as ∇_{tol} . The ground state of the composite system Hamiltonian is extracted by exact diagonalization for less than 11 spins and with the Lanczos method, whose implementation from the package KrylovKit.jl[16] is used, for more than ten spins. The monitored observables are $\{Z_i Z_{i+1} | 1 \leq i < N_A - 1\}$.

A large percentage of the work was focused on runtime optimization, especially since the computation cost increases exponentially with the system size. The following sections list the most important parts, which led to significant performance increases.

3.1 General considerations

This section presents three measures, which are followed throughout the code and should be followed in most of the codes written in **julia**.

The first thing to mention is the type stability. Julia's type system is dynamic, and thus, the types are not known until runtime [23]. If the source code is not written properly, it can happen that the type of a local variable within a function cannot be inferred, which leads to tremendously increased runtimes. On the one hand, type stability can be ensured by e.g. making sure that the type of elements within an array can be inferred or explicitly stating the types of the elements within a struct. On the other hand, it is necessary to ensure that all types can be inferred, by

writing functions, which always return values of the same type. To give an example, consider the following function `f`.

```
function f(x::Float64, y::Int64)
    if x >= y
        return x
    else
        return y
    end
end
```

If `y` is larger than `x`, `y` is returned, which is an `Int64`. Otherwise `x`, a `Float64`, is returned. Thus, the function does not always return the same type and is said to be “type unstable”. This can be simply fixed by e.g. converting `y` to a `Float64`. If the type can always be inferred throughout the code, the code is said to be “type stable”.

The second point is memory allocation. Especially at the beginning of writing `Julia` code, it can happen that many temporary arrays or matrices are allocated. Memory allocation itself and its resulting garbage collection often act as bottlenecks. The function

```
function E(rho::Matrix{ComplexF64}, O::Matrix{ComplexF64})
    return tr(rho*O)
end
```

simply returns a quantum mechanical expectation value of an observable. The bottleneck is the matrix multiplication `*(Matrix{ComplexF64}, Matrix{ComplexF64})`, since it allocates an intermediate `Matrix`. This can be fixed by passing a preallocated buffer, `temp` in this case, to the function, in which the result of the matrix multiplication can be saved.

```
function E(rho::Matrix{ComplexF64}, O::Matrix{ComplexF64},
           temp::Matrix{ComplexF64})
    mul!(temp, rho, O)
    return tr(temp)
end
```

The function `mul!` is the in-place version of the function `*`, which mutates the first argument. As a result, the function `E` produces zero allocations and is therefore optimized. This is a tremendous performance improvement if this function is looped over hundreds or thousands of times, which in fact happens in this algorithm.

The third keypoint is the correct choice of implementations for matrix multiplication. Since most of the runtime is spent on matrix multiplication of dense and complex

matrices, scaling cubically with the Hilbert space dimension, it is crucial to choose the best implementation for the matrix multiplication, which are the routines of openBLAS[49, 50] in this work. openBLAS is a package, mainly written in Fortran and optimized over decades, offering highly optimized matrix and vector operations.

3.2 Numerical integration method

The cornerstone of the algorithm is the cost function. As stated in Section 2.6, the cost function boils down to the mean value of a continuous function, i.e., an integral over time. Since the optimizer calls the cost function, depending on the problem size, tens up to thousands of times, it is naturally important to obtain the most efficient evaluation of the integral. The previously mentioned general considerations in Section 3.1 ensure that the integrand $c(\mathbf{g}, t)$ is efficiently evaluated at one time point t .

The next challenge is to evaluate the integral as quickly and accurately as possible. Thus, a numerical integration technique, which needs as few as possible evaluations of the integrand to converge, is required. At the beginning of the work, the right point rule was used to determine the cost function as in Equation (2.11). The reason behind is that, in Reference [25], it was stated, as already pointed out in Section 2.6, that the cost function is evaluated at a few arbitrary time points. Thus, the first idea was to conveniently sample at a few equidistant time points. With that method, no convergence regarding the optimal parameters $\mathbf{g}^{\text{opt}}$ could have been achieved in an adequate computation time. Over time, the idea emerged to interpret the cost function as an integral and up until that point, one of the worst integration methods was chosen with the right point rule for Riemann sums, which converges linearly with the number evaluation points. A better approach is to use the midpoint rule, which converges quadratically with the number of evaluation points [24]. Since the midpoint rule, like the right point rule, is a rectangular integration method, it will be used for future benchmarks for comparing different integration methods. Especially since the integrand $c(\mathbf{g}, t)$ is an oscillatory function, as illustrated in Figure 3.1, a better approach is needed for maximum performance. Additionally, the integrand changes with each new parameters $\mathbf{g}$, and thus, the integral can be seen as a blackbox, which is why an adaptive or iterative integration method with a good error estimation scheme is required. The best method for the integrand $c(\mathbf{g}, t)$, as it turns out, is the Tanh-sinh quadrature, which is one quadrature formula of a whole family, the Double Exponential Formulas[46] (DE Formulas).

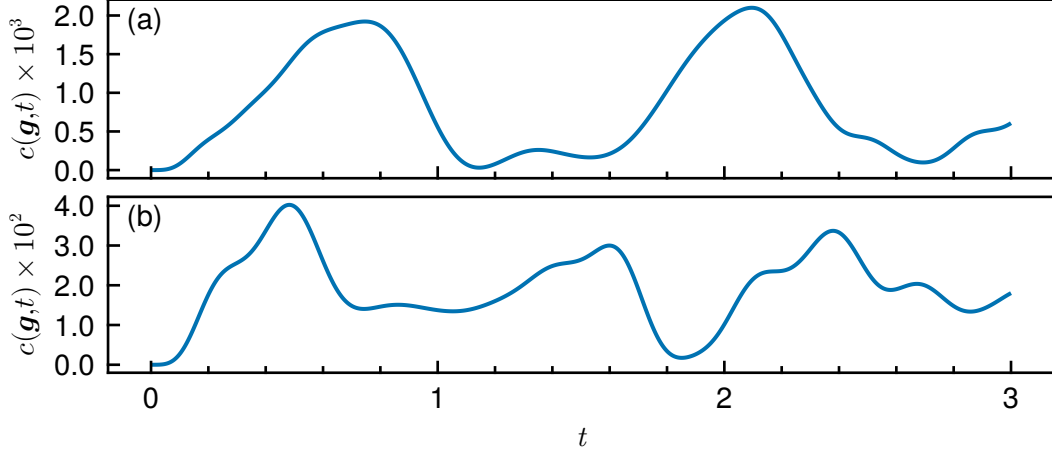


Figure 3.1: Integrand $c(\mathbf{g}, t)$ for the TFIM with $\Gamma = 1$, the Ansatz $\hat{H}_A^{\text{BW}}$, $N = 10$, $N_A = 5$ and OBC for (a) $\mathbf{g} = (1\ 2\ 3\ 4\ 5)^T$ and (b) $\mathbf{g} = (4\ 3\ 2\ 3\ 4)^T$.

Starting from an integral over the interval $[-1, 1]$

$$I = \int_{-1}^1 f(x) \, dx,$$

the DE Formulas utilize a variable transformation $x = \Phi(u)$ mapping the boundaries to infinity, i.e., $\phi(-\infty) = -1$ and $\phi(\infty) = 1$. That is, the integral reads

$$I = \int_{-\infty}^{\infty} f(\Phi(u)) \Phi'(u) \, du. \quad (3.1)$$

Applying the trapezoidal rule with a step size h to Equation (3.1) yields

$$I_h = h \sum_{j=-\infty}^{\infty} f(\Phi(jh)) \Phi'(jh)$$

with the abscissae $x_j = \Phi(jh)$ and weights $w_j = \Phi'(jh)$. To be able to compute the integrand with a computer, the sum needs to be truncated, which gives

$$I_h^M = h \sum_{j=-M}^M f(\Phi(jh)) \Phi'(jh).$$

The error ΔI_h^M stems from the error $E_D = I - I_h$ due to the discretization and the error $E_T = I_h - I_h^M$ caused by truncation of the infinite sum [36]. The best balance between the discretization error E_D and the truncation error E_T is achieved by a

variable transformation $x = \Phi(u)$, for which the integrand has a double exponential decay [36, 46]

$$f(\Phi(u))\Phi'(u) \approx e^{-\frac{\pi}{2}e^{|u|}}, \quad u \rightarrow \pm\infty,$$

giving the DE Formulas their name. The double exponential decay is achieved by the variable transformation

$$\Phi(u) = \tanh\left(\frac{\pi}{2} \sinh(u)\right), \quad (3.2)$$

caused by the derivative decaying as

$$\Phi'(u) \approx e^{-\frac{\pi}{2}e^{|u|}}, \quad u \rightarrow \pm\infty.$$

The specific variable transformation (3.2) gives the Tanh-sinh quadrature its name. The total error is roughly estimated as

$$|I - I_h^M| \approx e^{-C \frac{M}{\log(M)}}$$

with some $C > 0$ [36]. For the implementation of the Tanh-sinh quadrature, parts of the package `DoubleExponentialFormulas.jl`[37] are used and then fit to the purpose of this work. It estimates the integral iteratively at a level n (maximum of 12 levels used) with a step size $h_n = h_0/2^n$ beginning with $h_0 = 1$. After each iteration the error is estimated and checks whether the desired accuracy is achieved. If the desired accuracy has been reached, the integration is stopped and continued with the next level otherwise. This requires a good error estimation, which is not given in the package and was added manually, to get trustworthy results. For that, the heuristic error estimation scheme from Reference [1] is used and the following information is taken from it. The approximated integral at level k up to level n is denoted as S_k and the estimated error E_n at level n then is one if $n \leq 2$, zero if $S_n = S_{n-1}$ and 10^d otherwise, where $d = \max(d_1^2/d_2, 2d_1, d_3, d_4)$. The quantities d_i are given by

$$\begin{aligned} d_1 &= \log_{10} |S_n - S_{n-1}| \\ d_2 &= \log_{10} |S_n - S_{n-2}| \\ d_3 &= \log_{10} (\epsilon \cdot \max_j |w_j f(x_j)|) \\ d_4 &= \log_{10} \max(|w_l f(x_l)|, |w_r f(x_r)|) \end{aligned}$$

with $\epsilon = 10^{-p}$ and p as the precision in digits. For this work, $p = 15$ holds. Here, x_l and x_r are the closest abscissae to the left and right endpoint, respectively. The term d_1^2/d_2 is a multiplicative projection based on the differences between the result at the current level n and the past two levels. The fact that the optimal convergence rate achievable is quadratic, motivates $2d_1$. That means that the number of correct digits

can never be more than double the previous level. The quantity d_3 is motivated by the fact that the error cannot be less than the current precision ϵ times the largest product of abscissae and weights. The fourth term, d_4 , accounts for the truncation. For this work, this error estimation scheme is slightly modified. The logarithm in the definitions of the quantities d_i is dismissed, e.g. $d_1 = |S_n - S_{n-1}|$ is used instead $d_1 = \log_{10} |S_n - S_{n-1}|$. The estimated error then reads $E_n = d$. No problems have been observed with the modified error estimation scheme so far, which will be backed up in the benchmarks section (Section 4.4). In order to stop the integration, the relative error needs to be less than $\sqrt{\epsilon}$, where $\epsilon = 2.220\,446\,049\,250\,313 \cdot 10^{-16}$ as the machine epsilon for double precision. That is, $E_n \leq S_n \sqrt{\epsilon}$ must hold.

3.3 Gradient

Since the used optimizer is a gradient-based optimizer, the gradient of the cost function in respect to the variational parameters $\mathbf{g}$ is required. Differentiating the cost function as in Equation 2.12 in respect to a parameter g_k yields

$$\begin{aligned} \frac{\partial}{\partial g_k} C(\mathbf{g}) &= \frac{1}{T_{\max}} \int_0^{T_{\max}} \frac{1}{N_O} \sum_{j=1}^{N_O} \frac{\partial}{\partial g_k} \left(\underbrace{\langle \hat{\mathcal{O}}_j^A \rangle_t - \langle \hat{\mathcal{O}}_j^A \rangle_0}_{:=\delta_j(t)} \right)^2 dt \\ &= \frac{2}{T_{\max}} \int_0^{T_{\max}} \frac{1}{N_O} \sum_{j=1}^{N_O} \left(\frac{\partial}{\partial g_k} \langle \hat{\mathcal{O}}_j^A \rangle_t \right) \delta_j(t) dt. \end{aligned}$$

Thus, the derivative of the expectation value of the time evolved observables is needed, which, using the product rule, is given by

$$\begin{aligned} \frac{\partial}{\partial g_k} \langle \hat{\mathcal{O}}_j^A \rangle_t &= \frac{\partial}{\partial g_k} \text{Tr}_A \left[\hat{\mathcal{O}}_j^A \hat{U}_A(\mathbf{g}, t) \hat{\rho}_A \hat{U}_A^\dagger(\mathbf{g}, t) \right] \\ &= \text{Tr}_A \left[\hat{\mathcal{O}}_j^A \left(\frac{\partial}{\partial g_k} \hat{U}_A(\mathbf{g}, t) \right) \hat{\rho}_A \hat{U}_A^\dagger(\mathbf{g}, t) + \hat{\mathcal{O}}_j^A \hat{U}_A(\mathbf{g}, t) \hat{\rho}_A \frac{\partial}{\partial g_k} \hat{U}_A^\dagger(\mathbf{g}, t) \right]. \quad (3.3) \end{aligned}$$

That is, the derivative of the time evolution operator is required, which is not trivial, since the variational Ansatz in the exponent and its derivative with respect to the variational parameters do not commute. As a result, the chain rule can not be applied. As an alternative, finite differences could be used. However, its computation cost scales as $2N_P$. Above all, this performs extremely poorly for a large number of parameters and is not very accurate. The solution to this problem is the Fréchet derivative.

3.3.1 Fréchet derivative

The Fréchet derivative is an expression stemming from matrix calculus, which can be seen as a generalization of scalar and vector calculus (scalar derivatives, gradients, etc...) to matrices. It is of great importance in e.g. machine learning, where cost functions of matrix functions often appear and thus, to make use of gradient-based optimizers, the derivatives of matrix functions are needed. In this work, the underlying matrix function is the time evolution operator $\hat{U}_A$, a matrix exponential, with the variational Ansatz as an input. The Fréchet derivative is formally introduced as the derivative of a matrix function $f : \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$ at a point $X \in \mathbb{C}^{n \times n}$ as a linear mapping

$$\begin{aligned} \mathbb{C}^{n \times n} &\xrightarrow{\mathcal{L}_f} \mathbb{C}^{n \times n} \\ E &\mapsto \mathcal{L}_f(X, E) \end{aligned}$$

s.t. for all $E \in \mathbb{C}^{n \times n}$ [18]

$$f(X + E) - f(X) - \mathcal{L}_f(X, E) = o(\|E\|).$$

That reads, the Fréchet derivative $\mathcal{L}(X, E)$ of f at point X in the direction of or acting on E . The expression $h = o(\|E\|)$ means that $\|h\|/\|E\| \rightarrow 0$ as $\|E\| \rightarrow 0$ [18]. The usual sum or product rule is still valid. For a good overview of rules for the Fréchet derivative, the reader is referred to Reference [34]. Another, less formal, way of thinking is to interpret the Fréchet derivative as a linear operator stemming from the derivative as a linearization

$$df(X) = f(X + dX) - f(X) = \mathcal{L}_f(X, dX) + \mathcal{O}(dX^2),$$

dropping higher order terms beyond dX [22]. The Fréchet derivative of $f(X) = X^2$ is then obtained via the product rule

$$d f(X) = dX^2 = dX X + X dX = \mathcal{L}_{X^2}(X, dX).$$

Generalizing it to the n -th power yields

$$dX^n = \sum_{l=0}^{n-1} X^l dX X^{n-l-1} = \mathcal{L}_{X^n}(X, dX).$$

Applying it to the matrix exponential by expanding it into a power series gives

$$de^X = \sum_{n=0}^{\infty} \frac{1}{n!} dX^n = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{l=0}^{n-1} X^l dX X^{n-l-1} = \mathcal{L}_{e^X}(X, dX).$$

Finally, the partial derivative of the time evolution operator $\hat{U}_A(\mathbf{g}, t) = \exp(-i\hat{H}_A^{\text{Var}}(\mathbf{g})t)$ in respect to the parameters $\mathbf{g}$ reads

$$\frac{\partial}{\partial g_k} \hat{U}_A = \sum_{n=0}^{\infty} \frac{1}{n!} (-it)^n \frac{\partial}{\partial g_k} (\hat{H}_A^{\text{Var}})^n \quad (3.4)$$

$$\begin{aligned} &= \sum_{n=0}^{\infty} \frac{1}{n!} (-it)^n \sum_{l=0}^{n-1} (\hat{H}_A^{\text{Var}})^l \left(\frac{\partial}{\partial g_k} \hat{H}_A^{\text{Var}} \right) (\hat{H}_A^{\text{Var}})^{n-l-1} \\ &= -it \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{l=0}^{n-1} (-it \hat{H}_A^{\text{Var}})^l \left(\frac{\partial}{\partial g_k} \hat{H}_A^{\text{Var}} \right) (-it \hat{H}_A^{\text{Var}})^{n-l-1} \end{aligned} \quad (3.5)$$

$$= -it \mathcal{L}_{\text{ex}} \left(-it \hat{H}_A^{\text{Var}}, \frac{\partial}{\partial g_k} \hat{H}_A^{\text{Var}} \right). \quad (3.6)$$

That is, the Fréchet derivative of the matrix exponential at $-it\hat{H}_A^{\text{Var}}$ in the direction of $\partial/\partial g_k \hat{H}_A^{\text{Var}}$ needs to be computed. The partial derivative of the variational Ansatz is rather trivial

$$\frac{\partial}{\partial g_k} \hat{H}_A^{\text{Var}} = \frac{\partial}{\partial g_k} \sum_i g_i \hat{h}_i = \hat{h}_k.$$

Now, of course, the challenge is to compute the Fréchet derivative. However, there are exact formulas involving exponentiation of an augmented block triangular matrix (see Reference [35]), where the augmented vector space is double in size. An even more enlarged vector space is needed for the method presented in Reference [47]. The most efficient algorithm, for this purpose, is the algorithm presented in Reference [35], utilizing the Padé approximation with the scaling and squaring method s.t. the computation is performed directly on the original vector space, without enlargement of matrices. For that, the implementation of the package `ChainRules.jl`[6] is used.

3.3.2 Gradient in a Forward mode fashion

After having written down the partial derivative of the time evolution operator in respect to the parameters in Equation (3.6) in terms of the Fréchet derivative, a final expression for the gradient of the expectation value (Equation (3.3)) can be found. One last preparation still needs to be done, namely finding the derivative of the adjoint of the time evolution operator, which is read off as

$$\frac{\partial}{\partial g_k} \hat{U}_A^\dagger = it \mathcal{L}_{\text{ex}}^\dagger \left(-it \hat{H}_A^{\text{Var}}, \hat{h}_k \right). \quad (3.7)$$

Inserting both derivatives into Equation (3.3) yields

$$\begin{aligned}
\frac{\partial}{\partial g_k} \langle \hat{\mathcal{O}}_j^A \rangle_t &= -it \operatorname{Tr}_A \left[\hat{\mathcal{O}}_j^A \mathcal{L}_{e^x} \left(-it \hat{H}_A^{\text{Var}}, \hat{h}_k \right) \hat{\rho}_A \hat{U}_A^\dagger \right] \\
&\quad + it \operatorname{Tr}_A \left[\hat{\mathcal{O}}_j^A \hat{U}_A \hat{\rho}_A \mathcal{L}_{e^x}^\dagger \left(-it \hat{H}_A^{\text{Var}}, \hat{h}_k \right) \right] \\
&= -it \operatorname{Tr}_A \left[\mathcal{L}_{e^x} \left(-it \hat{H}_A^{\text{Var}}, \hat{h}_k \right) \hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A \right] \\
&\quad + it \operatorname{Tr}_A^* \left[\mathcal{L}_{e^x} \left(-it \hat{H}_A^{\text{Var}}, \hat{h}_k \right) \hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A \right] \\
&= 2t \operatorname{Im} \left\{ \operatorname{Tr}_A \left[\mathcal{L}_{e^x} \left(-it \hat{H}_A^{\text{Var}}, \hat{h}_k \right) \hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A \right] \right\}, \tag{3.8}
\end{aligned}$$

where from line one and two to line three and four, the hermicity of $\hat{\rho}_A$ and $\hat{\mathcal{O}}_j^A$, the cyclic invariance of the trace and $\operatorname{Tr}_A[X] = \operatorname{Tr}_A^*[X^\dagger]$ for some square matrix X were used. The expressions Tr_A^* and $\operatorname{Im}\{x\}$ denote the complex conjugate of the trace and the imaginary part of some complex number x , respectively. Finally, an expression for the derivative of the expectation value is found. However, the formula obtained in Equation (3.8) corresponds to a “Forward mode fashion”, since for each parameter, the Fréchet derivative needs to be computed, and thus, the runtime will scale strongly linearly with the number of parameters. Again, especially for a large number of parameters, this is not efficient enough. Since the gradient of a scalar cost function is sought and the input (the parameters $\mathbf{g}$) is a vector, the most efficient approach is a “Reverse mode fashion”.

3.3.3 Gradient in a Reverse mode fashion

Expanding the Fréchet derivative in Equation (3.8) yields

$$\frac{\partial}{\partial g_k} \langle \hat{\mathcal{O}}_j^A \rangle_t = 2t \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{l=0}^{n-1} \operatorname{Im} \left\{ \operatorname{Tr}_A \left[(-it \hat{H}_A^{\text{Var}})^l \hat{h}_k (-it \hat{H}_A^{\text{Var}})^{n-l-1} \hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A \right] \right\}.$$

Now, the trick is to use the cyclic invariance of the trace

$$\begin{aligned}
\frac{\partial}{\partial g_k} \langle \hat{\mathcal{O}}_j^A \rangle_t &= 2t \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{l=0}^{n-1} \operatorname{Im} \left\{ \operatorname{Tr}_A \left[(-it \hat{H}_A^{\text{Var}})^{n-l-1} \hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A (-it \hat{H}_A^{\text{Var}})^l \hat{h}_k \right] \right\} \\
&= 2t \operatorname{Im} \left\{ \operatorname{Tr}_A \left[\mathcal{L}_{e^x} \left(-it \hat{H}_A^{\text{Var}}, \hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A \right) \hat{h}_k \right] \right\}.
\end{aligned}$$

That is, a Fréchet derivative needs to be computed only one time, namely at the point $-it \hat{H}_A^{\text{Var}}$ in the direction of $\hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A$, independently of the number of parameters. Still, for each parameter, one additional matrix multiplication and one additional trace needs to be computed, s.t. the computation time still scales linearly with the number of parameters. However, in comparison to the Forward mode approach,

the runtime scales very weakly linearly. Summing up every intermediate result, the complete k -th entry of the gradient of the cost function reads

$$\begin{aligned}
& \frac{\partial}{\partial g_k} C(\mathbf{g}) \\
&= \frac{4}{T_{\max}} \int_0^{T_{\max}} \frac{t}{N_O} \sum_{j=1}^{N_O} \text{Im} \left\{ \text{Tr}_A \left[\mathcal{L}_{\text{ex}} \left(-it \hat{H}_A^{\text{Var}}, \hat{\rho}_A \hat{U}_A^\dagger \hat{\mathcal{O}}_j^A \right) \hat{h}_k \right] \right\} \delta_j(t) dt \\
&= \frac{4}{T_{\max} N_O} \int_0^{T_{\max}} t \text{Im} \left\{ \text{Tr}_A \left[\mathcal{L}_{\text{ex}} \left(-it \hat{H}_A^{\text{Var}}, \hat{\rho}_A \hat{U}_A^\dagger \hat{\Xi}_A \right) \hat{h}_k \right] \right\} dt,
\end{aligned}$$

where $\hat{\Xi}_A = \sum_{j=1}^{N_O} \hat{\mathcal{O}}_j^A \delta_j(t)$, which can be used, because the Fréchet derivative is linear in $\hat{\mathcal{O}}_j^A$, resulting in only one computation of the Fréchet derivative for all observables instead of one computation for each observable. Finally, the most efficient approach is found. Its runtime is almost independent of the number of observables and parameters (it scales very weakly linearly with the number of observables and parameters, to be exact). Additionally, it does not need any augmented vector space s.t. everything is limited to the original Hilbert space and all intermediate calculations, e.g. $\delta_j(t)$ for all j , from the cost function can be reused.

4 Results: Convergence

This chapter provides detailed information about the convergence of the results. It will be investigated whether it is enough to sample the integrand at a few arbitrary time points and how strong the influence of the maximum observation time $T_{\max}$ is. The convergence properties will be demonstrated with the TFIM and both Ansätze, $\hat{H}_A^{\text{BW}}$ and $\hat{H}_A^{\text{BWV}}$. In the following, only the ratios of parameters will be compared, since the algorithm does not determine the scale of the optimal parameters (see Section 2.5). Additionally, benchmarks are included, suggesting that the numerical methods used are the best for this algorithm.

4.1 Cost function and convergence of the midpoint rule for the BW-like Ansatz

This section provides an understanding of the cost function and proves that it is not sufficient to monitor the observables at a few arbitrary time points if the variational Ansatz cannot reconstruct the exact EH accurately, by using the midpoint rule with a gradually decreasing time step size Δt . Throughout this section, the TFIM is used with $N = 8$, $N_A = 4$, $I = 1$, OBC and the variational Ansatz $\hat{H}_A^{\text{BW}}$. Figure 4.1 shows the behaviour of the cost function in the parameter space around one minimum in the different directions by plotting $\mathcal{C}_i := \mathcal{C}(\mathbf{g}^{\text{opt}} + \varepsilon \mathbf{e}_i)$ vs. ε , where $\mathbf{e}_i$ is the unit vector in the i -th direction in the parameter space. The cost function was minimized using the Tanh-sinh quadrature. A maximum time of $T_{\max} = 20$ was used for both, the minimization and the plots of the cost function landscape. The cost function shows significant oscillations for a higher time step size of $\Delta t = 2$, which vanish as the time step size decreases. At $\Delta t = 0.01$ the oscillations seem to be completely disappeared and the landscape looks like the landscape obtained with the Tanh-sinh quadrature. This could lead to the assumption that the cost function is converged for $\Delta t = 0.01$. Additionally, the convexity of the cost function in the neighbourhood of the obtained minimum is mentionable. Of course, the depicted landscapes in Figure 4.1 are only slices in a high dimensional parameter space and statements about the form of the landscape must be made with caution. However, the landscape was plotted for multiple random points and even 3d-plots were made. Neither strong oscillations nor discontinuities have been observed. These additional

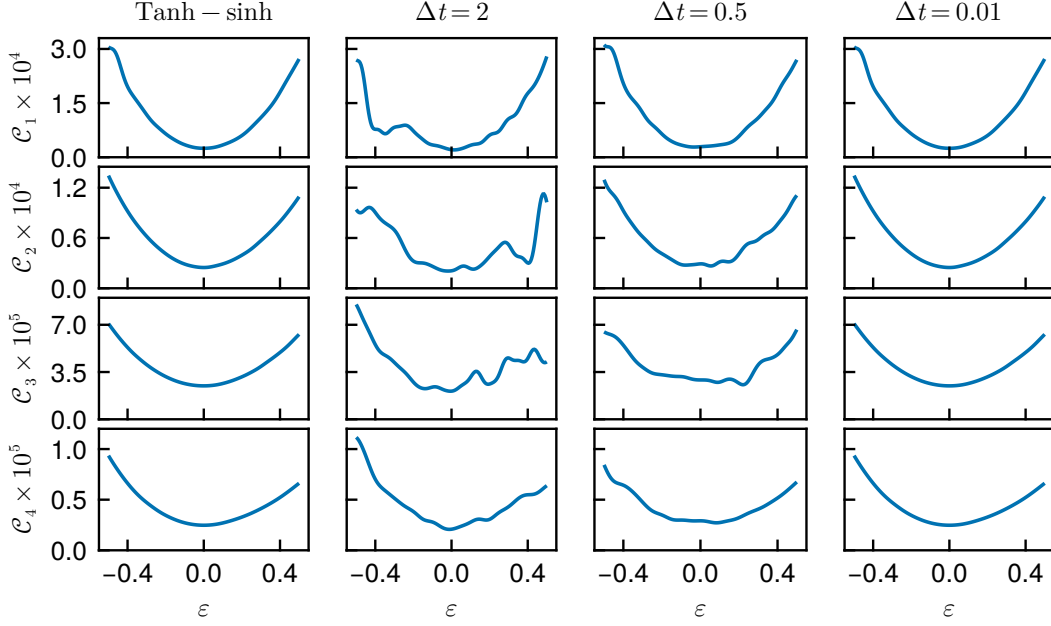


Figure 4.1: Slices of the cost function landscape $\mathcal{C}_i = \mathcal{C}(\mathbf{g}^{\text{opt}} + \varepsilon \mathbf{e}_i)$ in the i -th direction of the parameter space around a minimum at $\mathbf{g}^{\text{opt}}$ for the Tanh-sinh quadrature and the midpoint rule with different Δt . The variable ε controls how far the landscape is sampled. The cost function was minimized with the Tanh-sinh quadrature with $T_{\text{max}} = 20$, which is also used for the plots.

plots are omitted for brevity. It is important to note that the minimum of the cost function is significantly greater than zero, namely in the vicinity of $2 \cdot 10^{-5}$. Thus the variational Ansatz is not good enough.

The target values are the optimal variational parameters $\mathbf{g}^{\text{opt}}$, which is why the convergence can and should be checked in detail via the optimal parameters. Figure 4.2 shows the optimal parameters $\mathbf{g}^{\text{opt}}$ normalized to the first parameter g_1^{opt} vs. Δt , which were obtained by minimizing the cost function for each Δt with $T_{\text{max}} = 1$. The first parameter is not displayed, since $g_1^{\text{opt}}/g_1^{\text{opt}} = 1$. For $\Delta t \geq 0.2$, the parameters show sudden jumps (Figure 4.2(a)) and seem to have converged for $\Delta t \approx 0.1$ and smaller. However, zooming in (see Figure 4.2(b),(c) and (d)) shows that the parameters still steadily increase even for very small Δt and all parameters exhibit the same behaviour. Table 4.1 contains the normalized optimal parameters $\mathbf{g}^{\text{opt}}/g_1^{\text{opt}}$ rounded to 15 decimal places obtained via the Tanh-sinh quadrature and the midpoint rule with different Δt , again with $T_{\text{max}} = 1$. As already shown in Figure 4.2 it can be seen that the optimal parameters obtained with the midpoint rule approach the optimal parameters calculated with the Tanh-sinh quadrature asymptotically from below as the time step size Δt decreases, except for $\Delta t = 10^{-6}$, where the ratios

4.1 Cost function and convergence of the midpoint rule for the BW-like Ansatz

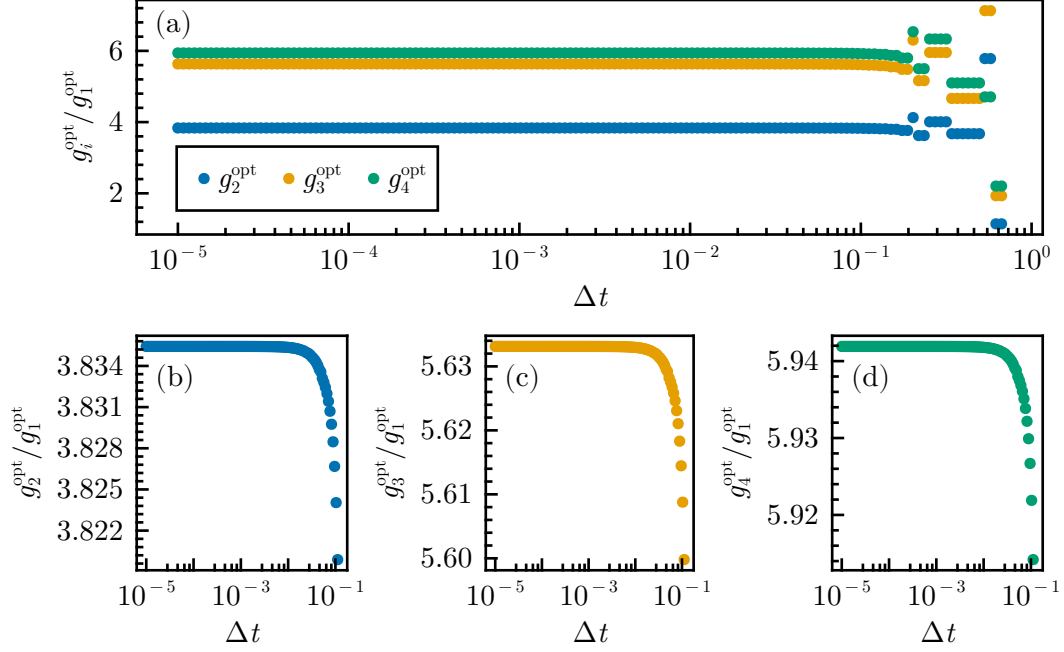


Figure 4.2: Normalized optimal parameters $g^{\text{opt}}/g_1^{\text{opt}}$ vs. time step size Δt for the midpoint rule over the interval of (a) $\Delta t \in [2/3, 10^{-5}]$. (b), (c) and (d) show the normalized optimal parameters $g_2^{\text{opt}}/g_1^{\text{opt}}$, $g_3^{\text{opt}}/g_1^{\text{opt}}$ and $g_4^{\text{opt}}/g_1^{\text{opt}}$, respectively, for $\Delta t \in [10^{-5}, \approx 0.11]$. The minimizer was run for each Δt with the midpoint rule and $T_{\text{max}} = 1$ to obtain g^{opt} .

Table 4.1: Normalized optimal parameters $g^{\text{opt}}/g_1^{\text{opt}}$ obtained with the midpoint rule with different Δt and the Tanh-sinh quadrature, both with $T_{\text{max}} = 1$, rounded to 15 decimal places. The convergence criterium was set to $\nabla_{\text{tol}} = 9 \cdot 10^{-17}$.

method	$g_2^{\text{opt}}/g_1^{\text{opt}}$	$g_3^{\text{opt}}/g_1^{\text{opt}}$	$g_4^{\text{opt}}/g_1^{\text{opt}}$
$\Delta t = 10^{-1}$	3.824 039 596 426 903	5.608 777 483 165 399	5.921 893 959 511 652
$\Delta t = 10^{-2}$	3.835 351 726 582 212	5.632 970 094 804 400	5.941 793 208 331 469
$\Delta t = 10^{-3}$	3.835 421 398 526 617	5.633 117 585 461 839	5.941 909 050 130 465
$\Delta t = 10^{-4}$	3.835 422 093 758 030	5.633 119 057 108 413	5.941 910 205 524 904
$\Delta t = 10^{-5}$	3.835 422 100 710 176	5.633 119 071 824 524	5.941 910 217 078 521
$\Delta t = 10^{-6}$	3.835 422 100 781 501	5.633 119 071 974 829	5.941 910 217 197 190
Tanh-sinh	3.835 422 100 780 429	5.633 119 071 973 224	5.941 910 217 195 280

are actually larger than the ratios obtained with the Tanh-sinh quadrature. Table 4.1 shows remarkable results, because for $\Delta t = 10^{-2}$ the integrand is evaluated 100 times but only two decimal places match the results of the Tanh-sinh quadrature, while for $\Delta t = 10^{-6}$ eleven decimal places are in agreement with the Tanh-sinh quadrature but the integrand is evaluated at 100000 time points. The Tanh-sinh quadrature shows more efficiency, since it evaluates the integrand only 101 times at the minimum of the cost function but seems to give very accurate results already.

These remarkable results demonstrate that, to get accurate results for an Ansatz, which does not reconstruct the exact EH accurately, it is not sufficient to monitor the observables at a few arbitrary time points and underline the statements in Section 2.6 that the cost function should not be treated as a discrete sum over a few time points but rather as an integral over the time domain. The fact that the ratios obtained with $\Delta t = 10^{-6}$ are larger than the ratios obtained with the Tanh-sinh quadrature implies that the midpoint rule overshoots the correct results with $\Delta t = 10^{-6}$ or the Tanh-sinh quadrature only delivers an accuracy up to eleven decimal points in this case. To verify this, the evaluation points for the Tanh-sinh quadrature were increased multiple times, but the first thirteen decimal places did not change. A benchmark for the accuracy of the Tanh-sinh quadrature is given in Section 4.4. Thus, the midpoint rule probably overestimates the optimal ratios $g^{\text{opt}}/g_1^{\text{opt}}$ with $\Delta t = 10^{-6}$.

4.2 Influence of the maximum observation time for the BW-like Ansatz

This section will show that no convergence of the parameters in regard to T_{max} can be achieved if the variational Ansatz is not accurate enough. For this purpose, the TFIM with $N = 8$, $N_A = 4$, $I = 1$, OBC and the variational Ansatz $\hat{H}_A^{\text{BW}}$ is used. Figure 4.3(a) shows the optimal parameters if T_{max} is varied. Here, for each T_{max} a minimization run was done. Each consecutive run was initialized with the optimal parameters from the previous run with the previous (higher) T_{max} . The optimal parameters clearly exhibit a $1/T_{\text{max}}$ dependence. However, the normalized parameters, $g_i^{\text{opt}}/g_1^{\text{opt}}$, are constant for all T_{max} up to $T_{\text{max}} \approx 45.47$ (Figure 4.4(b)). Up until the aforementioned T_{max} , the fluctuations in the normalized optimal parameters $g_i^{\text{opt}}/g_1^{\text{opt}}$ are in the vicinity of 10^{-12} . After $T_{\text{max}} \approx 45.47$ (shaded region in Figure 4.5(b)), the normalized optimal parameters show significantly higher fluctuations, which accounts to the loss of accuracy in the integration. This tendency can be observed in the cost function $\mathcal{C}(g^{\text{opt}})$ at its minimum as well (Figure 4.4). Like for the ratios $g_i^{\text{opt}}/g_1^{\text{opt}}$, the fluctuations of $\mathcal{C}(g^{\text{opt}})$ are in the vicinity of 10^{-12} before $T_{\text{max}} \approx 45.47$, after

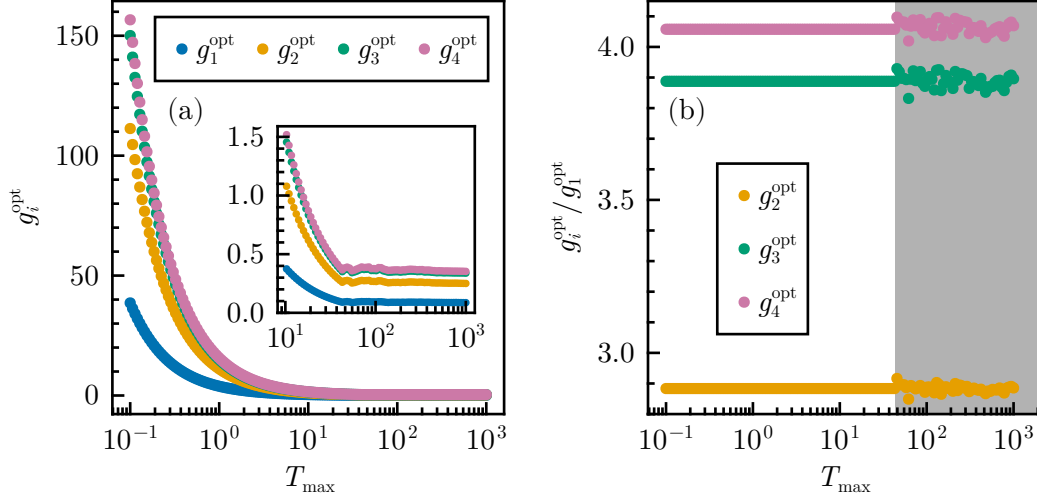


Figure 4.3: (a) Optimal parameters g^{opt} for each T_{max} . (b) Optimal parameters normalized to g_1^{opt} . The grey region indicates that the integral is not evaluated accurately enough for the corresponding maximum integration times T_{max} .

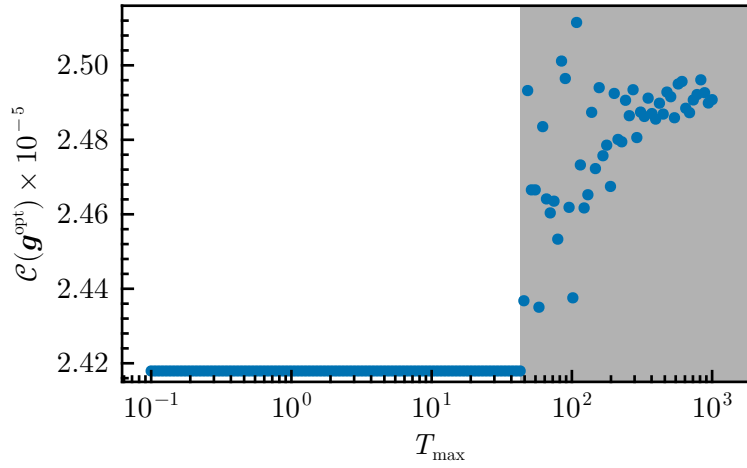


Figure 4.4: Cost function $\mathcal{C}(g^{\text{opt}})$ at its minimum vs. T_{max} . The grey region indicates that the integral is not evaluated accurately enough for the corresponding maximum integration times T_{max} .

which (shaded region in Figure 4.4), the cost function shows larger fluctuations, too. Figure 4.5(a) and (b) give an impression how the integrand looks for $T_{\max} \approx 45.47$

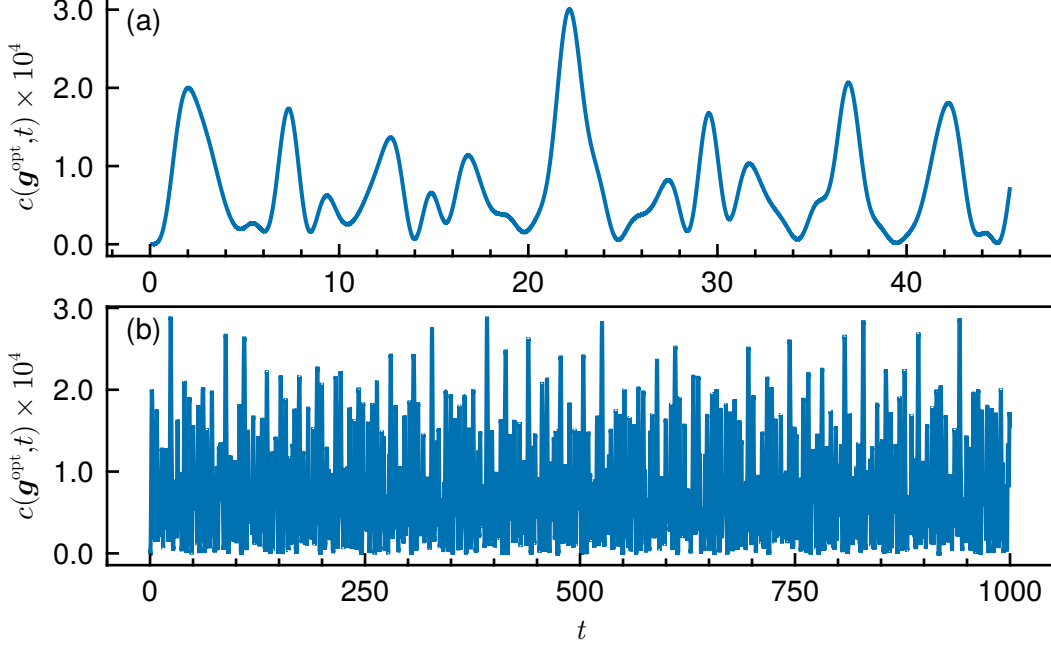


Figure 4.5: Integrand $c(\mathbf{g}^{\text{opt}}, t)$ for (a) $T_{\max} \approx 45.47$ and (b) $T_{\max} = 10^3$ at the obtained minimum $\mathbf{g}^{\text{opt}}$ of the cost function.

and $T_{\max} = 10^3$, respectively. For $T_{\max} \approx 45.47$ the integrand already exhibits a large amount of oscillations. However, for $T_{\max} = 10^3$, integrating the integrand seems almost hopeless if it needs to be done in an appropriate computation time. Now the open question is, how the algorithm and the integrand behave if the minimization run is initialized with different initial parameters. First of all, it is important to understand the $\propto 1/T_{\max}$ dependence of the optimal parameters (see Figure 4.3) in the case of initialization with the optimal parameters of the previous run. Table 4.2 shows these optimal parameters for a few $T_{\max}$, which will be referred to as the higher optimal parameters, $\mathbf{g}^{\text{opt,h}}$, in this section. There, the same tendency shows up, namely that the higher $T_{\max}$ is, the smaller $\mathbf{g}^{\text{opt,h}}$ is. Now, to understand the $\propto 1/T_{\max}$ dependence and how this translates into the integrand, Figure 4.6 displays the integrand for these higher optimal parameters $\mathbf{g}^{\text{opt,h}}$ with its corresponding $T_{\max}$. It becomes apparent that the form of integrand looks exactly the same, independent of $T_{\max}$, and it is only stretched if $T_{\max}$ is higher. If the higher optimal parameters $\mathbf{g}^{\text{opt,h}}$ in Table 4.2 are uniformly scaled by a factor and used as initial parameters for new minimization runs for each $T_{\max}$, the optimal parameters are expected to be lower and the integrand is expected to look different

Table 4.2: Higher optimal parameters $\mathbf{g}^{\text{opt,h}}$ for each T_{max} rounded to three decimal places, displayed in Figure 4.3, where each run was initialized with the optimal parameters from the previous run.

T_{max}	$g_1^{\text{opt,h}}$	$g_2^{\text{opt,h}}$	$g_3^{\text{opt,h}}$	$g_4^{\text{opt,h}}$
0.1	38.603	111.311	150.077	156.644
1	3.860	11.131	15.008	15.664
10	0.386	1.113	1.501	1.566
40	0.097	0.278	0.375	0.392

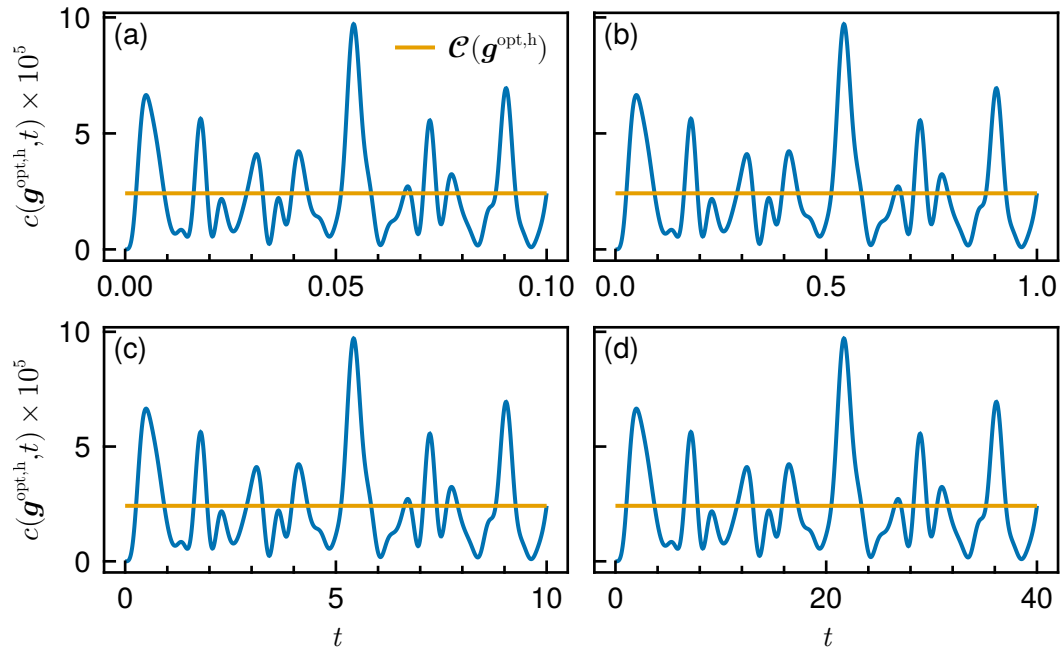


Figure 4.6: Integrand $c(\mathbf{g}^{\text{opt,h}}, t)$ for higher initial parameters for (a) $T_{\text{max}} = 0.1$, (b) $T_{\text{max}} = 1$, (c) $T_{\text{max}} = 10$ and (d) $T_{\text{max}} = 40$. The orange line is the cost function value $\mathcal{C}(\mathbf{g}^{\text{opt,h}})$ at the obtained minimum $\mathbf{g}^{\text{opt,h}}$, i.e., the mean value of the integrand over the interval $t \in [0, T_{\text{max}}]$.

4 Results: Convergence

as well. Table 4.3 shows the optimal parameters $\mathbf{g}^{\text{opt},l}$, which will be referred to as the lower optimal parameters. These are obtained by initializing new minimization runs with the higher optimal parameters $\mathbf{g}^{\text{opt},h}$, shown in Table 4.3, uniformly scaled by a factor 10^{-1} , i.e. $\mathbf{g}^{\text{init},l} = 10^{-1} \mathbf{g}^{\text{opt},h}$, where $\mathbf{g}^{\text{init},l}$ are the initial parameters. Still,

Table 4.3: Optimal parameters $\mathbf{g}^{\text{opt},l}$ for each T_{max} rounded to three decimal places, where each run was initialized with the higher optimal parameters $\mathbf{g}^{\text{init},h}$ from Table 4.2 scaled uniformly by a factor 10^{-1} , i.e. $\mathbf{g}^{\text{init},l} = 10^{-1} \mathbf{g}^{\text{opt},h}$.

T_{max}	$g_1^{\text{opt},l}$	$g_2^{\text{opt},l}$	$g_3^{\text{opt},l}$	$g_4^{\text{opt},l}$
0.1	4.188	16.063	23.592	24.885
1	0.419	1.606	2.359	2.488
10	0.042	0.161	0.236	0.249
40	0.010	0.040	0.059	0.062

the form of the integrand should look exactly the same for all T_{max} but differs from the form of integrand for $\mathbf{g}^{\text{opt},h}$. And this is exactly what happens, as depicted in Figure 4.7.

If the variational Ansatz is not good enough, dynamics still happen even at the minimum, and thus, as seen previously in this section, T_{max} has an influence on the found solution. If a run with a slightly lower maximum observation time, call it $\bar{T}_{\text{max}}$, is initialized with the optimal parameters $\mathbf{g}^{\text{opt}}$ from the previous run with a $T_{\text{max}} = \gamma \bar{T}_{\text{max}}$, where γ is bigger than but almost one, the optimizer finds a solution $\bar{\mathbf{g}}^{\text{opt}} = \gamma \mathbf{g}^{\text{opt}}$, where the same dynamics happen, but just over a smaller time scale. In terms of the EH, the relation $\hat{H}_A^{\text{Var}}(\bar{\mathbf{g}}^{\text{opt}}) = \hat{H}_A^{\text{Var}}(\gamma \mathbf{g}^{\text{opt}}) = \gamma \sum_i g_i^{\text{opt}} \hat{h}_i = \gamma \hat{H}_A^{\text{Var}}(\mathbf{g})$ then holds. That is, the energy scale (strictly speaking, it is an “entanglement energy scale”, since the eigenvalues of the EH build the ES) is slightly higher by a factor γ but the form of the EH is the same, which is why the same dynamics happen, but just faster, which is monitored over a smaller time scale $\bar{T}_{\text{max}}$. The same behaviour can be observed if a consecutive run with $\bar{T}_{\text{max}} = T_{\text{max}}/\gamma$ is initialized with the parameters $\gamma \mathbf{g}^{\text{opt}}$, where $\mathbf{g}^{\text{opt}}$ is a solution from a run with T_{max} (see Figure 4.7 and 4.6, and Table 4.3 and 4.2). Here, γ does not necessarily need to be slightly smaller or bigger than one but can take any arbitrary positive value.

The optimal parameters, their ratios to be exact, appear to have converged according to Figure 4.3(b). However, there is still a bias in those results, namely that each consecutive run was initialized with the optimal parameters from the previous run. The ratios from the higher parameters $\mathbf{g}^{\text{opt},h}$ and lower parameters $\mathbf{g}^{\text{opt},l}$ differ. For example in the case of $T_{\text{max}} = 1$, $g_5^{\text{opt},h}/g_1^{\text{opt},h} \approx 4.058$ and $g_5^{\text{opt},l}/g_1^{\text{opt},l} \approx 5.938$ hold. That is, two runs with significantly different initial parameters but the same T_{max} yield different ratios of the optimal parameters, which would hint to no convergence

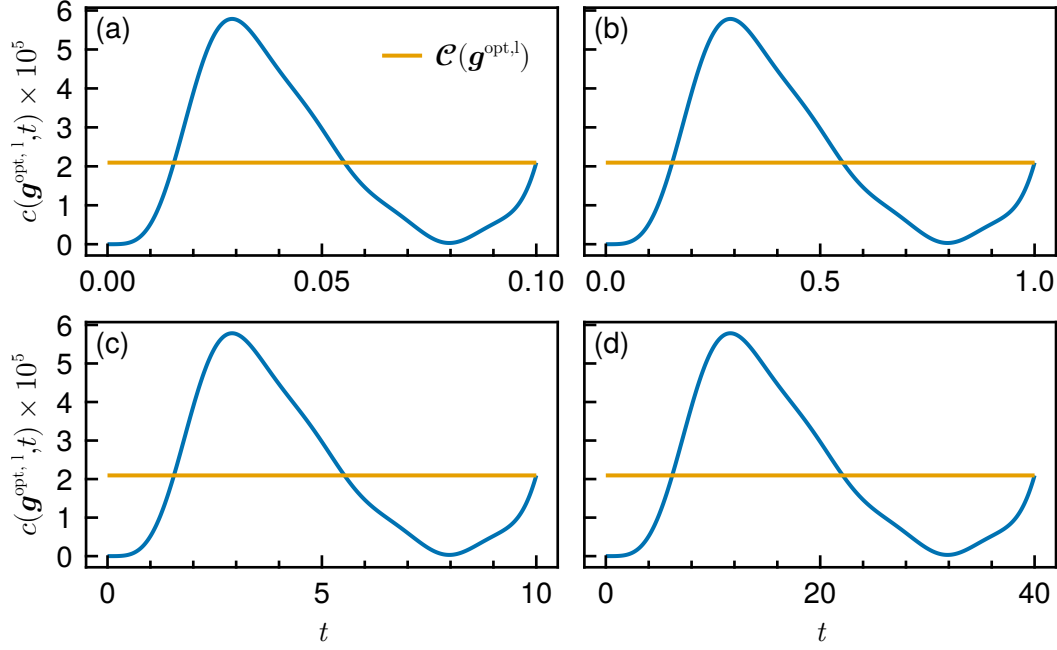


Figure 4.7: Integrand $c(\mathbf{g}^{\text{opt},l}, t)$ for lower initial parameters for (a) $T_{\max} = 0.1$, (b) $T_{\max} = 1$, (c) $T_{\max} = 10$ and (d) $T_{\max} = 40$. The orange line is the cost function value $\mathcal{C}(\mathbf{g}^{\text{opt},l})$ at the obtained minimum $\mathbf{g}^{\text{opt},l}$, i.e., the mean value of the integrand over the interval $t \in [0, T_{\max}]$.

of the optimal parameters. To check this, Figure 4.8 shows the ratios of the optimal parameter vs. $T_{\max}$. Here, each run was initialized with the same initial parameters, $\mathbf{g}^{\text{opt},h}$ for $T_{\max} = 1$ from table 4.2. It can be seen that the optimal parameters exhibit oscillations and are not converged for the displayed $T_{\max}$. The parameters are not expected to converge, since dynamics are always present if the variational Ansatz cannot capture the exact EH. Figure 4.9 shows the minimum of the cost function for each run. It clearly underlines the statement that the optimal parameters are not converged.

To conclude, $T_{\max}$ will always have an influence on the optimal parameters s.t. no convergence will be reached if the Ansatz is not an accurate representation of the true EH.

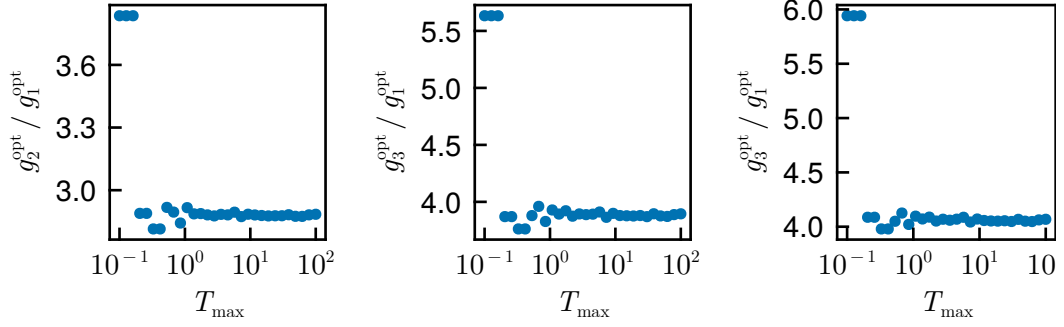


Figure 4.8: Optimal parameters normalized to g_1^{opt} vs. $T_{\max}$ if each run is initialized with the same initial parameters, the parameters $\mathbf{g}^{\text{opt,h}}$ for $T_{\max} = 1$ from table 4.2.

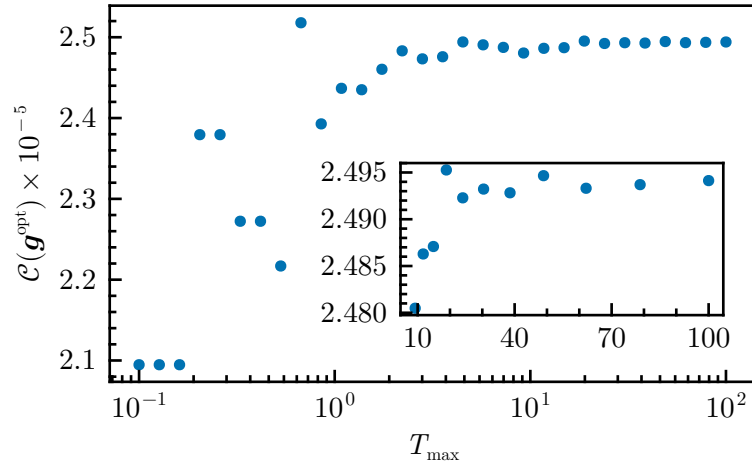


Figure 4.9: Cost function $\mathcal{C}(\mathbf{g}^{\text{opt}})$ at its minimum for the corresponding maximum integration times $T_{\max}$. Each run was initialized with the same initial parameters, the parameters $\mathbf{g}^{\text{opt,h}}$ for $T_{\max} = 1$ from table 4.2.

4.3 Convergence properties of the BW-violating Ansatz

The previous discussions on convergence were made with the BW-violating Ansatz $\hat{H}_A^{\text{BW}}$, where minimum of the cost function value is finite. This section deals with the convergence properties if a proper Ansatz is chosen, where the cost function is numerically zero at its minimum. For this section, the TFIM with $N = 8$, $N_A = 4$, OBC, $\Gamma = 1$ and the Ansatz $\hat{H}_A^{\text{BWV}}$ is chosen.

The first study concerns the number of observation times, where each run uses the midpoint rule with different time steps Δt . Here, $T_{\text{max}} = 1$ and the initial parameters $\mathbf{g}_{\text{init}} = (3.5810121415)^T$ as a good initial guess were used for all runs. It is important to note that the index i of a parameter g_i is not directly related to the i -th lattice site for the Ansatz $\hat{H}_A^{\text{BWV}}$. Table 4.4 shows the optimal parameters

Table 4.4: Optimal parameters normalized to Γ_1^{opt} for the Ansatz $\hat{H}_A^{\text{BWV}}$ and its corresponding minimum of the cost function for different time steps Δt for the midpoint rule. All runs were initialized with the same initial parameters.

method	$J_{1,2}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$I_2^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{2,3}^{\text{opt}}/\Gamma_1^{\text{opt}}$
$\Delta t = 0.25$	1.965 946 199 367 860	2.864 944 458 808 847	3.666 380 470 827 243
$\Delta t = 10^{-1}$	1.965 946 199 367 813	2.864 944 458 808 732	3.666 380 470 827 035
$\Delta t = 10^{-2}$	1.965 946 199 367 794	2.864 944 458 808 709	3.666 380 470 827 011
$\Delta t = 10^{-3}$	1.965 946 199 367 796	2.864 944 458 808 713	3.666 380 470 827 013
$\Delta t = 10^{-4}$	1.965 946 199 367 799	2.864 944 458 808 718	3.666 380 470 827 019
$\Delta t = 10^{-5}$	1.965 946 199 367 800	2.864 944 458 808 720	3.666 380 470 827 025
Tanh-sinh	1.965 946 199 367 803	2.864 944 458 808 727	3.666 380 470 827 037
$\Gamma_3^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{3,4}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\Gamma_4^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\mathcal{C}(\mathbf{g}^{\text{opt}})$
4.342 962 293 250 305	4.871 649 743 585 922	5.234 439 004 803 615	$1.438 \cdot 10^{-30}$
4.342 962 293 250 064	4.871 649 743 585 611	5.234 439 004 803 229	$1.135 \cdot 10^{-30}$
4.342 962 293 250 016	4.871 649 743 585 476	5.234 439 004 803 051	$5.790 \cdot 10^{-31}$
4.342 962 293 250 017	4.871 649 743 585 489	5.234 439 004 803 071	$4.818 \cdot 10^{-31}$
4.342 962 293 250 025	4.871 649 743 585 498	5.234 439 004 803 082	$4.564 \cdot 10^{-31}$
4.342 962 293 250 033	4.871 649 743 585 508	5.234 439 004 803 094	$4.516 \cdot 10^{-31}$
4.342 962 293 250 049	4.871 649 743 585 528	5.234 439 004 803 114	$4.460 \cdot 10^{-31}$

normalized to Γ_1^{opt} for all different time steps Δt . It can be seen that all normalized parameters agree up to twelve decimal places. The second study concerns the

4 Results: Convergence

influence of the maximum observation time $T_{\max}$. Here, the Tanh-sinh quadrature is used and each run uses a different $T_{\max}$. Each run is initialized with the same initial parameters as in the first study. Table 4.5 shows the optimal parameters

Table 4.5: Optimal parameters normalized to Γ_1^{opt} for the Ansatz $\hat{H}_A^{\text{BWV}}$ and its corresponding minimum of the cost function for different $T_{\max}$. The Tanh-sinh quadrature was used to evaluate the cost function. All runs were initialized with the same initial parameters.

$T_{\max}$	$J_{1,2}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\Gamma_2^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{2,3}^{\text{opt}}/\Gamma_1^{\text{opt}}$
0.1	1.965 946 199 360 310	2.864 944 458 793 144	3.666 380 470 804 380
1	1.965 946 199 367 804	2.864 944 458 808 729	3.666 380 470 827 036
10	1.965 946 199 367 801	2.864 944 458 808 722	3.666 380 470 827 032
100	1.965 946 199 367 835	2.864 944 458 808 774	3.666 380 470 827 074
$\Gamma_3^{\text{opt}}/\Gamma_1^{\text{opt}}$	$J_{3,4}^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\Gamma_4^{\text{opt}}/\Gamma_1^{\text{opt}}$	$\mathcal{C}(\mathbf{g}^{\text{opt}})$
4.342 962 293 221 468	4.871 649 743 552 242	5.234 439 004 766 612	$1.978 \cdot 10^{-30}$
4.342 962 293 250 046	4.871 649 743 585 524	5.234 439 004 803 110	$4.876 \cdot 10^{-31}$
4.342 962 293 250 044	4.871 649 743 585 517	5.234 439 004 803 101	$4.013 \cdot 10^{-29}$
4.342 962 293 250 076	4.871 649 743 585 551	5.234 439 004 803 123	$4.877 \cdot 10^{-27}$

normalized to Γ_1^{opt} for all different $T_{\max}$. All normalized parameters agree up to 13 decimal places, except for $T_{\max} = 0.1$, where only nine decimal places of the optimal ratios with $T_{\max} = 0.1$ agree with the optimal ratios of the runs with higher $T_{\max}$. Both results demonstrate a robust behaviour with respect to the number of observation times and the maximum observation time.

To conclude, the results seem to be converged for all Δt and almost all $T_{\max}$ if the correct Ansatz is chosen. Here, the term “correct Ansatz” refers to an Ansatz that leads to a cost function that is numerically zero at its minimum, which is the case for the BW-violating Ansatz for the TFIM as can be seen in table 4.4 and 4.5. The open question now is why the results seem to have converged if the proper Ansatz is used, although small $T_{\max}$ is used and the cost function is sampled only at a few time points. If the cost function drops to zero with the correct Ansatz, the subsystem is constant in time. That is, there are no more dynamics present in the subsystem at the optimal solution $\mathbf{g}^{\text{opt}}$, which always exists regardless of Δt and $T_{\max}$. If no dynamics are present, then it is completely irrelevant how often or how long the subsystem is sampled, since it does not change over time. That is why the algorithm yields converged results for an arbitrary Δt or $T_{\max}$. Thus, the cost function can be interpreted as a discrete sum up to an arbitrary maximum

observation time $T_{\max}$. On the other hand, if an Ansatz, which cannot represent the true EH, is chosen, the cost function will be finite at the minimum. Thus, even at the minimum, dynamics will be present s.t. it is not irrelevant how often or how long the system is sampled. No convergence in regard to $T_{\max}$ can be achieved (see Section 4.2) and the cost function needs to be interpreted as an integral (see Section 4.1). The remarkable result that the minimum of the cost function is in the vicinity of 10^{-30} for the BW-violating Ansatz for the TFIM together with the form of the parameters will be discussed in Chapter 5. Since it is not known whether an Ansatz is the correct one, the cost function should still be interpreted as an integral. Thus, all future calculations will be done with the cost function as an integral.

It needs to be mentioned that one parameter can be fixed throughout the optimization if the Ansatz is good and the cost function drops to zero. Fixing one parameter effectively fixes the “entanglement energy scale”, which has an influence on the time scale. But as previously seen, the optimizer always finds a converged optimal solution if the Ansatz is accurate, regardless of the maximum observation time. Since it is not known whether the Ansatz is good, no parameter will be fixed throughout this work.

4.4 Benchmarks

This section shows that the used methods for integration and optimization are the best methods for this kind of problem among all the methods, which are used for the comparison. The model used is the TFIM with $N = 8$, $N_A = 4$, OBC and $\Gamma = 1$ with the variational Ansatz $\hat{H}_A^{\text{BW}}$. The first benchmark concerns the integration, for which only the Gauß-Kronrod quadrature is used for comparison. The reason behind that is, as already pointed out in Section 3.2, an iterative or adaptive integration method with a good error estimation is needed. In addition to the Tanh-sinh quadrature, the only good candidate found for such type of integration problems is the aforementioned Gauß-Kronrod quadrature. To this end, the implementation from the package `QuadGK.jl`[21] is used. In this test, the cost function is computed 25 times at different random parameters $\mathbf{g}$ between zero and ten and the number of integrand evaluations is averaged over these 25 different cost function computations for each $T_{\max}$. For both, the Tanh-sinh quadrature and the Gauß-Kronrod quadrature, the maximum relative error is set to the square root of the machine epsilon for double precision. The results are depicted in Figure 4.10, where the error bands are given by the standard deviation over the 25 cost function evaluations with the different random parameters $\mathbf{g}$. It can be clearly seen that the Tanh-sinh quadrature outperforms the Gauß-Kronrod quadrature, especially for large $T_{\max}$. At $T_{\max} = 20$, the Tanh-sinh quadrature requires ≈ 1950 fewer evaluations on average.

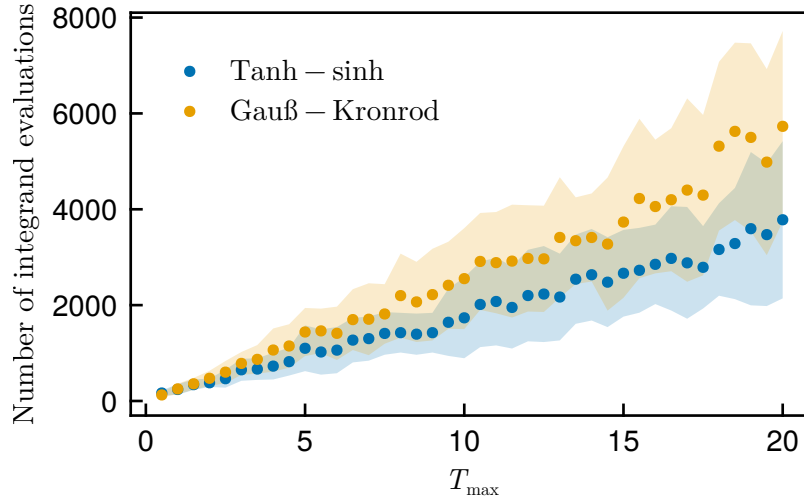


Figure 4.10: Comparison of the Tanh-sinh and Gauß-Kronrod quadrature. For each $T_{\max}$, 25 different random parameters $\mathbf{g}$ are sampled and the cost function is evaluated with both quadrature rules. The average number of integrand evaluation of the 25 samples is plotted together with the standard deviation (error bands).

Only at $T_{\max} = 0.5$, the Gauß-Kronrod needs fewer integrand evaluations. This probably amounts to the fact that the Gauß-Kronrod quadrature is implemented as an adaptive integration method, where the number of integrand evaluation can vary in arbitrary steps. Meanwhile, the Tanh-sinh quadrature halves the step size, s.t. the number of integrand evaluations cannot vary arbitrarily, and thus, uses too many evaluation points for such a small integration interval. Of course, both quadrature rules should yield the same cost function value for each new parameter set $\mathbf{g}$. Both rules showed no deviation greater than $\approx 4.441 \cdot 10^{-15}$ from one another over all 1000 samples, which is a remarkable result, since the Tanh-sinh quadrature clearly shows more efficiency while the accuracy does not suffer. An additional argument for the Tanh-sinh quadrature is that the abscissae and the weights are always the same for all integrands, and thus, the weights and abscissae can be calculated once at the beginning of a minimization run and can be reused in each iteration. On the other hand, the adaptivity of the Gauß-Kronrod quadrature requires the abscissae and weights to be calculated for each new parameters $\mathbf{g}$.

The second test concerns the optimization algorithm, where the algorithms used for comparison are Conjugate Gradient (ConjGrad), BFGS and Nelder-Mead (NM), which are all implemented in the package `Optim.jl` as well. The maximum integration time is set to $T_{\max} = 1$. The first three algorithms are gradient-based, whereas the Nealder-Mead algorithm is a direct search method and therefore does not require information about the gradient or Hessian. Initial parameters are chosen

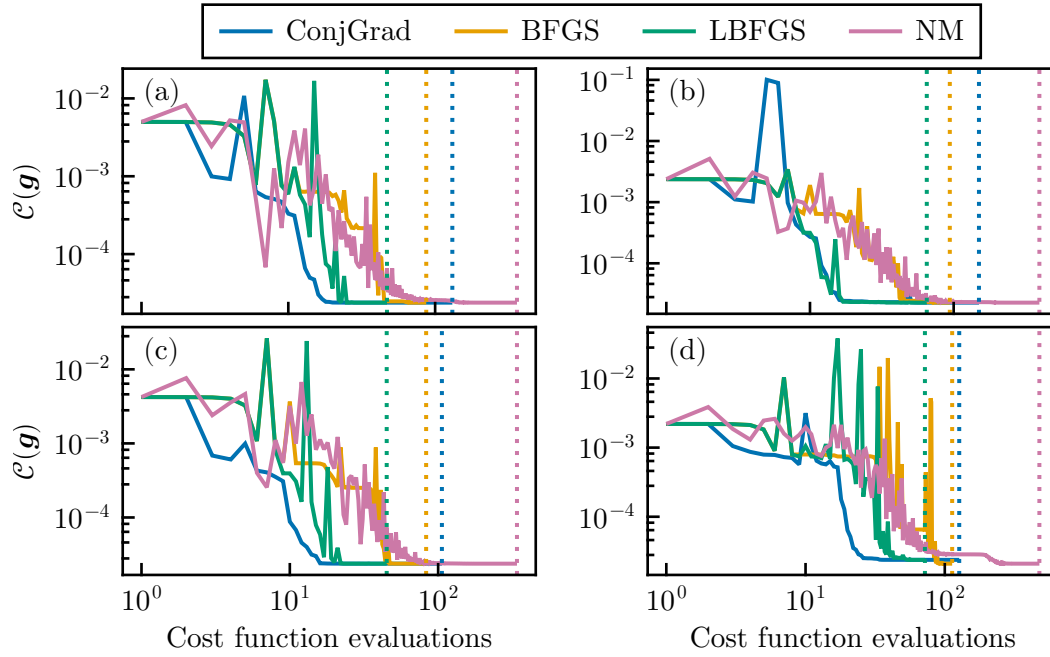


Figure 4.11: Cost function value vs. number of cost function evaluations in one minimization run. Random initial parameters are sampled and the cost function is minimized with the different algorithms. This procedure is done four times (run one, two, three and four in (a),(b),(c) and (d) respectively). The vertical dotted lines indicate the points, where the optimizers are converged.

at random and the cost function is minimized with all different algorithms. The convergence criterium for the gradient-based methods is set to $\nabla_{\text{tol}} = 10^{-16}$. NM uses a so called simplex, which consists of multiple points in the parameter space. The convergence criterium here is the standard deviation of the cost function value at these points in the current simplex, since the cost function value at the points of the simplex should be equal in the vicinity of the minimum. In this test, the standard deviation is required to be less than 10^{-16} , too. The cost function value vs. the number of cost function evaluations in one minimization run is shown in Figure 4.11. For all four different initial parameters, the LBFGS algorithm needed the fewest cost function evaluations, as indicated by the vertical dotted lines. However, in the fourth run (Figure 4.11(d)), the BFGS and Nelder-Mead algorithms found a slightly lower minimum ($\approx 3 \cdot 10^{-5}$ smaller) than the BFGS and Conjugate Gradient algorithms. Table 4.6 lists the time it took to minimize the cost function among all runs for all different algorithms. The minimization was repeated ten times for one set of initial parameters and the smallest time over these ten repetitions is listed. In all runs, the LBFGS algorithm won again in terms of runtime. To conclude,

Table 4.6: Runtime of the algorithms for the four sets of random initial parameters (i.e. four different runs).

algorithm	runtime / s			
	run 1	run 2	run 3	run 4
ConjGrad	2.400	3.088	1.898	2.271
BFGS	1.727	1.914	1.625	2.494
LBFGS	0.940	1.372	0.909	1.377
NM	1.873	2.161	1.779	2.505

even tho the LBFGS algorithm did not find the lowest minimum in the fourth run, the efficiency is very convincing. Additionally, this happened only one out of four times with random initial parameters. With a good initial guess this should not happen. The Gradient descent, ADAM and Simulated annealing algorithm have also been tested but not listed, because their performance were much worse than the algorithms included in the benchmark. It needs to be mentioned that a model with only 4 parameters has been used. To get deeper insights into the performance of the optimization algorithms, a model with significantly more parameters could prove helpful.

5 Results: Physics

The previous chapter dealt with the convergence properties of the algorithm, leading results that are as converged as possible. In this chapter, results in regard to the accuracy of the variational Ansatz are demonstrated and the explicit form of the EH as well as the CFT extensions will be discussed. The main result is the violation of the BW theorem on lattice systems in the XXZ model and the TFIM. Additionally, results, extrapolated into the thermodynamic limit for the XXZ model, are presented.

5.1 Violation of the BW theorem

This section shows that there are significant deviations from the BW theorem in the XXZ model and the TFIM. Additionally, an understanding about the CFT extensions, introduced in Section 2.4.1, will be provided. In the following, the XXZ model with $\Delta = 0.5$ and the TFIM with $\Gamma = 1$ will be used with OBC and PBC for both Ansätze, $\hat{H}_A^{\text{BW}}$ and $\hat{H}_A^{\text{BWV}}$, together with $N = 10$, $N_A = 5$ and $T_{\text{max}} = 1$. The minimization is run 50 times for the same model but with different random initial parameters between two and six. Outliers are filtered out, based on the value of the cost function and the ratios of the optimal parameters. If for example the cost function value is much larger, or a ratio is negative or deviates several magnitudes from the expected ratio, the results are not taken into the statistics. The optimal parameters from the QCFL are compared to the optimal parameters from the cost function

$$\mathcal{C}^{\text{comm}}(\mathbf{g}) = \frac{||[\hat{\rho}_A, \hat{H}_A^{\text{Var}}(\mathbf{g})]||_{\text{F}}}{2||\hat{\rho}_A||_{\text{F}}||\hat{H}_A^{\text{Var}}(\mathbf{g})||_{\text{F}}},$$

which measures the commutativity of the exact RDM and the variational Ansatz. The notation $||X||_{\text{F}}$ with some $n \times m$ matrix X denotes the Frobenius norm of X defined as[15]

$$||X||_{\text{F}} = \left(\sum_{i=1}^n \sum_{j=1}^m |x_{ij}|^2 \right)^{\frac{1}{2}}.$$

Here, $n = m = 2^{N_A}$ holds. No filtering of outliers is done with the results of the aforementioned commutator as a cost function $\mathcal{C}^{\text{comm}}(\mathbf{g})$.

5.1.1 TFIM

If the Ansatz $\hat{H}_A^{\text{BW}}$ is used in the case of OBC, the ratio $g_2^{\text{opt}}/g_1^{\text{opt}}$ obtained with the QCFL shows good agreement with the ratio obtained via the commutator (Figure 5.1(a)), exhibiting an absolute difference of 0.0237 ± 0.1037 . However, the third, fourth and fifth ratios deviate more with an absolute difference of up to 1.1021 ± 0.2858 . A linear rise near the entanglement cut can be observed and

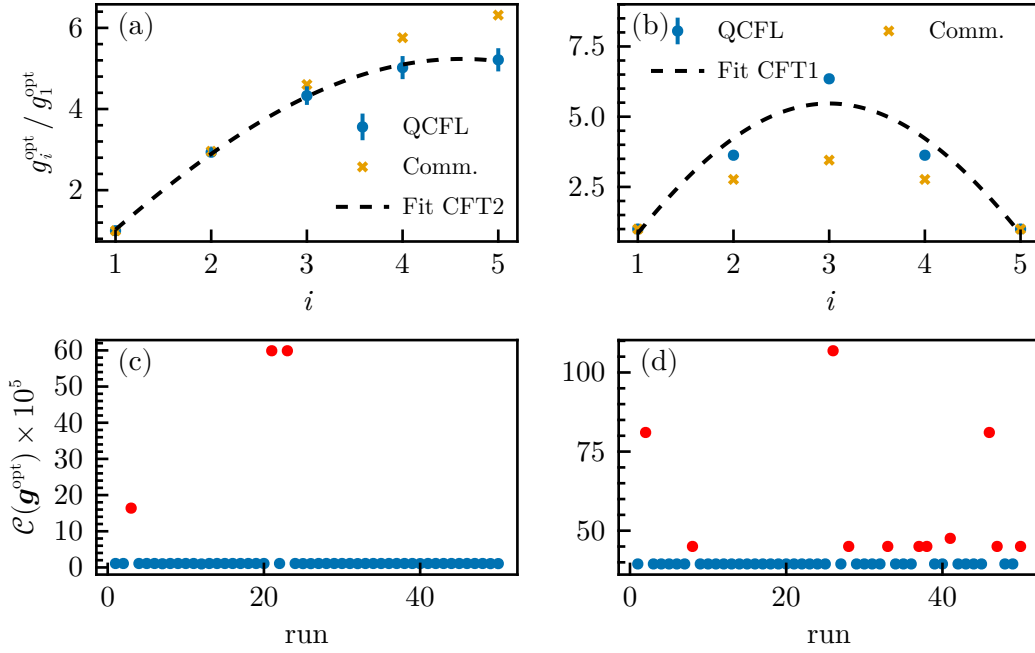


Figure 5.1: 50 runs with different random initial parameters for the TFIM with the Ansatz $\hat{H}_A^{\text{BW}}$. (a) The optimal parameters from the QCFL and from the commutator (Comm.) as a cost function for the case of OBC and (c) the corresponding cost function value at the minimum of all 50 runs of the QCFL. (b) and (d) show the same for PBC. The errorbars are given by the standard deviation of the 50 runs and the red points in the cost, (c) and (d), indicate outliers from the QCFL, which were not taken into account in (a) and (b). The dashed lines are fits in accordance to the CFT extensions in Equation (2.6) and (2.7) to guide the eye.

bending in accordance to the second CFT extension (Equation (2.7)) becomes apparent. In the case of PBC, the optimal parameters exhibit large deviations from the parameters obtained via the commutator and do not follow the first CFT extension (Equation 2.6) but rather a triangular form, while the parameters yielded with the commutator do show the behaviour predicted by the first CFT extension (Figure 5.1(b)). The maximum difference of the ratios obtained via the

QCFL and the commutator is given by the third ratio, $g_3^{\text{opt}}/g_1^{\text{opt}}$, and takes the value 2.894852 ± 0.000028 . In both cases, especially in the case of PBC, the standard deviation is very small, indicating the minimizer almost always finds the same solution. This statement is supported by the cost function value at its minimum agreeing in almost all runs (Figure 5.1(c),(d)). No errorbars are depicted for the ratios obtained with the commutator as a cost function, since no standard deviation is larger than $4 \cdot 10^{-5}$ for both, OBC and PBC. On the other hand, if the Ansatz

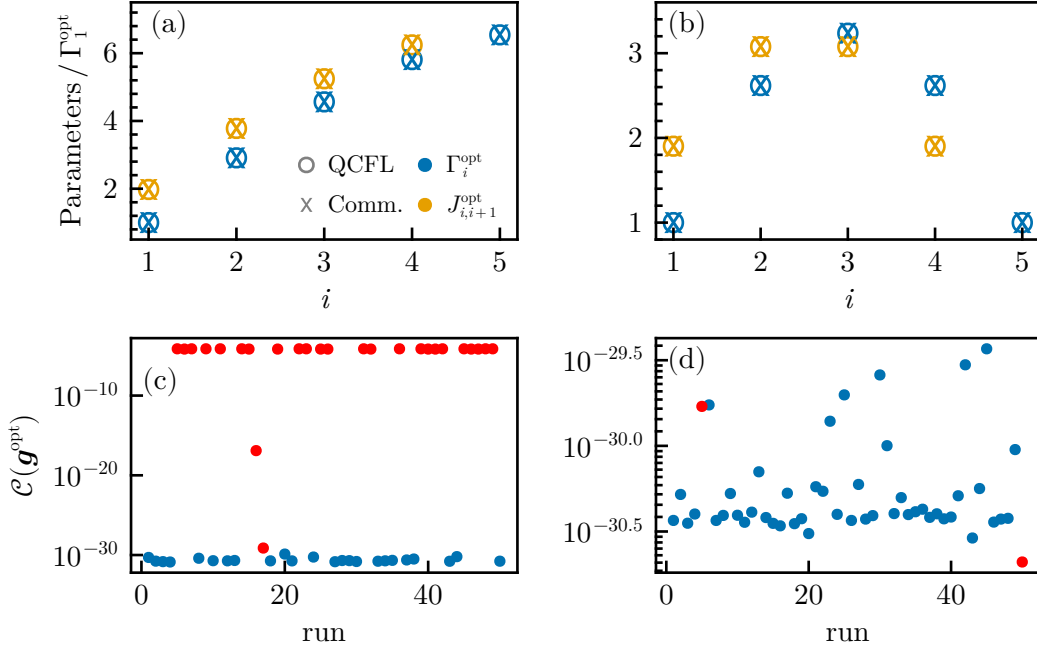


Figure 5.2: 50 runs with different random initial parameters for the TFIM with the Ansatz $\hat{H}_A^{\text{BWV}}$. (a) The optimal parameters from the QCFL and from the commutator (Comm.) as a cost function for the case of OBC and (c) the corresponding cost function value at the minimum of all 50 runs of the QCFL. (b) and (d) show the same for PBC. The red points in the cost, (c) and (d), indicate outliers, which were not taken into account in (a) and (b).

$\hat{H}_A^{\text{BWV}}$ is used, the optimal parameters obtained via the QCFL match perfectly with the parameters obtained via the commutator for both, OBC and PBC (see Figure 5.2(a),(b)). The ratios obtained via the QCFL and the commutator as a cost function exhibit a maximum deviation of $(1.2665 \pm 5.8603) \cdot 10^{-9}$ for OBC and PBC. In addition to the perfect match, after filtering out the outliers, the standard deviation of each parameter obtained via the QCFL is in the vicinity of 10^{-14} , indicating that there is one good solution, while the ratios obtained via the commutator are no greater than $6 \cdot 10^{-9}$, which is why no errorbars are depicted.

The two outliers in Figure 5.2(d) account to significantly different ratios of the optimal parameters, e.g. $\Gamma_3^{\text{opt}}/\Gamma_1^{\text{opt}} \approx -2.88 \cdot 10^{-17}$. What is very remarkable is that the cost function value at its minimum is in the vicinity of 10^{-30} . Comparing it to the minimum of the cost function of $\approx 10^{-5}$ for the BW-like Ansatz, $\hat{H}_A^{\text{BW}}$, underlines the fact that the Ansatz $\hat{H}_A^{\text{BWV}}$ works much better. This means that the BW theorem does not hold for the TIFM with $\Gamma = 1$. Additionally, as can be seen, the only corrections to the BW theorem are that there is not just one parameter per lattice, but two. Further corrections such as long range interactions or higher-body interactions are not needed, since the cost function is already numerically zero at its minimum.

5.1.2 XXZ model

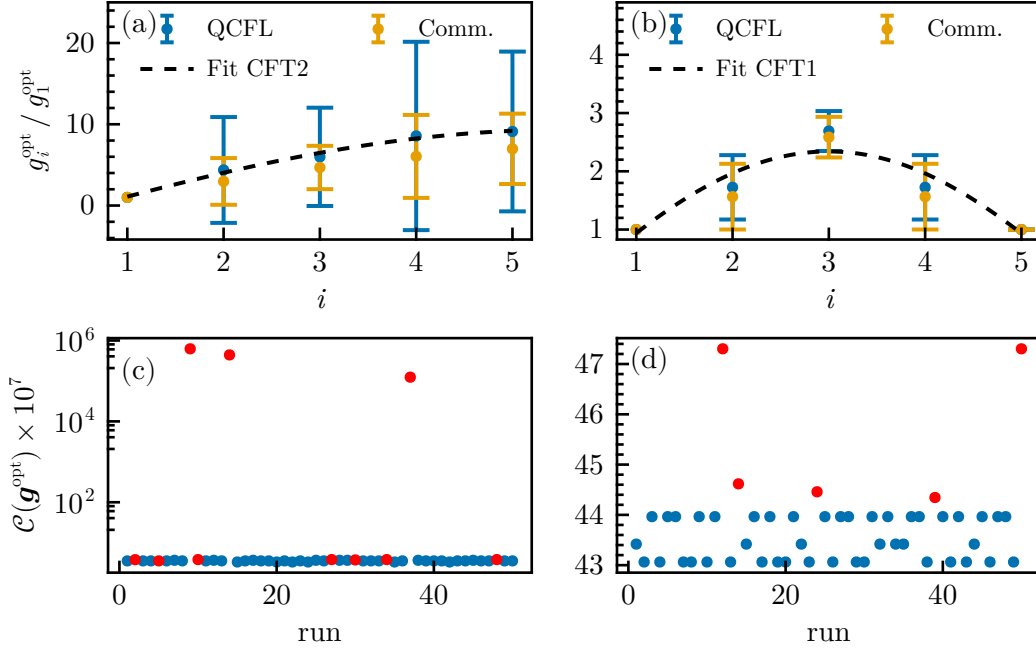


Figure 5.3: 50 runs with different random initial parameters for the XXZ model with the Ansatz $\hat{H}_A^{\text{BW}}$. (a) The optimal parameters from the QCFL and from the commutator (Comm.) as a cost function for the case of OBC and (c) the corresponding cost function value at the minimum of all 50 runs of the QCFL. (b) and (d) show the same for PBC. The errorbars are given by the standard deviation of the 50 runs and the red points in the cost, (c) and (d), indicate outliers from the QCFL, which were not taken into account in (a) and (b). The dashed lines are fits in accordance to the CFT extensions in Equation (2.6) and (2.7) to guide the eye.

The optimal parameters exhibit huge standard deviations in the case of the XXZ model with OBC and the BW-like Ansatz $\hat{H}_A^{\text{BW}}$ (Figure 5.3(a)), with a standard deviation of up to 11.6. The differences of the optimal ratios obtained via the QCFL and the commutator range from 1.3111 ± 6.6087 ($g_3^{\text{opt}}/g_1^{\text{opt}}$) up to 2.5180 ± 12.6613 ($g_4^{\text{opt}}/g_1^{\text{opt}}$). In the case of PBC, these differences range from $(0.0857 \pm 2.0805) \cdot 10^{-6}$ ($g_5^{\text{opt}}/g_1^{\text{opt}}$) up to 0.1602 ± 0.7895 ($g_2^{\text{opt}}/g_1^{\text{opt}}$) (Figure 5.3(b)). Again, the linear rise near the entanglement cut and bending at the right border becomes apparent in the case of OBC, while the symmetric behaviour can be observed if the composite system obeys PBC. However, the large standard deviation indicate that there are many suboptimal or local minima, which are found by the optimizer. Since the minimum of the cost function is far from zero, namely $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 10^{-7}$ (OBC) and $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 5 \cdot 10^{-6}$ (PBC) (Figure 5.3(c),(d)), these local minima are not the real solution. This means that there are, again, corrections to the BW theorem. Figure 5.4(a),(b) show the optimal parameters after filtering obtained with the Ansatz $\hat{H}_A^{\text{BWV}}$ for OBC and PBC, respectively. The parameters obtained with the

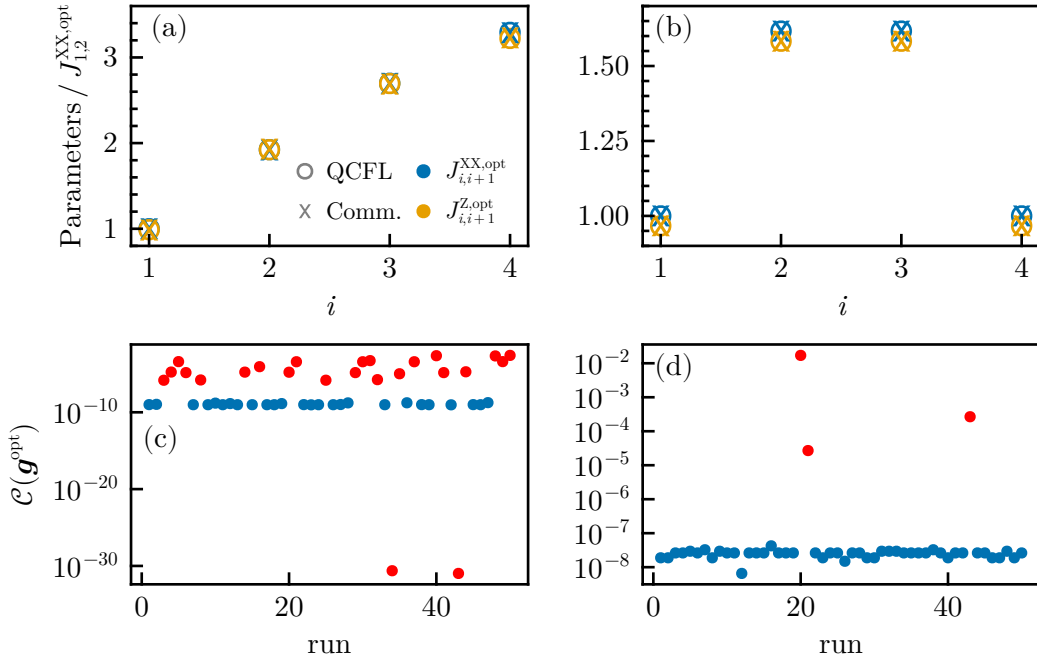


Figure 5.4: 50 runs with different random initial parameters for the XXZ model with the Ansatz $\hat{H}_A^{\text{BWV}}$. (a) The optimal parameters from the QCFL and from the commutator (Comm.) as a cost function for the case of OBC and (c) the corresponding cost function value at the minimum of all 50 runs of the QCFL. (b) and (d) show the same for PBC. The red points in the cost, (c) and (d), indicate outliers, which were not taken into account in (a) and (b).

QCFL match the parameters obtained with the commutator very well, exhibiting a maximum deviation of 0.0078 ± 0.0011 . For OBC and PBC, the standard deviations of the ratios obtained via the QCFL over the 50 runs, after filtering out, and the ratios obtained via the commutator are no greater than $2 \cdot 10^{-3}$, s.t. no error bars given in Figure 5.4(a),(b). Remarkably, in the case of OBC, the parameters only deviate slightly from the BW theorem. That is, the ratios of the coupling in the x - and y -direction, and the coupling in z -direction, $J_{i,i+1}^{\text{XX,opt}}/J_{i,i+1}^{\text{Z,opt}}$, is 1.0219 ± 0.0005 at most, which agrees up to one decimal place with the BW theorem, predicting a ratio of one. The deviation of the parameters from the BW theorem in the case of PBC is similar. However, the deviation from the BW theorem is not negligible, underlined by the drop in the cost function value by approximately three (OBC) and two (PBC) orders of magnitude if the BW-violating Ansatz is used (see Figure 5.4(c),(d)). Although the cost function is noticeably lower with the Ansatz $\hat{H}_A^{\text{BWV}}$ and parameters obtained via the QCFL and the commutator match very well, the minimum of the cost function is still $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 10^{-10}$ and $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 10^{-8}$ for OBC and PBC, respectively. That means, on top of multiple parameters per lattice site, additional corrections to the BW theorem need to be taken into account. That is, long range interactions and more-body interactions need to be included into the variational Ansatz. Long range interactions included in the Ansatz $\hat{H}_A^{\text{BWV}}$ for the XXZ model are discussed in the next section.

To conclude, the algorithm approximately finds the progression of the optimal parameters predicted by the CFT extensions. In case of OBC the linear rise at the entanglement cut and the bending at the right border could be found. The systems with PBC exhibit the symmetric behaviour of the optimal parameters with linear rises at both entanglement cuts. The main finding in this section is that there are significant corrections to the BW theorem, signalling that the BW theorem is violated for lattice systems. For the TFIM, instead of a deformed Hamiltonian, the real EH needs two parameters per site, but no additional corrections are needed. On the other hand, the XXZ model needs additional long range or more-body interactions on top of the two parameters per site.

5.2 Long range corrections in the XXZ model

One implication of the BW theorem is that if the system Hamiltonian contains only nearest neighbour interactions, then the EH will, too. However, it was already shown that the BW theorem is not exact on lattices. Besides more than one variational parameter per lattice site, long range interactions as further corrections to the BW theorem for the XXZ model will be investigated. The XXZ model with OBC is used exactly as in the previous section (Section 5.1) and it will be examined how much

further the cost function can be minimized. The variational Ansatz, together with the corrections $\hat{H}_A^{\text{corr}}$, reads

$$\begin{aligned}\hat{H}_A^{\text{Var}} &= \hat{H}_A^{\text{BWV}} + \hat{H}_A^{\text{corr}} \\ &= \sum_{r=1}^{r_{\text{max}}} \sum_{i=1}^{N_A-r} (J_{i,i+r}^{\text{XX}} (X_i X_{i+r} + Y_i Y_{i+r}) + J_{i,i+r}^{\text{Z}} \Delta Z_i Z_{i+r}),\end{aligned}$$

where $\{J_{i,i+r}^{\text{XX}}, J_{i,i+r}^{\text{Z}}\}$ act as variational parameters. The quantity r_{max} determines the maximum range of interaction. Every term beyond $r = 1$ is a part of the long range interactions, and thus, part of the corrections. The optimal parameters from Section 5.1 of the BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ are used for initialization of the parameters for $r = 1$. All long range couplings (beyond $r = 1$) are initialized to zero, as these are expected to be small. Table 5.1 lists the optimal parameters for a run with $r_{\text{max}} = 4$,

Table 5.1: Optimal parameters $\{J_{i,i+r}^{\text{XX}}, J_{i,i+r}^{\text{Z}}\}$ including all long range interactions for $r_{\text{max}} = 4$ rounded to five decimal places.

	$r = 1$	$r = 2$	$r = 3$	$r = 4$
$J_{1,1+r}^{\text{XX}}$	0.843 46	0.005 56	0.001 25	0.003 22
$J_{1,1+r}^{\text{Z}}$	0.882 05	0.048 18	0.044 73	0.047 37
$J_{2,2+r}^{\text{XX}}$	1.623 81	−0.008 05	0.002 66	
$J_{2,2+r}^{\text{Z}}$	1.637 85	0.011 11	0.026 54	
$J_{3,3+r}^{\text{XX}}$	2.322 76	0.067 28		
$J_{3,3+r}^{\text{Z}}$	2.326 81	0.097 54		
$J_{4,4+r}^{\text{XX}}$	2.843 32			
$J_{4,4+r}^{\text{Z}}$	2.680 56			

i.e., all long range terms included. It can be seen that the corrections are at least one magnitude smaller in comparison to the parameters for $r = 1$. Additionally, the parameters exhibit antiferromagnetic behaviour (positive parameters) and one antiferromagnetic coupling can be seen, although only antiferromagnetic couplings are included in the system Hamiltonian. The parameters show, as expected, a decay with the range of interaction r . Of course, if the variational Ansatz contains long range corrections, the found solution should represent the exact EH better, and thus, the cost function value at the minimum should be lower, which is exactly the case, as can be seen in Figure 5.5, where the minimizer was run for all possible maximum interactions ranges r_{max} . The case $r_{\text{max}} = 1$ corresponds to the absence of any long range corrections. It needs to be mentioned that convergence problems

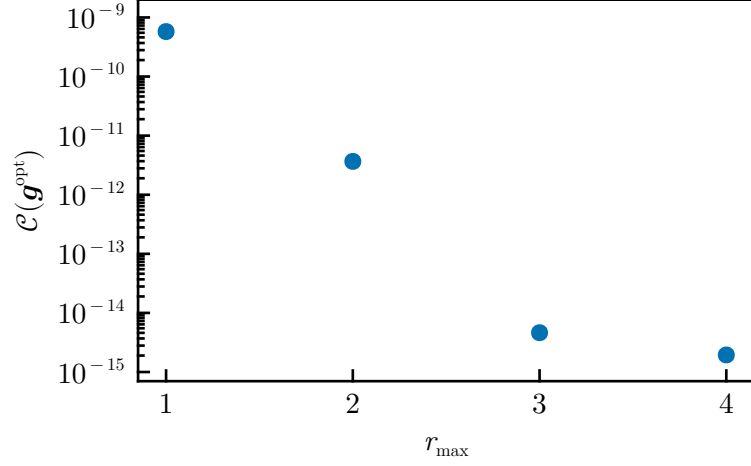


Figure 5.5: Cost function $\mathcal{C}(g^{\text{opt}})$ at the found solution vs. the maximum range of interaction $r_{\max}$ included in the correction term.

were present for $r_{\max} = 3$ and $r_{\max} = 4$. The infinity norm of the gradient was in the vicinity of 10^{-14} most of the time, and thus, the optimization procedure was very slow. After 100000 iterations, the minimization run has been terminated. To conclude, the XXZ model, in case of PBC and $\Delta = 0.5$, needs long range interaction included in the variational Ansatz to reconstruct the exact EH. With the help of long range corrections, the minimum of the cost function could be reduced by approximately five orders of magnitude.

5.3 Comparison of the Entanglement spectra

This section compares the universal ratios (see Equation (2.10)) of the variational solutions of the two previous sections, since the ES (equivalently the universal ratios) is the quantity of main interest. For the universal ratios, $\alpha_0 = 1$ and $\alpha_1 = 5$ is used. The BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ appeared to be exact for the TFIM with OBC and $\Gamma = 1$ (see Section 5.1). Thus, only the universal ratios from the BW-like Ansatz $\hat{H}_A^{\text{BW}}$ and BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ will be compared. Since the Ansatz $\hat{H}_A^{\text{BWV}}$ is not exact for the XXZ model for OBC and $\Delta = 0.5$, the universal ratios of the BW-violating Ansatz with all long range corrections ($r_{\max} = 4$) (see Section 5.2) will be taken into account. The exact ES is computed by diagonalizing the exact EH, given by $\hat{H}_A = -\ln(\hat{\rho}_A)$, exactly.

In the case of the TFIM, the low-lying universal ratios match the exact universal ratios better if the BW-violating Ansatz is used, as expected (see Figure 5.6(a),(b)).

In contrast, in the higher part of the spectrum, the universal ratios obtained via

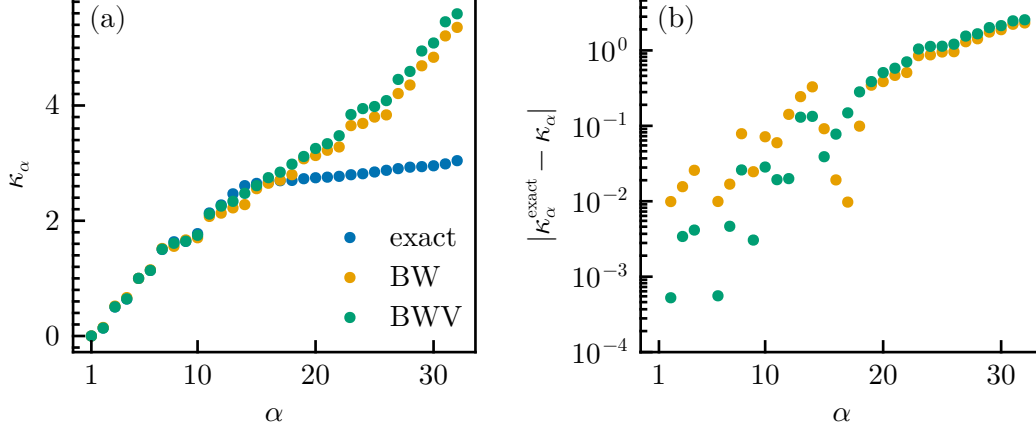


Figure 5.6: (a) Universal ratios and (b) deviations from the exact universal ratios for the BW-like Ansatz $\hat{H}_A^{\text{BW}}$ (BW) and BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ (BWV) in the TFIM.

variation show significant deviations from the exact universal ratios (Figure 5.6(b)). This is mostly a numerical artifact, which is due to the double precision used. The RDM contains (eigen-)values, which are so small s.t. they cannot be captured with double precision accurately. Since the ES (and thus the universal ratios) is obtained by taking the logarithm of the RDM, the lowest eigenvalues of the RDM are mapped to the highest universal ratios. Thus, to compare the universal ratios, only the first ten universal ratios are taken into account. The first ten universal ratios exhibit a mean absolute deviation from the exact universal ratios of $\overline{\Delta\kappa_\alpha^{\text{BW}}} = 0.0253$ and $\overline{\Delta\kappa_\alpha^{\text{BWV}}} = 0.0071$ for the Ansatz $\hat{H}_A^{\text{BW}}$ and $\hat{H}_A^{\text{BWV}}$, respectively. That is, the low-lying spectrum (here, the first ten universal ratios) is reconstructed more than three times more accurately on average if the BW-violating Ansatz is used.

The universal ratios in the case of the XXZ model are given in Figure 5.7(a). As can be seen in Figure 5.7(b), the low-lying spectrum is not significantly better reconstructed if the Ansatz $\hat{H}_A^{\text{BWV}}$ or the Ansatz $\hat{H}_A^{\text{BWV}}$ with all long range corrections is used. For comparison, two additional measures are listed here, which involve the RDM on subsystem A obtained through the minimization

$$\hat{\rho}_A^{\text{Var}} = \frac{1}{\text{Tr} [e^{-\hat{H}_A^{\text{Var}}(\mathbf{g}^{\text{opt}})}]} e^{-\hat{H}_A^{\text{Var}}(\mathbf{g}^{\text{opt}})}.$$

The first measure is the trace distance

$$\mathcal{J}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}}) = \frac{1}{2} \text{Tr} \left[\sqrt{(\hat{\rho}_A - \hat{\rho}_A^{\text{Var}})^\dagger (\hat{\rho}_A - \hat{\rho}_A^{\text{Var}})} \right] = \frac{1}{2} \text{Tr} \left[\sqrt{(\hat{\rho}_A - \hat{\rho}_A^{\text{Var}})^2} \right],$$

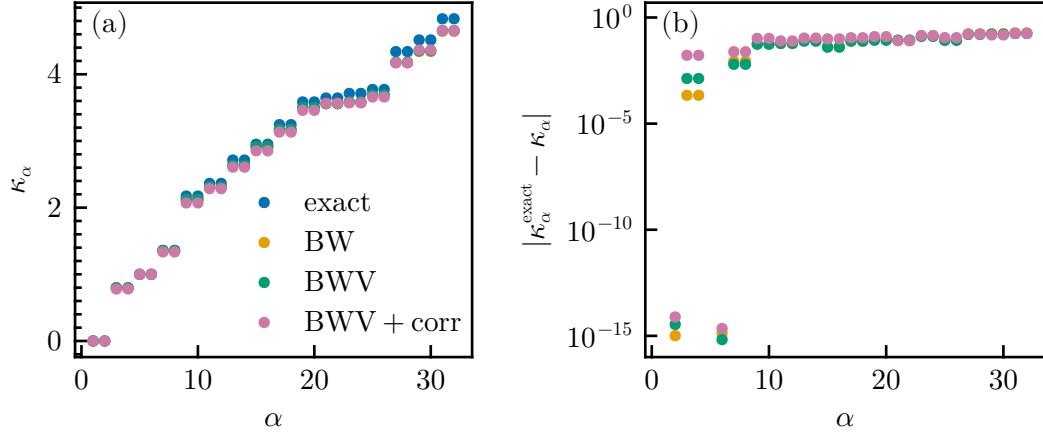


Figure 5.7: (a) Universal ratios and (b) deviations from the exact universal ratios for the BW-like Ansatz $\hat{H}_A^{\text{BW}}$ (BW), BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ (BWV) and the BW-violating Ansatz with all long range corrections (BWV+corr) in the XXZ model.

which measures how close two quantum states are and ranges from 0 (identical states) to 1 (maximally distant states) [38]. The last equality holds since both density matrices are hermitian. The second measure utilizes the commutator of $\hat{\rho}_A$ and $\hat{\rho}_A^{\text{Var}}$

$$\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}}) = \frac{\|[\hat{\rho}_A, \hat{\rho}_A^{\text{Var}}]\|_{\text{F}}}{2\|\hat{\rho}_A\|_{\text{F}}\|\hat{\rho}_A^{\text{Var}}\|_{\text{F}}}, \quad (5.1)$$

which, again, ranges from 0 (completely commuting) to 1 (maximally non-commutative) [7]. This measure is included, since the cost function is based on the commutativity of the variational Ansatz and the exact RDM. Table 5.2 lists the mean of the

Table 5.2: All included measures to compare the accuracy of the variational Ansätze, BW-like Ansatz (BW), BW-violating Ansatz (BWV) and the BW-violating Ansatz with all long range corrections (BWV+corr) for the XXZ model.

Ansatz	$\overline{\Delta\kappa_\alpha}$	$\mathcal{T}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$	$\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$	$\mathcal{C}(\mathbf{g}^{\text{opt}})$
BW	0.013 09	0.007 51	$4.173\,79 \cdot 10^{-4}$	$3.550\,22 \cdot 10^{-7}$
BWV	0.012 56	0.019 83	$9.321\,76 \cdot 10^{-5}$	$5.768\,21 \cdot 10^{-10}$
BWV+corr	0.028 87	0.030 30	$1.373\,17 \cdot 10^{-6}$	$1.945\,40 \cdot 10^{-15}$

absolute deviations of the first ten universal ratios from the exact universal ratios $\overline{\Delta\kappa_\alpha}$, the trace distance $\mathcal{T}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$, the norm of the commutator $\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$ and the cost function value at its minimum $\mathcal{C}(\mathbf{g}^{\text{opt}})$. It can be seen that $\mathcal{F}(\hat{\rho}_A, \hat{\rho}_A^{\text{Var}})$

is lower if the Ansatz $\hat{H}_A^{\text{BWV}}$ is used and the lowest if the long range corrections are included. This is in agreement with the cost function value, which shows the same trend. This observation makes sense, since the cost function is based on the commutativity of the exact RDM and the variational Ansatz. However, the trace distance is the highest for the Ansatz $\hat{H}_A^{\text{BWV}}$ with long range interactions included and the lowest for the BW-like Ansatz $\hat{H}_A^{\text{BW}}$. Thus, the universal ratios are not reconstructed more accurately with the Ansatz $\hat{H}_A^{\text{BWV}}$ with long range interactions, since the trace distance is a measure for how close two quantum states are.

5.4 TFIM and XXZ model across the phase diagram

This section investigates the TFIM and XXZ model across the respective phase diagrams to analyse how the algorithm performs when the systems are not critical and whether the algorithm can indicate the critical points or specific phases. Both physical models obey OBC and the chain lengths are varied, while $N = 2N_A$ always holds. The Ansatz $\hat{H}_A^{\text{BWV}}$ is used without any other corrections for both models and the maximum integration time is set to $T_{\text{max}} = 1$.

Figure 5.8 shows the minimum of the cost function for varying the transverse field strength in the TFIM. At $\Gamma = 0$ the minimum of the cost function is below 10^{-29}

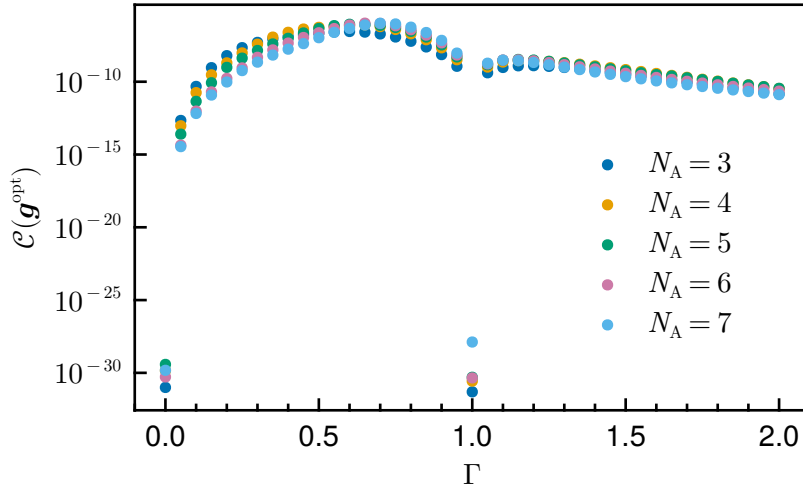


Figure 5.8: Minimum of the cost function in dependence on the transversal field strength Γ for the TFIM.

for all N_A . Assuming the exact EH takes the form

$$\hat{H}_A = \sum_{i=1}^{N_A-1} \tilde{J}_{i,i+1} Z_i Z_{i+1},$$

at $\Gamma = 0$, the exact EH always commutes with the variational Ansatz, since only the Pauli matrix Z occurs in both, the exact EH and the variational Ansatz. That is, the cost function is numerically zero. The second eye-catching point is $\Gamma = 1$, where the cost function drops near zero again. This is due to the fact that the quantum phase transition, in the thermodynamic limit, is at exactly at that point.

The minimum of the cost function has three interesting points in the case of the XXZ model, namely $\Delta = -1$, $\Delta = 0$ and $\Delta = 1$ (see Figure 5.9). The abrupt decrease

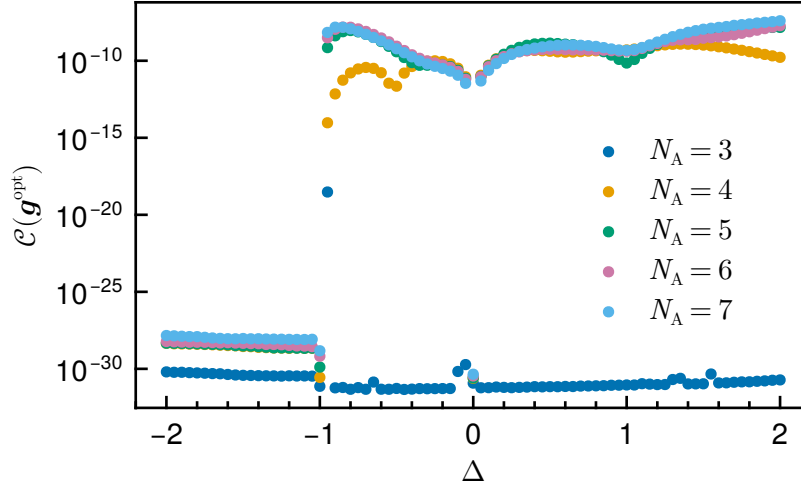


Figure 5.9: Minimum of the cost function in dependence on the anisotropy Δ for the XXZ model.

of the minimum at $\Delta = -1$ can be explained by the first order phase transition occurring. For $\Delta < -1$ the ground state of the XXZ model is a simple product state with all spins pointing in the same direction, and thus, no entanglement is present in the composite system. The second order phase transition can be recognized by the cusp-like behaviour of the cost function at $\Delta = 1$. For $N_A = 3$, the BW-violating Ansatz works very well, since the cost function value at its minimum is below 10^{-30} except at $\Delta = -0.95$, where $\mathcal{C}(\mathbf{g}^{\text{opt}}) \approx 10^{-19}$ holds. For $\Delta \leq -1$, the couplings in the z -direction match the couplings in the x - and y -direction very accurately up to at least thirteen digits for all N_A .

To conclude, the algorithm delivers indications for the quantum phase transitions and classical states. Additionally, the accuracy of the BW-violating Ansatz varies

across the phase diagram. Thus, for certain values of Γ and Δ , where the cost function is very large at its minimum, the variational Ansatz needs to be adjusted.

5.5 Violation of the BW theorem in the XXZ model in the thermodynamic limit

This section shows the violation of the BW theorem in the XXZ model in the thermodynamic limit (TDL). The idea was initialized by Reference [39], where it was theoretically shown that the BW theorem is violated in the XXZ model in the TDL. A global minus sign is included in the Hamiltonian, the system obeyed PBC and the anisotropy was set to $\Delta = -0.5$

$$\hat{H}_{\text{XXZ}} = - \sum_{i=1}^N (X_i X_{i+1} + Y_i Y_{i+1} + \Delta Z_i Z_{i+1}). \quad (5.2)$$

It was stated that the couplings of the Ansatz $\hat{H}_A^{\text{BWV}}$ obeyed a ratio of $J_{i,i+1}^{\text{XX}}/J_{i,i+1}^{\text{Z}} \approx 1.1$ instead of, predicted by the BW theorem, $J_{i,i+1}^{\text{XX}}/J_{i,i+1}^{\text{Z}} = 1$. No long range or more-body interactions are taken into account, because the authors in Reference [39] mentioned that these contributions are more than one order of magnitude smaller than the contributions of the nearest neighbour coupling of two spins. To obtain the couplings $J_{i,i+1}^{\text{XX}}$ and $J_{i,i+1}^{\text{Z}}$ in the TDL, the procedure is the following:

1. Extract the ground state of the system Hamiltonian (Equation 5.2) for different lattice sizes N (up to $N = 29$ could have been achieved with the Lancos algorithm).
2. Construct the RDM with the ground state obtained in step one for a subsystem chain length N_A .
3. Run the algorithm with the BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$ with $J_{i,i+1}^{\text{XX}}$ and $J_{i,i+1}^{\text{Z}}$ as variational parameters for the different RDMs for each composite system size N from step 2 for a subsystem chain length N_A .
4. Plot the ratios of the obtained parameters $J_{i,i+1}^{\text{XX}}/J_{1,2}^{\text{XX}}$ and $J_{i,i+1}^{\text{Z}}/J_{1,2}^{\text{XX}}$ vs. $1/N$ (even N) and $1/N^2$ (odd N) and extrapolate for $1/N \rightarrow 0$, i.e., into the TDL.
5. Repeat step two to four for different subsystem chain lengths N_A .

This procedure is done once for odd and once for even N and the previously mentioned settings together with $T_{\text{max}} = 5$ are used in the following two subsections, to see whether this algorithm can reproduce the findings in Reference [39]. Odd and even N are treated separately, since the optimal parameters exhibit a different

dependence on N if N is odd or even. The subsystem chain length ranges from $N_A = 4$ to $N_A = 7$. To measure the deviation from the BW theorem, the quantity

$$\theta_i = \frac{J_{i,i+1}^{XX}}{J_{i,i+1}^Z} - 1 \quad (5.3)$$

is defined, which will be referred to as the discrepancy. All runs were initialized with a good initial guess.

5.5.1 Odd number of lattice sites in the composite system

Figure 5.10 shows the obtained optimal parameters normalized to $J_{1,2}^{XX}$ vs. $1/N^2$ for $N_A = 7$. The index i indicates the lattice site and the solid lines are the corresponding fits. For brevity, the plots for other N_A are omitted. A linear fit was

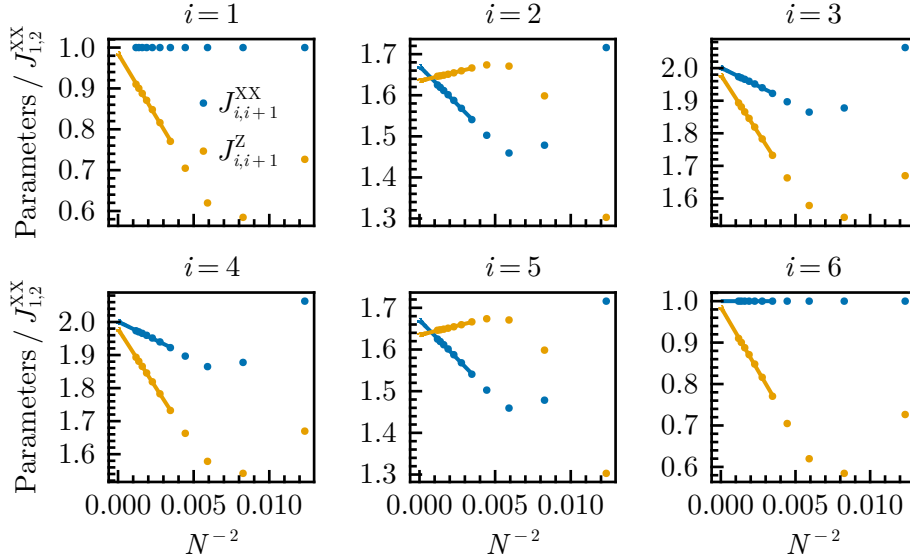


Figure 5.10: Optimal parameters $J_{i,i+1}^{XX}$ and $J_{i,i+1}^Z$ normalized to $J_{1,2}^{XX}$ vs. $1/N^2$ for each lattice site i in the subsystem A for $N_A = 7$. The solid lines are linear fits.

used

$$\frac{J_{i,i+1}^\Omega}{J_{1,2}^{XX}}(N) = p_1 \frac{1}{N^2} + p_2, \quad \Omega = XX, Z \quad (5.4)$$

where p_1 and p_2 act as parameters for the fit. The optimal parameters normalized to $J_{1,2}^{XX}$ extrapolated in to the TDL ($N \rightarrow \infty$) then are

$$\frac{J_{i,i+1}^\Omega}{J_{1,2}^{XX}}(N \rightarrow \infty) = p_2, \quad \Omega = XX, Z.$$

5.5 Violation of the BW theorem in the XXZ model in the thermodynamic limit

Finally, the extrapolated parameters normalized to $J_{1,2}^{\text{XX}}$ are given in Figure 5.11, where the solid lines are there to guide the eye and take the quadratic form $\propto i \frac{N_A - i}{N_A}$, as suggested by the conformal extension $H_A^{\widehat{\text{CFT}}3}$ (see Equation 2.8), although the CFT extensions apply only to the BW theorem. No error bars are given, since the propagated estimated uncertainties of the parameters of the fit (see Equation (5.4)) are not larger than $2 \cdot 10^{-3}$. The gap between the x - and y -couplings, and

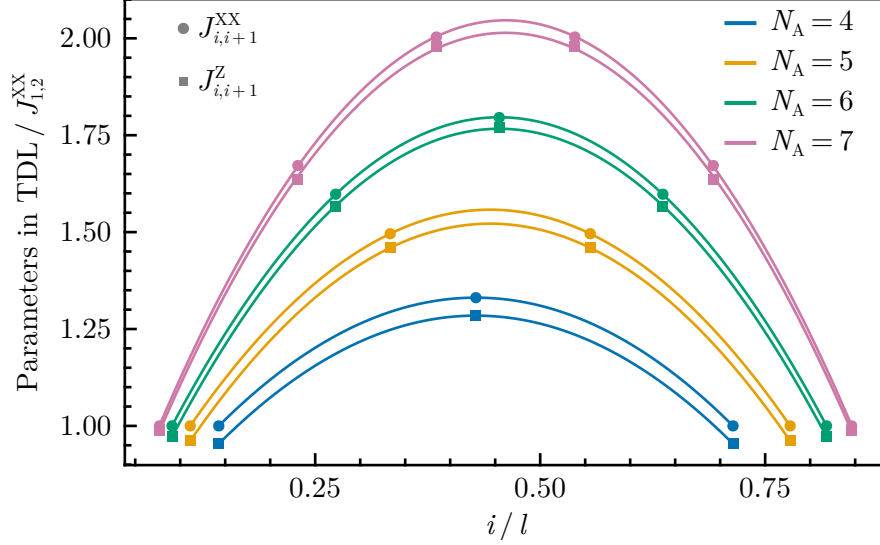


Figure 5.11: Optimal parameters $J_{i,i+1}^{\text{XX}}$ and $J_{i,i+1}^Z$ normalized to $J_{1,2}^{\text{XX}}$ extrapolated into the TDL vs. the lattice site i in units of the subsystem chain length l for each number of sites N_A in the subsystem A.

the z -couplings can be seen. Table 5.3 lists the mean value of the discrepancies, as

Table 5.3: The mean value of the discrepancies as defined in Equation (5.3) in the TDL for each subsystem size N_A .

N_A	$\bar{\theta}_i$
4	0.0443 ± 0.0002
5	0.0326 ± 0.0006
6	0.0224 ± 0.0004
7	0.0154 ± 0.0006

defined in Equation (5.3), over the lattice sites i for all subsystem lattice sizes N_A in the TDL. The errors given in Table 5.3 are the propagated errors, stemming from

the estimated uncertainties of the fit.

5.5.2 Even number of lattice sites in the composite system

Figure 5.12 shows the obtained optimal parameters normalized to $J_{1,2}^{XX}$ vs. $1/N$ for $N_A = 7$. For even number of lattice sites N in the composite system, the fit function

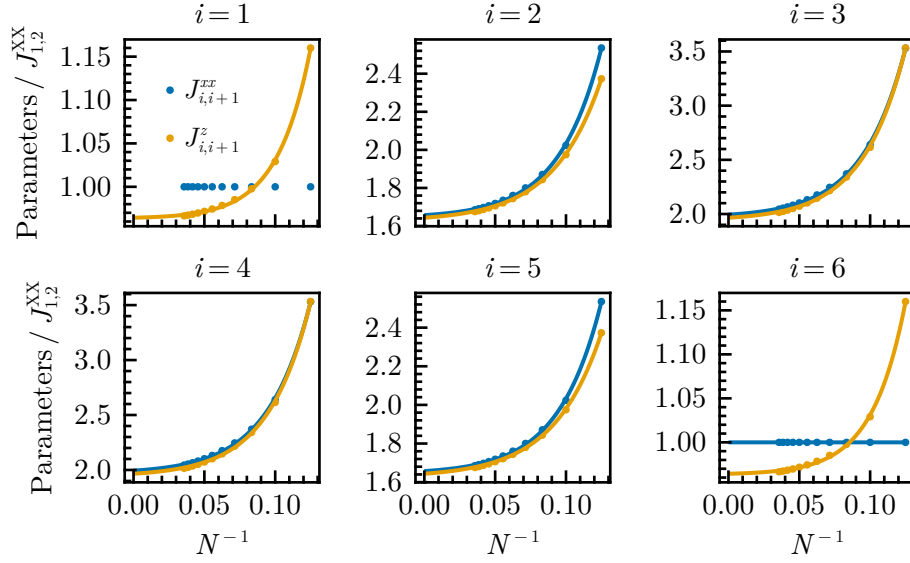


Figure 5.12: Optimal parameters $J_{i,i+1}^{XX}$ and $J_{i,i+1}^Z$ normalized to $J_{1,2}^{XX}$ vs. the composite system chain length N for each lattice site i in the subsystem A for $N_A = 7$. The solid lines are fits according to Equation (5.5).

is chosen as

$$\frac{J_{i,i+1}^{\Omega}}{J_{1,2}^{XX}}(N) = p_1 e^{\frac{p_2}{N}} + p_3, \quad \Omega = XX, Z, \quad (5.5)$$

where p_1, p_2, p_3 and p_4 act as parameters for the fit, which turned out to work well, as can be seen in Figure 5.12. The optimal parameters extrapolated in to the TDL ($N \rightarrow \infty$) then are

$$\frac{J_{i,i+1}^{\Omega}}{J_{1,2}^{XX}}(N \rightarrow \infty) = p_1 + p_3, \quad \Omega = XX, Z$$

and are depicted in Figure 5.13. The propagated errors stemming from the estimated uncertainties of the fit are not displayed, since no error is bigger than $3 \cdot 10^{-3}$. Again, differences between $J_{i,i+1}^{XX}$ and $J_{i,i+1}^Z$ can be seen. The discrepancies, as defined in

5.5 Violation of the BW theorem in the XXZ model in the thermodynamic limit

Table 5.4: The mean value of the discrepancies as defined in Equation (5.3) in the TDL for each subsystem size N_A .

N_A	$\bar{\theta}_i$
4	0.0459 ± 0.0021
5	0.0383 ± 0.0017
6	0.0271 ± 0.0014
7	0.0199 ± 0.0020

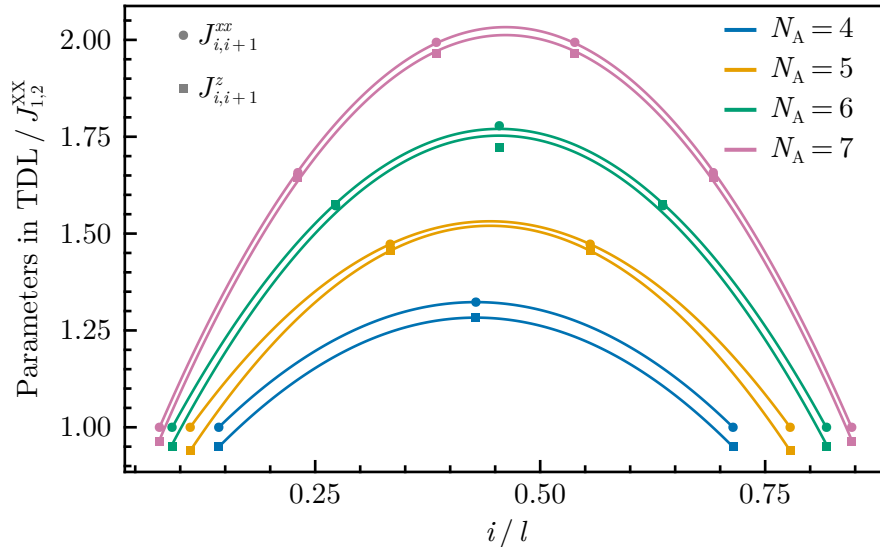


Figure 5.13: Optimal parameters $J_{i,i+1}^{XX}$ and $J_{i,i+1}^Z$ normalized to $J_{1,2}^{XX}$ extrapolated into the TDL vs. the lattice site i in units of the subsystem chain length l for each number of sites N_A in the subsystem A.

Equation (5.3), are given in Table 5.4. Like in the case of odd N , the discrepancies decrease as N_A increases.

To conclude, a discrepancy could have been observed in the TDL. However, instead of the discrepancies of ≈ 0.1 from Reference [39], the discrepancies found with the algorithm of this work are two to five times lower, depending on the subsystem chain length N_A . To reach the same result, long range or many-body corrections could be included.

6 Summary and outlook

The algorithm presented in Reference [26] delivers a good way to obtain the Entanglement Hamiltonian of lattice systems, which is very difficult to determine otherwise. The algorithm was successfully implemented for this work, resulting in a well documented and optimized `julia` package. The main goal of the package is to be easily extensible to other lattice models and to be easy to use with as few as possible lines of code for the user. At first, the derivation of the correct formula for the gradient was a challenge, but with the concept of the Fréchet derivative, an exact expression was found and the gradient is implemented in an efficient Reverse mode approach. Additionally, the optimal integration and optimization algorithm for this problem was identified through benchmarking. Thus, the EH can be obtained with minimal runtime.

The main goal of this thesis was to provide a fundamental understanding about the convergence properties. Huge improvements in regard to convergence of the results could be achieved. On the one hand, a new interpretation of the cost function has been employed. Instead of monitoring observables at a few arbitrary time points, i.e., interpreting the cost function as a discrete sum, the cost function is promoted to a continuous integral over the time domain. The convergence of the optimal parameters, at least in respect to the number of observation times, was observed systematically with that improvement. An efficient numerical integration method, the Tanh-sinh quadrature, is used to evaluate the cost function. In addition to the number of observation times, the convergence in respect to the maximum observation time has been investigated. It could be shown that no convergence can be achieved if the variational Ansatz is not capable of reconstructing the exact EH accurately, while the optimal parameters converge perfectly for a good variational Ansatz, where the maximum observation time and the number of observation times are irrelevant. With the aforementioned investigations of convergence, results for the actual form of the EH could be yielded. The main result of this work is that there are significant corrections to the BW theorem. With a modified version of the variational Ansatz, the BW-violating Ansatz $\hat{H}_A^{\text{BWV}}$, it was possible to push down the cost function to numerically zero. Furthermore, the violation of the BW theorem in the XXZ model could be shown in the TDL.

A useful property of the algorithm is that, with the help of the BW-violating Ansatz, the phase diagram could be investigated and interesting regimes and points across the phase diagram were signaled by a drop of the cost function. Classical states, i.e.

simple product states, and phase transitions could be recognized. Additionally, the applicability of the conformal extensions of the BW theorem was investigated by tuning the Hamiltonian parameters in such a way that the systems exhibit quantum critical behavior for OBC and PBC. For the OBC, a bending, as the second CFT extension suggests, was observed. Especially the results for the PBC were satisfying, since the parameters accurately matched the expectation of a symmetric behavior, predicted by the second CFT extension. Even for the BW-violating Ansatz, such bending and symmetry could be observed.

In this work, only $\{Z_i Z_{i+1} | 1 \leq i < N_A - 1\}$ were used as observables. The comparison of different observables would be of interest. Most of the runtime is spent on matrix multiplication, especially with dense matrices. This presents a huge hurdle in regard to computation time due to the exponential scaling of the Hilbert space dimension, leading to an exponential scaling of the computation time with the number of spins in the subsystem. The scaling prohibits studying larger subsystem sizes. One way to simulate larger system sizes is porting the implementations from the CPU onto GPUs.

A further outlook is to tackle the problem with the ES, which is the target quantity. Although long range corrections were included in the variational Ansatz $\hat{H}_A^{\text{BWV}}$ for the XXZ model and the minimum of the cost function decreased, the exact ES was not reconstructed more accurately. The problem was identified by using the trace distance as an accuracy measure. Although the norm of the commutator of the variational Ansatz and the exact RDM was smaller, the trace distance increased. This could mean that the eigenbasis of the exact RDM is reconstructed better, while the eigenvalues are not. That is, the applicability of the algorithm to learn the eigenvalues can be questioned. As the algorithm targets the commutator of the exact EH and the variational Ansatz, the eigenbasis is correctly captured by the algorithm, but the eigenvalues might not be correctly reproduced. To make accurate statements, further investigations are needed. Finally, the higher ES cannot be learnt with double precision, since very small values cannot be accurately represented by the used double precision. On the one hand, a higher precision could prove helpful. On the other hand, this would increase the computation time even more.

Bibliography

- [1] David H. Bailey, Karthik Jeyabalan, and Xiaoye S. Li. “A comparison of three high-precision quadrature schemes.” In: *Experimental Mathematics* 14.3 (2005), pp. 317–329. DOI: [10.1080/10586458.2005.10128931](https://doi.org/10.1080/10586458.2005.10128931).
- [2] Jeff Bezanson et al. “Julia: A fresh approach to numerical computing.” In: *SIAM review* 59.1 (2017), pp. 65–98. DOI: <https://doi.org/10.1137/141000671>.
- [3] Joseph J. Bisognano and Eyvind H. Wichmann. “On the duality condition for a Hermitian scalar field.” In: *Journal of Mathematical Physics* 16.4 (Apr. 1975), pp. 985–1007. DOI: <https://doi.org/10.1063/1.522605>.
- [4] Joseph J. Bisognano and Eyvind H. Wichmann. “On the duality condition for quantum fields.” In: *Journal of Mathematical Physics* 17.3 (Mar. 1976), pp. 303–321. DOI: <https://doi.org/10.1063/1.522898>.
- [5] M. Cerezo et al. “Variational quantum algorithms.” In: *Nature Reviews Physics* 3.9 (Aug. 2021), pp. 625–644. DOI: [10.1038/s42254-021-00348-9](https://doi.org/10.1038/s42254-021-00348-9).
- [6] *ChainRules*. Version 1.72.2. 2024. URL: <https://juliadiff.org/ChainRulesCore.jl/stable/>.
- [7] Man-Duen Choi. “Almost Commuting Matrices Need not be Nearly Commuting.” In: *Proceedings of the American Mathematical Society* 102.3 (1988), pp. 529–533. DOI: <https://doi.org/10.1090/S0002-9939-1988-0928973-3>.
- [8] Marcello Dalmonte et al. “Entanglement Hamiltonians: From Field Theory to Lattice Models and Experiments.” In: *Annalen der Physik* 534.11 (2022), p. 2200064. DOI: <https://doi.org/10.1002/andp.202200064>.
- [9] G. De Chiara et al. “Entanglement Spectrum, Critical Exponents, and Order Parameters in Quantum Spin Chains.” In: *Phys. Rev. Lett.* 109 (23 Dec. 2012), p. 237208. DOI: [10.1103/PhysRevLett.109.237208](https://doi.org/10.1103/PhysRevLett.109.237208).
- [10] Ming-Ming Du et al. “Visualizing quantum phase transitions in the XXZ model via the quantum steering ellipsoid.” In: *Phys. Rev. A* 104 (1 July 2021), p. 012418. DOI: [10.1103/PhysRevA.104.012418](https://doi.org/10.1103/PhysRevA.104.012418).

- [11] Benedikt Fauseweh. “Analysis of the transverse field Ising model by continuous unitary transformations.” Publicly available at https://cmt.physik.tu-dortmund.de/storages/cmt-physik/r/uhrig/master/master_Benedikt_Fauseweh_2012.pdf. MA thesis. Technische Universität Dortmund, 2012.
- [12] Benedikt Fauseweh. “Quantum many-body simulations on digital quantum computers: State-of-the-art and future challenges.” In: *Nature Communications* 15.1 (Mar. 2024), p. 2123. DOI: 10.1038/s41467-024-46402-9.
- [13] Lukasz Fidkowski. “Entanglement Spectrum of Topological Insulators and Superconductors.” In: *Phys. Rev. Lett.* 104 (13 Apr. 2010), p. 130502. DOI: <https://doi.org/10.1103/PhysRevLett.104.130502>.
- [14] G. Giudici et al. “Entanglement Hamiltonians of lattice models via the Bisognano-Wichmann theorem.” In: *Phys. Rev. B* 98 (13 Oct. 2018), p. 134403. DOI: <https://doi.org/10.1103/PhysRevB.98.134403>.
- [15] Gene H. Golub and Charles F. Van Loan. *Matrix Computations - 4th Edition*. Philadelphia, PA: Johns Hopkins University Press, 2013. ISBN: 978-1-4214-0794-4.
- [16] Jutho Haegeman. *KrylovKit*. Version 0.7.0. 2024. DOI: 10.5281/zenodo.10622234.
- [17] Philipp Hauke et al. “Measuring multipartite entanglement through dynamic susceptibilities.” In: *Nature Physics* 12.8 (Aug. 2016), pp. 778–782. DOI: 10.1038/nphys3700.
- [18] Nicholas J. Higham. *Functions of Matrices*. Society for Industrial and Applied Mathematics, 2008. DOI: 10.1137/1.9780898717778.
- [19] Ryszard Horodecki et al. “Quantum entanglement.” In: *Rev. Mod. Phys.* 81 (2 June 2009), pp. 865–942. DOI: 10.1103/RevModPhys.81.865.
- [20] Lukas Janssen. *Quantum Phase Transitions*. <https://tu-dresden.de/mn/physik/itp/ressourcen/dateien/arbeitsgruppe-quantum-critical-matter/qpt-ss24/qpt-pokart.pdf?lang=en>. 2024.
- [21] Steven G. Johnson. *QuadGK.jl: Gauss–Kronrod integration in Julia*. <https://github.com/JuliaMath/QuadGK.jl>. 2013.
- [22] Steven G. Johnson and Alan Edelman. *Matrix Calculus for Machine Learning and Beyond*. MIT course. Visited: 2024-12-24. 2024. URL: <https://github.com/mitmath/matrixcalc/tree/main?tab=readme-ov-file>.
- [23] *Julia 1.11 Documentation*. <https://docs.julialang.org/en/v1/>. 2024.
- [24] Jan Kierfeld. *Computational Physics*. https://cmt.physik.tu-dortmund.de/storages/cmt-physik/r/kierfeld/lecture_notes/kierfeld_CompPhys.pdf. 2022.

-
- [25] Christian Kokail et al. “Entanglement Hamiltonian tomography in quantum simulation.” In: *Nature Physics* 17.8 (Aug. 2021), pp. 936–942. DOI: <https://doi.org/10.1038/s41567-021-01260-w>.
 - [26] Christian Kokail et al. “Quantum Variational Learning of the Entanglement Hamiltonian.” In: *Phys. Rev. Lett.* 127 (17 Oct. 2021), p. 170501. DOI: <https://doi.org/10.1103/PhysRevLett.127.170501>.
 - [27] David Lay, Steven Lay, and Judi McDonald. *Linear Algebra and Its Applications Global Edition*. Pearson Deutschland, 2016, p. 576. URL: <https://elibrary.pearson.de/book/99.150005/9781292092249>.
 - [28] L. Lepori, G. De Chiara, and A. Sanpera. “Scaling of the entanglement spectrum near quantum phase transitions.” In: *Phys. Rev. B* 87 (23 June 2013), p. 235107. DOI: <https://doi.org/10.1103/PhysRevB.87.235107>.
 - [29] Hui Li and F. D. M. Haldane. “Entanglement Spectrum as a Generalization of Entanglement Entropy: Identification of Topological Order in Non-Abelian Fractional Quantum Hall Effect States.” In: *Phys. Rev. Lett.* 101 (1 July 2008), p. 010504. DOI: <https://doi.org/10.1103/PhysRevLett.101.010504>.
 - [30] Dong C. Liu and Jorge Nocedal. “On the limited memory BFGS method for large scale optimization.” In: *Mathematical Programming* 45.1 (Aug. 1989), pp. 503–528. DOI: 10.1007/BF01589116.
 - [31] Xiu-Zhe Luo et al. “Yao.jl: Extensible, Efficient Framework for Quantum Algorithm Design.” In: *arXiv preprint arXiv:1912.10877* (2019). DOI: <https://doi.org/10.48550/arXiv.1912.10877>.
 - [32] T. Mendes-Santos et al. “Entanglement Hamiltonian of quantum critical chains and conformal field theories.” In: *Phys. Rev. B* 100 (15 Oct. 2019), p. 155122. DOI: <https://doi.org/10.1103/PhysRevB.100.155122>.
 - [33] Patrick Kofod Mogensen and Asbjørn Nilsen Riseth. “Optim: A mathematical optimization package for Julia.” In: *Journal of Open Source Software* 3.24 (2018), p. 615. DOI: <https://doi.org/10.21105/joss.00615>.
 - [34] Awad H. Al-Mohy. *Generalizing the Fréchet Derivative Algorithm for the Matrix Exponential*. 2024. DOI: <https://doi.org/10.48550/arXiv.2410.03575>.
 - [35] Awad H. Al-Mohy and Nicholas J. Higham. “Computing the Fréchet Derivative of the Matrix Exponential, with an Application to Condition Number Estimation.” In: *SIAM Journal on Matrix Analysis and Applications* 30.4 (2009), pp. 1639–1657. DOI: 10.1137/080716426.
 - [36] Kazuo Murota and Takayasu Matsuo. *Double-Exponential transformation: A quick review of a Japanese tradition*. 2023. URL: <https://doi.org/10.48550/arXiv.2301.01920>.

- [37] Masaaki Nakamura. *DoubleExponentialFormulas*. Version 0.1.0. Mar. 24, 2021. URL: <https://machakann.github.io/DoubleExponentialFormulas.jl/stable/>.
- [38] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
- [39] Bernard Nienhuis, Massimo Campostrini, and Pasquale Calabrese. “Entanglement, combinatorics and finite-size effects in spin chains.” In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.02 (Feb. 2009), P02063. DOI: 10.1088/1742-5468/2009/02/P02063.
- [40] Jorge Nocedal. “Updating quasi-Newton matrices with limited storage.” In: *Mathematics of Computation* 35.151 (1980), pp. 773–782. DOI: <https://doi.org/10.1090/S0025-5718-1980-0572855-7>.
- [41] Frank Pollmann et al. “Entanglement spectrum of a topological phase in one dimension.” In: *Phys. Rev. B* 81 (6 Feb. 2010), p. 064439. DOI: <https://doi.org/10.1103/PhysRevB.81.064439>.
- [42] John Preskill. “Quantum Computing in the NISQ era and beyond.” In: *Quantum* 2 (Aug. 2018), p. 79. DOI: 10.22331/q-2018-08-06-79.
- [43] Nicolas Regnault. “Entanglement spectroscopy and its application to the quantum Hall effects.” In: *Topological Aspects of Condensed Matter Physics: Lecture Notes of the Les Houches Summer School: Volume 103, August 2014*. Oxford University Press, Jan. 2017. DOI: <https://doi.org/10.1093/acprof:oso/9780198785781.003.0004>.
- [44] Federico Rottoli, Michele Fossati, and Pasquale Calabrese. “Entanglement Hamiltonian in the non-Hermitian SSH model.” In: *Journal of Statistical Mechanics: Theory and Experiment* 2024.6 (June 2024), p. 063102. DOI: <https://doi.org/10.1088/1742-5468/ad4860>.
- [45] J. T. Schneider, S. J. Thomson, and L. Sanchez-Palencia. “Entanglement spectrum and quantum phase diagram of the long-range XXZ chain.” In: *Phys. Rev. B* 106 (1 July 2022), p. 014306. DOI: <https://doi.org/10.1103/PhysRevB.106.014306>.
- [46] Hidetosi Takahasi and Masatake Mori. “Double Exponential Formulas for Numerical Integration.” In: *Publications of the Research Institute for Mathematical Sciences* 9.3 (1973), pp. 721–741. DOI: 10.2977/PRIMS/1195192451.
- [47] C. Van Loan. “Computing integrals involving the matrix exponential.” In: *IEEE Transactions on Automatic Control* 23.3 (1978), pp. 395–404. DOI: 10.1109/TAC.1978.1101743.

- [48] Matthias Vojta. *Thermal and Quantum Phase Transitions*. https://tu-dresden.de/mn/physik/itp/ressourcen/dateien/arbeitsgruppe-quantum-critical-matter/qpt-ss24/vojta_quantum-phase-transitions.pdf?lang=en. 2015.
- [49] Qian Wang et al. “AUGEM: Automatically generate high performance Dense Linear Algebra kernels on x86 CPUs.” In: *SC '13: Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. 2013, pp. 1–12. DOI: 10.1145/2503210.2503219.
- [50] Zhang Xianyi, Wang Qian, and Zhang Yunquan. “Model-driven Level 3 BLAS Performance Optimization on Loongson 3A Processor.” In: *2012 IEEE 18th International Conference on Parallel and Distributed Systems*. 2012, pp. 684–691. DOI: 10.1109/ICPADS.2012.97.

Eidesstattliche Versicherung

(Affidavit)

Kind, Yanick Sebastian

Name, Vorname
(surname, first name)

218169

Matrikelnummer
(student ID number)

☐ Bachelorarbeit
(Bachelor's thesis)

☒ Masterarbeit
(Master's thesis)

Titel
(Title)

Exploring Entanglement Hamiltonians with
Variational Algorithms

Ich versichere hiermit an Eides statt, dass ich die vorliegende Abschlussarbeit mit dem oben genannten Titel selbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

I declare in lieu of oath that I have completed the present thesis with the above-mentioned title independently and without any unauthorized assistance. I have not used any other sources or aids than the ones listed and have documented quotations and paraphrases as such. The thesis in its current or similar version has not been submitted to an auditing institution before.

Dortmund, 05.03.2025

Ort, Datum
(place, date)


Unterschrift
(signature)

Belehrung:

Wer vorsätzlich gegen eine die Täuschung über Prüfungsleistungen betreffende Regelung einer Hochschulprüfungsordnung verstößt, handelt ordnungswidrig. Die Ordnungswidrigkeit kann mit einer Geldbuße von bis zu 50.000,00 € geahndet werden. Zuständige Verwaltungsbehörde für die Verfolgung und Ahndung von Ordnungswidrigkeiten ist der Kanzler/die Kanzlerin der Technischen Universität Dortmund. Im Falle eines mehrfachen oder sonstigen schwerwiegenden Täuschungsversuches kann der Prüfling zudem exmatrikuliert werden. (§ 63 Abs. 5 Hochschulgesetz - HG -).

Die Abgabe einer falschen Versicherung an Eides statt wird mit Freiheitsstrafe bis zu 3 Jahren oder mit Geldstrafe bestraft.

Die Technische Universität Dortmund wird ggf. elektronische Vergleichswerkzeuge (wie z.B. die Software „turnitin“) zur Überprüfung von Ordnungswidrigkeiten in Prüfungsverfahren nutzen.

Die oben stehende Belehrung habe ich zur Kenntnis genommen:

Official notification:

Any person who intentionally breaches any regulation of university examination regulations relating to deception in examination performance is acting improperly. This offense can be punished with a fine of up to EUR 50,000.00. The competent administrative authority for the pursuit and prosecution of offenses of this type is the Chancellor of TU Dortmund University. In the case of multiple or other serious attempts at deception, the examinee can also be unenrolled. Section 63 (5) North Rhine-Westphalia Higher Education Act (*Hochschulgesetz, HG*).

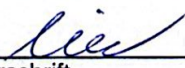
The submission of a false affidavit will be punished with a prison sentence of up to three years or a fine.

As may be necessary, TU Dortmund University will make use of electronic plagiarism-prevention tools (e.g. the "turnitin" service) in order to monitor violations during the examination procedures.

I have taken note of the above official notification:*

Dortmund, 05.03.2025

Ort, Datum
(place, date)


Unterschrift
(signature)

***Please be aware that solely the German version of the affidavit ("Eidesstattliche Versicherung") for the Bachelor's/ Master's thesis is the official and legally binding version.**